

# The Epistemology of Fake News

## ENGAGING PHILOSOPHY

This series is a new forum for collective philosophical engagement with controversial issues in contemporary society.

### *Disability in Practice*

*Attitudes, Policies, and Relationships*

Edited by Adam Cureton and Thomas E. Hill, Jr.

### *Taxation*

*Philosophical Perspectives*

Edited by Martin O'Neill and Shepley Orr

### *Bad Words*

*Philosophical Perspectives on Slurs*

Edited by David Sosa

### *Academic Freedom*

Edited by Jennifer Lackey

### *Lying*

*Language, Knowledge, Ethics, and Politics*

Edited by Eliot Michaelson and Andreas Stokke

### *Treatment for Crime*

*Philosophical Essays on Neurointerventions in Criminal Justice*

Edited by David Birks and Thomas Douglas

### *Games, Sports, and Play*

*Philosophical Essays*

Edited by Thomas Hurka

### *Effective Altruism*

*Philosophical Issues*

Edited by Hilary Greaves and Theron Pummer

### *Philosophy and Climate Change*

Edited by Mark Budolfson, Tristram McPherson, and David Plunkett

# The Epistemology of Fake News

*Edited by*

SVEN BERNECKER, AMY K. FLOWERREE,  
AND THOMAS GRUNDMANN

**OXFORD**  
UNIVERSITY PRESS

**OXFORD**  
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,  
United Kingdom

Oxford University Press is a department of the University of Oxford.  
It furthers the University's objective of excellence in research, scholarship,  
and education by publishing worldwide. Oxford is a registered trade mark of  
Oxford University Press in the UK and in certain other countries

© Oxford University Press 2021  
Introduction © Sven Bernecker, Amy K. Flowerree, and Thomas Grundmann 2021;  
Chapter 6 © Thomas Grundmann 2021; Chapter 13 © Sven Bernecker 2021

The moral rights of the authors have been asserted

First Edition published in 2021

Impression: 1

All rights reserved. No part of this publication may be reproduced, stored in  
a retrieval system, or transmitted, in any form or by any means, without the  
prior permission in writing of Oxford University Press, or as expressly permitted  
by law, by licence or under terms agreed with the appropriate reprographics  
rights organization. Enquiries concerning reproduction outside the scope of the  
above should be sent to the Rights Department, Oxford University Press, at the  
address above

You must not circulate this work in any other form  
and you must impose this same condition on any acquirer

Published in the United States of America by Oxford University Press  
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data

Data available

Library of Congress Control Number: 2020953013

ISBN 978-0-19-886397-7

DOI: 10.1093/oso/9780198863977.001.0001

Printed and bound by  
CPI Group (UK) Ltd, Croydon, CR0 4YY

Links to third party websites are provided by Oxford in good faith and  
for information only. Oxford disclaims any responsibility for the materials  
contained in any third party website referenced in this work.

# Preface

We are convinced that the epistemological pathologies of knowledge acquisition and dissemination in the age of the Internet, social media, and political polarization call for a readjustment of social epistemology. To help bring about this readjustment we organized an international conference on *Fake Knowledge* in June 2018 with more than thirty speakers. The conference was hosted by the Cologne Center for Contemporary Epistemology and the Kantian Tradition. We thank the Alexander-von-Humboldt Foundation for generously funding the conference and Sibel Schmidt and her team consisting of Karolin Meinert and Jakob Ohlhorst for running the conference.

Six chapters in this book are based on lectures given at the *Fake Knowledge* conference. The other chapters have been added to broaden the diversity of perspectives. We are deeply grateful to the authors for their excellent contributions and for their philosophical enthusiasm and patience during the review and production process. Furthermore, we would like to thank these colleagues for advice, comments, and support: Jason Baehr, Wout Bisschop, Kenneth Boyd, Quassim Cassam, Filippo Ferrari, Sanford Goldberg, Peter Graham, Alex Guerrero, Klemens Kappel, Brian Keeley, Brent Kyle, Pierre Le Morvan, Neil Levy, Lee McIntyre, Robin McKenna, Nikil Mukerji, Andy Müller, Nikolaj Nottelmann, Jakob Ohlhorst, Rik Peels, Jessica Pepp, Tommasso Piazza, Luis Rosa, Joseph Shieber, Rachel Sterken, Emily Sullivan, Joseph Uscinski, and René van Woudenberg. Last but not least, we would like to thank Oxford University Press, and especially Peter Momtchiloff, for supporting the idea of the volume and for making the editing process go smoothly.

Sven Bernecker  
Amy K. Flowerree  
Thomas Grundmann



# Notes on Contributors

**Michael Baurmann** studied sociology, philosophy and law. In 1997 he was appointed to a chair for sociology at the University of Düsseldorf. He has been visiting professor at the Research School of Social Sciences of the Australian National University, the Instituto Tecnológico Autónomo in Mexico and at the Department of Politics at the New York University. Currently he is scientific director of the Center for Advanced Internet Research (CAIS) in Bochum. He has been co-editor of *Analyse & Kritik: Journal for Social Theory* for forty years. His main research interests are rational choice theory, social epistemology, and digital democracy.

**Sven Bernecker** is Humboldt Professor of Philosophy at the University of Cologne and Professor of Philosophy at the University of California, Irvine. His main areas of research are epistemology, metaphysics, and philosophy of mind, and he has published numerous articles in these areas. He is the author of *Reading Epistemology* (Blackwell, 2006), *The Metaphysics of Memory* (Springer, 2008), and *Memory* (Oxford University Press, 2010). He is co-editor of *Knowledge* (Oxford University Press, 2000), *Companion to Epistemology* (Routledge, 2011), *Handbook of Philosophy of Memory* (Routledge, 2017), and *Medical Knowledge in a Social World*, special issue of *Synthese* (2019).

**David Coady** is a senior lecturer in philosophy at the School of Humanities at the University of Tasmania. He has published on a variety of topics in applied epistemology, including expertise, conspiracy theory, rumour, propaganda, Wikipedia, and the blogosphere. He has also published on metaphysics, the philosophy of law, scientific whaling, police ethics, the ethics of horror films, and the ethics of cricket. He is the author of *What to Believe Now: Applying Epistemology to Contemporary Issues* (2012), the co-author of *The Climate Change Debate: An Epistemic and Ethical Inquiry* (2013), the editor of *Conspiracy Theories: The Philosophical Debate* (2006), co-editor of the *Blackwell Companion to Applied Philosophy* (2017), and co-editor of the *Routledge Handbook of Applied Epistemology* (2018).

**Daniel Cohnitz** (PhD 2005, Düsseldorf) holds the Chair for Theoretical Philosophy at Utrecht University. His research interests include metaphilosophy, social epistemology, the philosophy of language, mathematics, logic, and related areas of philosophy and cognitive science. His most recent book is *An Introduction to the Philosophy of Logic* that he co-authored with Luis Estrada-González and which was published by Cambridge University Press in 2019.

**Catarina Dutilh Novaes** is professor of philosophy and University Research Chair at Vrije Universiteit Amsterdam, and Professorial Fellow at Arché, University of St Andrews. She is the author of *Formalizing Medieval Logical Theories* (Springer, 2007), *Formal Languages in Logic* (Cambridge University Press, 2012), and *The Dialogical Roots of Deduction* (Cambridge University Press, 2021), as well as of numerous journal articles and book

chapters. She is currently (2018–23) the Principal Investigator of the ERC-Consolidator project ‘The Social Epistemology of Argumentation’.

**Filippo Ferrari** was awarded his PhD by Aberdeen University in 2014. He is currently a Stars@UNIPD postdoc at the University of Padua and the Principal Investigator of the project ‘Logical Anti-Exceptionalism’. His main research interests lie at the intersection between philosophy of language, philosophy of logic, and epistemology. He has published several articles on a variety of topics—including the nature and epistemological significance of disagreement, the nature and value of truth, logical pluralism and logical normativity—in journals such as *Philosophical Quarterly*, *Synthese*, *Analysis*, *Mind*, *Inquiry*, *Canadian Journal of Philosophy*, and *American Philosophical Quarterly*. He has recently co-authored a book on post-truth and science denialism with Sebastiano Moruzzi (published in 2020 with1088 press).

**Amy K. Flowerree** is Assistant Professor of Philosophy at Texas Tech University. Her work is at the intersection of epistemology, ethics, and metaethics. She focuses on role of belief within agency. She has written on epistemic agency, practical reasons for belief, and the metaethical foundations of epistemic normativity. Her work has been published in *Philosophical Studies*, *Synthese*, and several edited volumes.

**Axel Gelfert** is a Professor of Theoretical Philosophy at the Technical University of Berlin, where he heads the Institute of History and Philosophy of Science, Technology, and Literature. Much of his work focuses on the intersection of social epistemology and philosophy of science and technology. He is the author of *A Critical Introduction to Testimony* (Bloomsbury, 2014) and *How to Do Science With Models* (Springer, 2016).

**Emmanuel J. Genot** is an Associate Researcher in Theoretical Philosophy at Lund University, Sweden. His areas of research include philosophical logic, game-theoretic semantics, and formal models of inquiry and discovery. His most recent publication is ‘The Holmesian Logician: Sherlock Holmes’ Science of Deduction and Analysis and the Logic of Discovery’ (*Synthese*, 2020). He is presently co-investigator in a project funded by The Swedish Foundation for Humanities and Social Sciences investigating personalized search, filter bubbles, and polarization in online search.

**Sanford C. Goldberg** is Chester D. Tripp Professor in the Humanities and Professor of Philosophy at Northwestern University. He works primarily in the areas of epistemology and philosophy of language. His books include *Collected Essays in Social Epistemology* (Oxford University Press, forthcoming), *Conversational Pressure* (Oxford University Press, 2020), *To the Best of Our Knowledge* (Oxford University Press, 2018), *Assertion* (Oxford University Press, 2015), *Relying on Others* (Oxford University Press, 2010), and *Anti-Individualism* (Cambridge University Press, 2007).

**Thomas Grundmann** is Professor of Philosophy at the University of Cologne, Germany. He has published numerous articles on epistemology, philosophical methodology, and analytic existential philosophy. Among his books are *Die philosophischen Wahrheitstheorien* (2018) and *Analytische Einführung in die Erkenntnistheorie* (2017). His current research addresses issues from social epistemology such as epistemic authority and experts, the limits of critical thinking, disagreement, and post-truth. He is the co-editor of *Experimental Philosophy and Its Critics* (2012).

**Romy Jaster** is a postdoctoral researcher at Humboldt University Berlin. Her work revolves around abilities, dispositions, and free will, as well as fake news, public debate, and the possibilities and limits of constructive discourse. She studied in Bielefeld, at the University of Notre Dame, and Humboldt University, which is also where she received her PhD. Her book *Agents' Abilities* was published by De Gruyter in 2020. Together with her co-author David Lanius, she wrote a book on fake news (in German), which was published by Reclam in 2019. Apart from her research, she co-heads (with David Lanius) the German 'Forum for Streitkultur', a think tank for the promotion of public discourse in the German public debate.

**Jennifer Lackey** is the Wayne and Elizabeth Jones Professor of Philosophy at Northwestern University, the Director of the Northwestern Prison Education Program, and Editor-in-Chief of *Philosophical Studies* and *Episteme*. Most of her research is in the area of social epistemology, with recent projects including work on the rationality of punishment, credibility and false confessions, eyewitness testimony and epistemic agency, epistemic reparations, and echo chambers and fake news. Lackey is the winner of the Dr. Martin R. Lebowitz and Eve Lewellis Lebowitz Prize for Philosophical Achievement and Contribution as well as the Young Epistemologist Prize, and she has received grants and fellowships from the Andrew W. Mellon Foundation, the American Council of Learned Societies, the National Endowment for the Humanities, and the Alice Kaplan Institute for the Humanities.

**David Lanius** is postdoctoral researcher at DebateLab of Karlsruhe Institute of Technology (KIT) with focus on strategic indeterminacy in law and politics, argumentative patterns of populism, fake news, public debate, and the possibilities and limits of constructive discourse. He studied philosophy, economics, and logic in Regensburg, Munich, Madrid, and Amsterdam and received his PhD in philosophy from Humboldt University Berlin. His book *Strategic Indeterminacy in the Law* was published by Oxford University Press in 2019. Together with his co-author Romy Jaster, he also wrote a book on fake news (in German), which was published by Reclam in 2019. Apart from his research, he co-heads (with Romy Jaster) the German 'Forum für Streitkultur', which is a think tank for the promotion of constructive discourse in the German public debate.

**Sebastiano Moruzzi** is associate professor in philosophy of language at the University of Bologna. He obtained the title of PhD from the Doctorate in Philosophy of Language at the University of Eastern Piedmont and then worked as research fellow at the research center of Arché at the University of St Andrews. He is scientific director of the Bologna Cogito research group in philosophy, and he has founded the research group Aiòn on teaching philosophy and philosophical practices. He has published several articles in international journals and volumes on vagueness, relativism, alethic pluralism, epistemology, and philosophy of logic. He has published an Italian book on the problem of vagueness and a book on post-truth with Filippo Ferrari. In addition to writing about philosophy, he collaborates with schools to conduct experimental sessions of philosophical practices for popularizing philosophy among children and teenagers.

**M. Giulia Napolitano** is a PhD student at the University of California, Irvine, where she is currently working on her dissertation on the philosophy of conspiracy theories. Before that, she was a PhD student at CONCEPT, University of Cologne, and she studied philosophy at

the University of Padua and at Utrecht University. Her interests lie mainly in epistemology, philosophy of language, and political philosophy.

**Erik J. Olsson** is Professor in Theoretical Philosophy at Lund University, Sweden. His areas of research include epistemology, philosophical logic, pragmatism, and, more recently, epistemological aspects of social networks and search engines. His books include *Against Coherence: Truth, Probability, and Justification* (Oxford University Press, 2005, paperback 2008), *Knowledge and Inquiry: Essays on the Pragmatism of Isaac Levi* (Cambridge University Press, 2006) and *Belief Revision Meets Philosophy of Science* (Springer, 2011). He is presently leading a research project funded by The Swedish Foundation for Humanities and Social Sciences investigating personalized search, filter bubbles, and polarization in Google.

**Maura Priest** is an assistant professor of philosophy and bioethicist at Arizona State University. She has a PhD in philosophy from UC Irvine (2016), an MS in bioethics from Columbia University (2018), and a certificate in pediatric bioethics from Children's Mercy hospital. She has published many journal articles and book chapters focused on the intersections of ethics and epistemology, applied ethics (especially bioethics), social and political philosophy, and social ontology. Her current research projects include two books, one on intellectual elites and another on the ethics of dating. She is the editor of ASU's applied philosophy webpage, chair of the diversity committee, and supervises several MA and PhD thesis/dissertation projects.

**Duncan Pritchard** is UC Distinguished Professor of Philosophy at the University of California, Irvine. His books include *Epistemic Luck* (Oxford University Press, 2005), *The Nature and Value of Knowledge* (co-authored, Oxford University Press, 2010), *Epistemological Disjunctivism* (Oxford University Press, 2012), *Epistemic Angst* (Princeton University Press, 2015), and *Scepticism: A Very Short Introduction* (Oxford University Press, 2019).

**Jeroen de Ridder** is Associate Professor of Philosophy at Vrije Universiteit Amsterdam and Professor (by special appointment) of Christian Philosophy at the University of Groningen. His research focuses on issues in social epistemology, philosophy of science, and philosophy of religion. His work has been published in *The Australasian Journal of Philosophy*, *Synthese*, *Erkenntnis*, and *Episteme* and funded by several grants from the Netherlands Research Council (NWO) and the Templeton World Charity Foundation. He is co-editor of *Scientism: Prospects and Problems* (Oxford University Press, 2018), *Scientific Challenges to Common Sense Philosophy* (Routledge, 2020), and *The Routledge Handbook of Political Epistemology* (Routledge, 2021).

**Sarah Wright** is an Associate Professor of Philosophy at the University of Georgia. Her research focuses on the normative aspects of epistemology, particularly on the epistemic virtues as virtues of our psychological character. She has also written on contextualism in epistemology, on social and group epistemology, and on environmental ethics. She is currently working on applications of virtue epistemology to particular real-life situations, including concerns with epistemic injustice and the politics of knowing. Her work has been published in numerous edited volumes as well as in *Episteme*, *Philosophical Issues*, *Acta Analytica*, *History of Philosophy Quarterly*, *Ethics and the Environment*, and *Metaphilosophy*.

# Introduction

*Sven Bernecker, Amy K. Flowerree, and Thomas Grundmann*

## 1. The Challenge from Fake News

News matters. Democracies need independent, fact-based news to provide a voice for a diverse range of people, to watchdog the powerful, and to keep members of a society informed. News is the basis for the public to make informed decisions in elections and referenda and for public officials to act on behalf of the public. By letting the public monitor government performance, the news serves as a mechanism for democratic accountability and socializes citizens into democratic attitudes and values. When the news has epistemic standing, citizens who pay attention to the news are likely to be able to cast more informed ballots, hold public officials accountable, and be more supportive of democratic processes and values (Goidel et al. 2017: 836; Goldman 1999: ch. 10).

The epistemic quality of news depends, in the first instance, on its truth value. Good news is accurate news. The accuracy of a news item, while necessary, is not sufficient to render it epistemically valuable. In order to be epistemically valuable, news must also be issued by a media channel (newspaper, television station, website, etc.) that meets the conditions for content and coverage reliability. Content reliability consists in a media channel's tendency to produce accurate contents. In other words, it is not enough that a media channel issues accurate news once in a while. Inaccurate news must be the exception. Coverage reliability, on the other hand, refers to the tendency of a media channel to keep its audience reliably apprised of the relevant facts in a certain domain and to reliably report regarding the obtaining these facts (Goldberg 2010: 157).

If we take 'news' in a broad sense, to cover all kinds of public information, not just organized reports of new developments, it is quite clear that much of the news surrounding us today does not, for one reason or the other, meet the standards of epistemically valuable news. Our media environments are polluted by inaccurate news and other forms of mis- and disinformation. In the United States, for instance, a significant portion of the publicly disseminated information in the context of the 2016 presidential election was false and was written either to make money or to mislead the public (Allcott & Gentzkow 2017). In the United Kingdom, a political consulting firm, Cambridge Analytica, used misinformation to help influence the outcome of the Brexit referendum (Rawlinson 2020).

Moreover, social media outlets, such as Facebook, have enabled unreliable news and conspiracy theories to emerge as dominant political tools. For instance, in the three months leading up to the 2016 U.S. elections, the top twenty fake news stories on Facebook got more shares than the top twenty real news stories (McIntyre 2018: 109).

It is an open question whether the flood of unreliable news we are encountering today is an old or a new phenomenon (see Dutilh Novaes and de Ridder, Chapter 7 in this volume). Even if the phenomenon is not entirely new, today's information technology makes the problem worse. There are a number of factors that contribute to the worsening of the threat posed by unreliable news. First, the Internet has led to a democratization of news production. Anyone can operate their own news channel. There is no filter in place that controls for quality (Goldman 1999: 187; McIntyre 2018: 95). Second, in the absence of effective quality control, news channels are not penalized for increasing, and may even be commercially incentivized to increase, their audience at the expense of the epistemic quality of the news they disseminate (Goldman 1999: 182–8). Third, social media (e.g., Twitter) spreads inaccurate news significantly farther, faster, deeper, and more broadly than genuine news. The reason is that false news tends to be more novel and that people are more likely to share novel information (Vosoughi et al. 2018). Fourth, search engines and social media, due to their recommendation and personalization algorithms, seem to lead to filter bubbles and echo chambers (Pariser 2011; Sunstein 2017), which are responsible for ideological polarization: We no longer encounter a balanced information diet, but only see information that targets our established interests and reinforces our existing world views. (More on this issue in Chapter 9 by Lackey and Chapter 10 by Genod and Olsson in this volume.) Fifth, there is a lack of diversity of news in the digital age. Even though there is a larger variety of news publishers, there is less variety of news because a large part of the news across a wide range of news publishers can be traced back to the same news agencies, or wire services (Welbers et al. 2018). Last but not least, the interaction and collaboration between journalists and public relations practitioners in the production of mass media news content has arguably led to a decline of professional journalism (Simons & Strovsky 2019).

## 2. The Need for a New Epistemology of Fake News

If there is a place within epistemology that should be dealing with the problem of unreliable news it might seem that it should be the epistemology of testimony. The epistemology of testimony concerns the epistemic status (justification, knowledge) of beliefs we form on the basis of what others tell us. Since some of our testimonial beliefs are based on news reports, it is natural to suppose that the issue of

unreliable news falls within the purview of the epistemology of testimony. However, the epistemology of testimony is ill-equipped to deal with the issue of this type of news. Let us explain.

The central debate in the epistemology of testimony concerns the nature of the justification or warrant of testimonially obtained beliefs. The two dominant views—reductionism and anti-reductionism—differ about the default rule for accepting testimony. Anti-reductionists hold that the norm of truth-telling is prevalent in society, such that by default, a testimony can be presumed to be correct unless the recipient has reasons to question it (Burge 1993; Coady 1992; Goldberg 2010). Reductionists, by contrast, hold that testimony is generally too unreliable to be believed by default, and argue that a recipient of testimony must check the trustworthiness of the source before they may believe it (Fricker 1995).

It is important to realize that reductionists and anti-reductionists make certain assumptions about the communication situation that do not (always) hold in the digital age. For starters, epistemologists of testimony generally assume that the recipient of testimony knows the testifier, at least by name. Many online news contents, however, are anonymous. When the recipient does not know who the testifier is (let alone whether they are a real person or an AI/chatbot), they cannot, in principle, vet the testifier's trustworthiness and competence (Goldberg 2013). This connects with the next point. The recipient of an electronic message may know and trust the sender but may be mistaken in thinking that the sender is also the author. When electronic messages and contents are forwarded or re-posted, the recipient of testimony can easily mix up the trustworthiness of the sender with the trustworthiness of the author. (More on this in Chapter 11 by Wright in this volume.) And even when the recipient is able to tell apart the sender from the author, they may still be in the dark about the relevant social context and the sender's communicative intention. Are re-posts and re-tweets assertions or some other kind of speech act (Pepp et al. 2019a; Rini 2017)? Finally, online news contents is often not text-based but in the format of visual imagery. The epistemology of testimony tends not to deal with the special issues connected with the acquisition of information through visual communication (Fallis 2019; Rini 2020).

The upshot of all of this is that the epistemology of testimony is ill-equipped to deal with the problem of unreliable news or, more specifically (as we will detail in Section 3), it is ill-equipped to deal with the problem of *fake news* because it focuses on the transfer of knowledge and information under ideal conditions. The new epistemology of fake news, which this volume launches, can be characterized as applied epistemology for knowledge communication under non-ideal conditions.

### 3. Contours of an Epistemology of Fake News

#### 3.1 A Fundamental Worry

We believe that it is advisable to use the term ‘fake news’ to cover all cases of epistemically corrupt news. Before the meaning of ‘fake news’ will be made transparent in the next section, a fundamental worry about the scientific use of this term must be addressed. A number of scholars have argued that the term ‘fake news’ is linguistically defective and should thus be abandoned from academic discourse (e.g., Chapter 3 by Coady in this volume; Oremus 2016; Sullivan 2017; Talisse 2018). In particular, Joshua Habgood-Coote launches three attacks against ‘fake news.’ According to the *objection from semantic instability* (Habgood-Coote 2019: 1036–47), the use of the term ‘fake news’ has significantly changed during the last decade and varies greatly among speakers, including experts. Until 2015, the term was typically used to refer to satiric news parodies of the kind found in *The Onion* and only rarely applied to false media reports. This changed with the 2016 U.S. presidential election (Brown 2019: 145–6). But even then, the meaning of the term continued to be unstable. Some experts use the term to refer to false or misleading news (Levy 2017), whereas others take the author’s deceptive intention to be the essential feature of fake news (Rini 2017). The variability of usage is thought to indicate that the term ‘fake news’ lacks a determinate descriptive meaning or is altogether nonsensical. Needless to say, academic discourse should avoid such a shaky basis.

According to the *objection from redundancy* (Habgood-Coote 2019: 1047–9), the term ‘fake news’ is not necessary to describe the pathologies of social epistemology it is meant to refer to. As Habgood-Coote puts it:

We already have plenty of words for talking about deceit, miscommunication, bullshitting, false assertion, false implicature, being unreliable, distorting the facts, being biased, propaganda, and so on. These terms have perfectly good meanings in ordinary language. (Habgood-Coote 2019: 1049)

Adding the term ‘fake news’ to our conceptual resources is superfluous. And finally, the *objection from propagandistic usage* (Habgood-Coote 2019: 1050–4) states that the term ‘fake news’ amounts to nothing more than an epistemic slur meant to discredit the opponent. Donald Trump’s ubiquitous use of this term is a case in point. Slurs and insults, so the argument goes, have no place in academic discourse.

What might be said in response to these objections? Let us start with the objection from semantic instability. First, while it is true that the term ‘fake news’ lacks a well-entrenched usage, there is a cluster of social phenomena that the term tries to characterize. The challenge is to find the relevant underlying phenomenon first and then to restrict the term’s use to this phenomenon. In this

respect, the definition of the term ‘fake news’ has to be partly stipulative, as it is usually the case with Carnapian explications (Pepp et al. 2019b: 7). After all, experts converge, at least to a certain extent, about what the relevant phenomena defining ‘fake news’ are (Brown 2019: 146). Second, the objection from semantic instability is unfair. There are other terms for social phenomena such as ‘gender,’ ‘sexual harassment,’ or ‘propaganda’ that exhibit the same semantic instability as does ‘fake news’ but whose role in academic discourse is not called into question (Brown 2019: 146; Pepp et al. 2019b: 4).

Next, consider the objection from redundancy. The objection can be interpreted in two distinct ways. According to one reading, what is driving the objection is the idea that a term that is definable in terms of independent terms should be replaced by its *definiens*. Yet if this is so, then we have to abandon not only ‘fake news’ but also ‘propaganda,’ ‘lie,’ or ‘false assertion’ since these terms are also definable in terms of more basic ones. Habgood-Coote’s plea for holding onto the latter terms while abandoning the former appears unreasonably partial (Pepp et al. 2019b: 8). Furthermore, if the initial terms are replaced with the defining terms, the connection to the terms used in public discourse is lost. This is why it does not seem to be a good idea to get rid of all definable terms in favor of the more basic ones. According to an alternative reading of the objection from redundancy, ‘fake news’ is nothing but a cluster term referring to a number of well-known epistemic pathologies such as deceit, miscommunication, bullshitting, false assertion, etc. Yet it is far from clear that ‘fake news’ does not pick out a specific and so far unnamed phenomenon of its own (Brown 2019: 148).

Finally, let us consider the objection from propagandistic usage. For terms to be exploitable for propagandistic purposes, their meaning must contain a negative evaluation. ‘Fake news’ clearly has a negative evaluative meaning. Labeling a piece of news ‘fake news’ means that it would be epistemically bad to rely on it. However, not all terms that have such an evaluative meaning are abandoned from academic discourse (Brown 2019: 151). As long as terms also have descriptive meaning that makes the (negative) evaluation reasonable, they are deemed acceptable. Consider, for example, the terms ‘murderer,’ ‘fraudster,’ ‘liar,’ or ‘propaganda.’ Each of these terms is such that its descriptive meaning grounds its evaluative meaning. For example, a murderer is a morally bad person (evaluative meaning) *because* they killed someone unlawfully and with premeditation (descriptive meaning). These kinds of terms are perfectly acceptable for the academic discourse. Why treat ‘fake news’ differently? We recognize that there is a problem with use of ‘fake news’ as our term for unreliable and/or misleading news: The literal meaning of the expression has been to some extent undermined by its cynical and insincere deployment for propagandist purposes. Nevertheless, we believe that ‘fake news’ has become and will remain the most commonly used label for the growing phenomenon of publicly shared information that is unreliable and/or misleading, and so we have decided to retain it.

In sum, even though ‘fake news’ is a disputed term, its use can be regulated and tweaked in such a way that epistemologists may keep talking about fake news as other academics do outside philosophy (see Lazer et al. 2018).

## 3.2 Approaches to the Epistemology of Fake News

An epistemological investigation of fake news needs to address three key questions: first, what is fake news; second, what are the mechanisms that foster the production and spread of fake news; and third, which therapies are available as an antidote to fake news? The chapters of the volume address these questions from different perspectives. The purpose of the following remarks is to set the stage.

### 3.2.1 What is Fake News?

Beyond the loose talk of fake news as false or misleading media stories (Levy 2017), there are three main accounts of fake news in the current literature. According to the generally prevailing *hybrid view*, fake news is news that is both lacking truth and truthfulness (Jaster & Lanius 2018 and Chapter 1 in this volume; see also Rini 2017; Gelfert 2018; McIntyre 2018; Mukerij 2018). News fails to be true if it is either literally false or conveys false information. News lacks truthfulness if it is produced with the intention to deceive or without any concern for the truth in the relevant domain. In the former case, fake news is closely related to public lies or disinformation. In the latter case, fake news is typically stories that have been fabricated with the aim of attracting attention in the Internet or social media. Here, the deception of the recipient is a byproduct as opposed to the target. This is illustrated by the case of the Macedonian teenagers who, back in 2016, fabricated exciting stories, including one according to which Pope Francesco endorsed Trump as the next U.S. president, and spread these stories via the Internet to attract as much attention as possible. The hybrid view of fake news is widely applicable and emphasizes the key role of the producer’s bad intention, but it lacks theoretical unity. There is no clear thread that unites the disjunctive account of the producer’s intentions. Whether someone lies or is simply not caring about the truth when making assertions are two very different things that, according to the hybrid view, can be both characteristic of fake news.

According to the *privative view*, fake news fails to be genuine news since it lacks the required pedigree of being produced by standard journalistic processes (Fallis & Mathiesen 2019; Lazer et al. 2018; Pepp et al. 2019a; Pritchard, Chapter 2 in this volume). Fake news thus consists of stories published by the media that falsely appear to be genuine news. This view captures an important aspect of fake news as mimicking real news. It is, however, less clear why this category should have any epistemological significance. For even stories that are mistakenly treated as news might still be epistemically valuable and reliable.

Finally, there is the *consumer-centered view* (Grundmann 2020), which defines fake news exclusively in terms of its systematic disposition to deceive its consumers. This view applies very broadly—maybe too broadly because even cases of deception through news bias or bad journalism are included. This view may be rendered even more inclusive if even accurate news can be fake news when the consumer assumes they have been informed about all relevant facts (see Chapter 13 by Bernecker in this volume).

### 3.2.2 What Are the Mechanisms that Foster the Production and Spread of Fake News?

For now, we are going to take our subject-matter to include anything covered by one of the different views. As the volume proceeds, different authors will sharpen the focus in different ways. But what explains the proliferation of fake news and what are the causally relevant factors? We can distinguish three structural features that tend to promote the proliferation of fake news: (i) features of communication technology, (ii) social conditions, and (iii) epistemic ideologies. It is these factors that render our epistemic ecosystem non-ideal.

How can features of the communication technology further the spread of fake news? Here are some suggestions: Internet platforms such as Google or Yahoo rank news through algorithm-driven search engines. These algorithms favor engagement with content over quality of content when ranking news (Chapter 10 by Genot & Olsson in this volume; Lazer et al. 2018). Users of search engines are more likely to be exposed to interesting news rather than quality information. Moreover, there is the worry that the algorithms of the Internet and social media generate filter bubbles (Pariser 2011; Sunstein 2017). The news consumer inhabits a news filter bubble if the news outlet selects for news that fits with what they already think about the world. This personalized exposure to news is based on the news consumer's recorded past search behavior on the net. It is still an open question whether the Internet and social media generate filter bubbles and if these effects are significant (see Curtois et al. 2018; Haim et al. 2017; Hannack et al. 2013). However, if they do, this would explain why people with fringe views have a hard time escaping their insular news universes. Finally, communication via social media such as Facebook, Twitter, or Instagram requires that the user quickly responds by liking or sharing received news. From an epistemological point of view, this generates at least two problems. First, users simply do not find the time to critically reflect before they re-distribute the news and thus often distribute garbage (Vosoughi et al. 2018; Chapter 11 by Wright in this volume). Second, the speech act of sharing is itself ambiguous between calling for attention and asserting what is shared and recipients may confuse one with the other (Rini 2017).

Social conditions also contribute to the proliferation of fake news. People who inhabit environments that are characterized by group polarization, identity-based cognition, a high degree of anxiety, and a felt loss of control, are less likely to give

equal weight to all available news. News outlets of the political ‘enemy,’ ‘the elites,’ the government, official sites, etc. will be generally discredited. Echo chambers and conspiracy theories will select what news is considered relevant. This can systematically insulate news consumers from correcting false beliefs. For empirical studies on the dynamics of these kinds of social situations, see Jamieson and Cappella (2008), Uscinski and Parent (2014), and Benkler et al. (2018). For an epistemological assessment, see Cassam (2019) and Lynch (2019). Lackey (Chapter 9 in this volume) thinks that, in principle, there is nothing bad about echo chambers.

Epistemic ideologies offer us norms that regulate which pieces of evidence should be taken seriously and which ones can be properly ignored. For example, the journalistic balance norm of reporting requires that every testifier should receive equal attention. A norm of intellectual tolerance might point in the same direction. The norm of accepting only what strikes one as plausible privileges one’s own critical thinking over expert authority. Epistemic norms like these ones are the product of history. They regulate our selection of news. Some of them may select for fake news and are, for this reason, objectionable (see Chapter 5 by Ferrari & Moruzzi and Chapter 6 by Grundmann, both in this volume).

### 3.2.3 What Therapies Are Available as an Antidote to Fake News?

Suppose the news is massively polluted with fake news. What can be done to prevent the individual consumer from relying on it when they form their judgment? Three measures suggest themselves. First, the individual news consumer can be trained in distinguishing news from fake news. This training will relate to critical thinking, media literacy, or reflection on biases. In the extreme, the consumer might come to the rational conclusion that they should ignore the media in general (see Chapter 13 by Bernecker in this volume). Second, the individual news producer or distributor can be taught to exhibit more responsibility when disseminating news (see Chapter 8 by Priest and Chapter 11 by Wright both in this volume). News producers can try to communicate unambiguously by making clear whether a message is meant as an assertion or as an item of interest. Modesty in tone and language can help to prevent the audience from polarizing even further. News distributors can self-impose the rule that they do not share or re-tweet a story unless they have finished reading it. Third, there can be structural regulations by the government or self-regulations by platforms. These regulations may concern the algorithms of platforms (e.g., adjusting search engines to offer only quality information), censorship with respect to news content (e.g., deleting specific propagandistic messages), or revising journalistic practices and communication settings. All these measures can be implemented individually or in combination. Which of these measures is most effective and which of them is politically and morally acceptable is an open question.

#### 4. Contributions to the Volume

The volume is arranged into three parts. The chapters in Part I are concerned with the meaning of ‘fake news’ and related notions such as ‘conspiracy theory’ as well as with the novelty of the phenomenon of fake news. Part II discusses various practices that promote or generate fake news or are purported to do so. Part III explores potential therapies for fake news.

In their chapter, “Speaking of Fake News: Definitions and Dimensions,” Romy Jaster and David Lanius argue that fake news is news that lacks truth and truthfulness. They propose seven dimensions that are part of the fake news phenomenon, and explore how competing accounts of fake news accommodate these seven dimensions. They argue that their account best captures the concept fake news, while also precisifying it in an illuminating way. It is important to capture the various dimensions of fake news, they argue, in order to engage in the ameliorative project of avoiding the epistemic problem of fake news. The problem of fake news involves ignorance and uninformed decision-making, distrust of information further undermining access to information, and inability to deliberate within democracy. The correctives they propose are structural (rather than individual). To counteract fake news, we need boundary work and paradigm repair. Boundary work establishes clear norms of what is and isn’t news. Paradigm repair works to build strong internal norms for vetting and promoting news.

In contrast, in “Good News, Bad News, Fake News,” Duncan Pritchard argues that, just as a decoy duck is not a duck, fake news is not news. More than just lacking truth and truthfulness, Pritchard maintains that fake news doesn’t meet the requirement of being aimed at conveying accurate information of the relevant kind. The ‘fake’ tag indicates that it is masquerading as real news in order to spread misinformation. Since fake news is not genuine news, it is insufficient to distinguish it from a good source of news merely by tracing its poor epistemic pedigree. This point is especially significant for curtailing fake news; since fake news is not news, Pritchard suggests, undermining fake news is not on par with undermining the free press. On a structural level, he advocates for democratic institutions to monitor and flag fake news. On an individual level, Pritchard agrees with other contributors, Sarah Wright and Maura Priest, that an important component to resisting fake news is to cultivate intellectual virtues. He argues that virtues such as intellectual humility, intellectual conscientiousness, and honesty help individuals to spot fake news.

In his chapter, “The Fake News about Fake News,” David Coady argues that fake news is nothing more than a term used by the politically powerful to quell democratic dissent. In contrast with Pritchard’s contention that fake news is not news, Coady argues that fake news is nothing more than a label used to discredit democratic disagreement. On Coady’s view, then, the only thing that distinguishes fake news from news is who is inconvenienced by it. False reports exist, but they

exist in respectable news outlets as well as non-traditional ones. The only problem with fake news, Coady concludes, is that we keep talking about fake news. This just plays into the power games that powerful people deploy in order to maintain their power.

Shifting from fake news to conspiracy theories, in her chapter, “Conspiracy Theories and Evidential Self-Insulation,” Giulia Napolitano presents a novel view of conspiracy theories. Contrary to contemporary treatment of conspiracy theories as theories *about* conspiracies, Napolitano contends that conspiracy theories are the content of a particular kind of belief (a conspiracy belief) that is resistant to counterevidence by being *self-isolating*. A belief is self-isolating if the believer takes the conspiracy itself to neutralize the relevant counterevidence that the believer might normally encounter. Using a Bayesian framework, Napolitano characterizes the self-isolating character of conspiracy beliefs such that the agent’s credence in the conspiracy theory,  $C$ , is  $P(C|E) = P(C)$ , for any evidence  $E$  that the agent might normally encounter. This could happen either because one is certain in the conspiracy theory,  $P(C = 1)$ , or because the evidence is disregarded as irrelevant. Napolitano argues that neither can be a rational response. Thus, conspiracy theories are irrational by their nature.

In Part II, our authors explore the mechanics of fake news.

In their chapter, “Enquiry and Normative Deviance: The Role of Fake News in Science Denialism,” Filippo Ferrari and Sebastiano Moruzzi take up the topic of science denialism. Science denialism, they point out, is not generally a rationally unintelligible practice. Instead, they argue, it embodies an aberrant form of enquiry they term *post-enquiry*. In their chapter, they develop a model that captures the role fake news plays in bolstering the normative deviance of post-enquiry practices. Their model captures the ways in which science denialism mimics enquiry by making use of rational processes, as well as the way science denialism is aberrant. Post-enquiry science denialism makes use of epistemic filters and fake news to discredit institutional sources of evidence, and also to amplify pseudo-scientific explanations.

In most of the chapters, our authors explore the way pathologies creep into our epistemic practices. But in his chapter, “Facing Epistemic Authorities: Where Democratic Ideals and Critical Thinking Mislead Cognition,” Thomas Grundmann argues that even respectable enlightenment principles can lead us disastrously astray. Grundmann argues that two enlightenment ideals—(i) to use our critical thinking without restriction and (ii) to respect the rational judgment of any rational agent as epistemically rational—ought to be qualified when we have access to expert testimony. Experts, Grundmann argues, are far more likely to be right than we are in their conclusions about their domain of expertise. One overlooked source of our post-truth malaise, Grundmann argues, is that enlightenment ideals have given us bad epistemic standards that incubate conspiracy theories. Unlike other authors of this volume who argue for the cultivation of

individual abilities to identify fake news, Grundmann contends we should defer to experts. While individual intellectual abilities may have other benefits, in the case of expert testimony, we should view expert testimony as *preempting* our own judgments.

Catarina Dutilh Novaes & Jeroen de Ridder consider, “Is Fake News Old News?” They explore whether the phenomenon of fake news is actually anything new. They argue that it is largely continuous with other forms of misinformation and propaganda, though its distribution and mechanism have shifted in the Internet age. They consider three models of manipulation of information that have been used historically, and examine how these models impact epistemic autonomy. In Model A, Pleasing and Seducing the Audience, epistemic autonomy is largely intact, since other voices aren’t silenced. In Model B, Propaganda and Censorship, epistemic autonomy is undermined by the elimination of voices. In Model C, Disinformation by Epistemic Pollution, viewers are manipulated into thinking they are epistemically autonomous, though really they are manipulated. Dutilh Novaes and de Ridder examine the use of these models throughout history and conclude current fake news campaigns are not novel in their aim or tactics. It’s unclear that there is more fake news, or that our news landscape is more egalitarian. Propaganda has always appealed to baser instincts, but now a considerable amount of effort goes into gaming algorithms.

While many authors focus on the way that individual vices lead to being susceptible to fake news, and individual virtues can mitigate the harm, in her chapter, “How Vice Can Motivate Distrust in the Elites and Trust in Fake News,” Maura Priest explores ways in which experts have hampered their own efforts to convey knowledge. Priest identifies two vices of experts, *epistemic insensitivity* and *epistemic obstruction*. *Epistemic insensitivity* involves a failure to recognize the way various environmental features impact the uptake of knowledge within a context. *Epistemic obstruction* involves presenting material in a way that is not digestible by a non-expert audience. Both of these vices, Priest argues, can feed distrust of experts.

Echo chambers are often listed as a mechanism for our collective epistemic woes. But in her chapter, “Echo Chambers, Fake News, and Social Epistemology,” Jennifer Lackey contends that the problem is not echo chambers; the problem is *unreliable* echo chambers. In order to diagnose what has gone wrong, we should not look to some structural feature, such as the number of sources or diversity of opinions. Instead, the problem is about the content, about whether our information environment is reliable. She argues that our epistemic plight calls for non-ideal social epistemology. Theories of testimony, for example, are ill-equipped to handle an epistemic environment of retweeting bots. While Lackey does not go so far as to endorse Bernecker’s notion of news abstinence, she does suggest that we are much better off with a single reliable news source than we are with a range of conflicting news reports, some of which are unreliable.

In their chapter, Emmanuel J. Genod and Erik J. Olsson examine the way algorithms can obscure scientific data, propagating discredited science, even when official channels have refuted it. In “The Dissemination of Fake Science: On the Ranking of Retracted Articles in Google,” they offer a case study of how the search engines Google and Google Scholar display discredited scientific data. They show that in a range of cases, search engines continue to show the retracted article in a more prominent place than its retraction. They hypothesize that this is a result of Google’s popularity driven search algorithms, in conjunction with what they term the Law of Retraction. The Law of Retraction holds that retractions are rarely as noteworthy and shared as the original study, so in a popularity ranked algorithm, retractions will rank *below* the original article. As a result, Google disseminates fake and misleading science through their ranking of search results.

Finally, Part III explores ways in which we might ameliorate our epistemic plight.

In her chapter, “The Virtue of Epistemic Trustworthiness and Re-Posting on Social Media,” Sarah Wright argues that we are individually responsible for the epistemic risks we incur when we promote information on social media. She explores how the epistemically trustworthy person ought to behave on social media. The epistemically trustworthy person ought to assess the risks of potentially passing on false information, since it is difficult to effectively retract false information. Also, the epistemically trustworthy person recognizes the ways others depend on us to share important information. The virtuous person will balance risk and dependence through vigilantly examining potential posts, and contextualizing information that may be misleading.

Like Wright, Sanford Goldberg focuses on the way individuals within an information system are crucial to the epistemic health of the whole. In his chapter, “Fake News and Epistemic Rot; or, Why We Are All in this Together,” Goldberg focuses on the way in which individual epistemic practices affect more than just speaker and hearer. In an information network, there are many subtle and important ways that members affect the epistemic status of a particular news report. In addition, Goldberg suggests, our ambient news environment affects which news reports we encounter. Others play an important role in what he calls *background gardening*. Much like gardeners in a public garden who pull out weeds and nurture healthy plants, others in our social networks will do epistemic work to squelch bad reports and nurture good ones. And we owe it to each other to engage in mutually beneficial background gardening. Continuing with the plant metaphor, Goldberg also introduces the notion of *epistemic rot*, the decay of informational systems through deviant mechanisms. Fake news constitutes one of these mechanisms, and it is incumbent on each of us to tend our garden carefully to contain the pathogen.

One potential way to contain epistemic rot might be to practice news abstinence. In his chapter, “An Epistemic Defense of News Abstinence,” Sven Bernecker argues that we may be propositionally justified in temporarily ignoring news,

either in a domain or from a source. Bernecker lays out two conditions that must be met: (i) we are in a fake news environment or are justified in believing that we are, and (ii) it is cognitively difficult or time consuming to discriminate genuine from fake news or to obtain genuine news. In this case, the consumption of the news leads us to acquire false or irrelevant beliefs and prevents us from acquiring true and relevant ones. Motivated ignorance of the news is epistemically justified when epistemic value can be secured better elsewhere. Bernecker's position contrasts strongly with both Pritchard (who argues that information is too important for abstinence to be the right move, even if the environment is risky) and Wright (who argues that others depend on us, and we have a role to play in our information network).

In his chapter, "Fake News, False Beliefs, and the Fallible Art of Knowledge Maintenance," Axel Gelfert argues that fake news presents a novel kind of social-epistemic dysfunction, arising from *systemic* distortions of established processes of creating and disseminating news-like content. The result is not just that there are false reports in our information environment, but also our environment lacks coverage reliability. What is needed, Gelfert contends, is epistemic routines that vindicate trust in reliable sources. While Gelfert does not use the language of gardening or virtue, his suggestions are consonant with those of Goldberg and Wright: we need epistemic practices that operate on a habitual level that promote a healthy information network.

Finally, in "Trust No One? The (Social) Epistemological Consequences of Belief in Conspiracy Theories," Michael Baurmann and Daniel Cohnitz argue that conspiracy theories threaten to undermine democratic institutions. They argue that democratic societies require a complex and open information-sharing system. Belief in conspiracy theories undermines social trust in information networks, and is easily exploited by actors seeking to undermine open societies. Baurmann and Cohnitz consider whether debunking or infiltrating conspiracy networks would be successful at undermining the potency of conspiracy theories. They argue that these strategies are *unlikely* to be successful, since conspiracy networks have very little trust towards outsiders, and undermining a conspiracy from the inside doesn't instill trust in the diminished institution (especially if the infiltration is discovered!). Instead, they suggest that a solution is to cultivate personal relationships of trust with members of conspiracy networks. Their proposal goes far beyond those made by others in Part III. Where Wright, Goldberg, and Gelfert focus on our duties to others within our information network, they are silent on whether we ought to excise the conspiratorially minded from our community. Baurmann and Cohnitz argue that we ought to positively cultivate relationships with conspiracy thinkers in order to maintain relationships of trust.<sup>1</sup>

<sup>1</sup> We thank Peter Momtchiloff for insightful comments on a previous draft of this chapter.

## References

Allcott, H. and Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives* 31: 211–36.

Benkler, Y., Faris, R., and Roberts, H. (2018). *Network Propaganda. Manipulation, Disinformation, and Radicalization in American Politics*. Oxford: Oxford University Press.

Brown, E. (2019). ‘Fake News’ and Conceptual Ethics. *Journal of Ethics and Social Philosophy* 16: 144–54.

Burge, T. (1993). Content Preservation. *Philosophical Review* 102: 457–88.

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

Coady, C.A.J. (1992). *Testimony: A Philosophical Study*. New York: Oxford University Press.

Curtois, C., Slechten, L., and Coenen, L. (2018). Challenging Google Search Filter Bubbles in Social and Political Information: Disconfirming Evidence from a Digital Methods Case Study. *Telematics and Information* 35: 2006–15.

Fallis, D. (2019). The Epistemic Threat of Deepfakes. *Philosophy & Technology*. <https://doi.org/10.1007/s13347-020-00419-2>.

Fallis, D. and Mathiesen, K. (2019). Fake News is Counterfeit News. *Inquiry*. <https://www.tandfonline.com/doi/full/10.1080/0020174X.2019.1688179>.

Fricker, E. (1995). Telling and Trusting: Reductionism and Anti-Reductionism in the Epistemology of Testimony. *Mind* 101: 393–411.

Gelfert, A. (2018). Fake News: A Definition. *Informal Logic* 38: 84–117.

Goidel, K., Gaddie, K., and Ehrl, M. (2017). Watching the News and Support for Democracy: Why Media Systems Matter. *Social Science Quarterly* 98: 836–55.

Goldberg, S.C. (2010). *Relying on Others: An Essay in Epistemology*. Oxford: Oxford University Press.

Goldberg, S.C. (2013). Anonymous Assertions. *Episteme* 10: 135–51.

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

Grundmann, T. (2020). Fake News: The Case for a Purely Consumer-Oriented Explication. *Inquiry*. <https://www.tandfonline.com/doi/full/10.1080/0020174X.2020.1813195>.

Habgood-Coote, J. (2019). Stop Talking About Fake News! *Inquiry* 62: 1033–65.

Haim, M., Arendt, F., and Scherr, S. (2017). Abyss or Shelter? On the Relevance of Web Search Engines’ Search Results When People Google for Suicide. *Health Communication* 32: 253–8.

Hannack, A., Sapiezynski, P., Kakhki, A.M., Krishnamurthy, B., Lazer, D., Mislove, A., and Wilson, C. (2013). Measuring Personalization of Web Search. *Proceedings of the 22nd International Conference on World Wide Web*, pp. 527–38. <https://dl.acm.org/doi/10.1145/2488388.2488435>.

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

Jaster, R. and Lanius, D. (2018). What is Fake News? *Versus* 2: 207–27.

Lazer, D.M., Baum, M.A., Benkler, Y., Berinsky, A.J., Greenhill, K.M., Menczer, F., Metzger, M.J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C.R., Thorson, E.A., Watts, D.J., and Zittrain, J.L. (2018). The Science of Fake News: Addressing Fake News Requires a Multidisciplinary Effort. *Science* 359: 1094–6.

Levy, N. (2017). The Bad News About Fake News. *Social Epistemology Review and Reply Collective* 6: 20–36.

Lynch, M.P. (2019). *Know-It-All Society: Truth and Arrogance in Political Culture*. New York and London: Liveright.

McIntyre, L. (2018). *Post-Truth*. Cambridge, MA: MIT Press.

Mukerji, N. (2018). What is Fake News? *Ergo* 5: 923–46.

Oremus, W. (2016). Stop Calling Everything Fake News. *Slate Magazine*, Dec. 6, 2016. <https://slate.com/technology/2016/12/stop-calling-everything-fake-news.html>.

Pariser, E. (2011). *The Filter Bubble: What the Internet is Hiding from You*. London & New York: Penguin.

Pepp, J., Michaelson, E., and Sterken, R.K. (2019a). What's New About Fake News? *Journal of Ethics and Social Philosophy* 16: 67–94.

Pepp, J., Michaelson, E., and Sterken, R.K. (2019b). Why We Should Keep Talking about Fake News. *Inquiry*. <https://www.tandfonline.com/doi/full/10.1080/0020174X.2019.1685231>.

Rawlinson, F. (2020). *How Press Propaganda Paved the Way to Brexit*. London: Palgrave Macmillan.

Rini, R. (2017). Fake News and Partisan Epistemology. *Kennedy Institute of Ethics Journal* 27: 43–64.

Rini, R. (2020). Deepfakes and the Epistemic Backstop. *Philosopher's Imprint* 20. <http://hdl.handle.net/2027/spo.3521354.0020.024>.

Simons, G. and Strovsky, D. (2019). The Interaction of Journalism and Public Relations in Russia: A Self-Perception. *Global Media and Communication* 15: 3–25.

Sullivan, M. (2017). It's Time to Retire the Tainted Term 'Fake News.' *Washington Post*: Jan. 8, 2017. [https://www.washingtonpost.com/lifestyle/style/its-time-to-retire-the-tainted-term-fake-news/2017/01/06/a5a7516c-d375-11e6-945a-76f69a399dd5\\_story.html](https://www.washingtonpost.com/lifestyle/style/its-time-to-retire-the-tainted-term-fake-news/2017/01/06/a5a7516c-d375-11e6-945a-76f69a399dd5_story.html).

Sunstein, C. (2017). *#Republic: Divided Democracy in the Age of Social Media*. Princeton, NJ and Oxford: Princeton University Press.

Talisse, R.B. (2018). There's No Such Thing as Fake News (and That's Bad News). *3:AM Magazine*: June 9, 2018. <https://www.3ammagazine.com/3am/theres-no-such-thing-as-fake-news-and-thats-bad-news/>

Uscinski, J. and Parent, J. (2014). *American Conspiracy Theories*. Oxford: Oxford University Press.

Vosoughi, S., Roy, D., and Aral, S. (2018). The Spread of True and False News Online. *Science* 359(6380): 1146–51.

Welbers, K., Atteveldt, W. van, Kleinnijenhuis, J., and Ruigrok, N. (2018). A Gatekeeper among Gatekeepers: News Agency Influence in Print and Online Newspapers in the Netherlands. *Journalism Studies* 19: 315–33.

# 1

## Speaking of Fake News

### Definitions and Dimensions

*Romy Jaster and David Lanius*

### 1. Introduction

There has been much discussion about what fake news is and what to do about it—among politicians and journalists, between academics and in the wider public. But it is not clear what fake news is and what people mean when they speak of “fake news.” So much so, that doubts have been raised that the term is useful at all (e.g. House of Commons: DCMS 2019; Habgood-Coote 2018; Wardle 2017; Wardle & Derakhshan 2017; Zuckerman 2017).

This chapter shows why a definition is urgently needed and what a suitable definition of “fake news” might look like. We take up Brown’s (2019) challenge to keep using the term “fake news” with care, since doing so enables us to raise important philosophical and societal questions.<sup>1</sup>

We begin by introducing our definition of “fake news” (§2) and employ it to set fake news apart from related phenomena that are often conflated with it (§3). We then extract seven potential dimensions of the concept of fake news from the literature (§4) and compare the most representative definitions that have been proposed so far along those dimensions (§5). In particular, we discuss the definitions by Rini, Gelfert, Dentith, Mukerji, and Zimmermann and Kohring, show up their merits and debits, and put them in relation to ours.<sup>2</sup>

So, although we take our definition as the starting point and argue for it on the sidelines, our primary aims are (i) to enable a systematic evaluation of prevalent definitions with respect to their extensional scope, practical utility, and conceptual transparency, (ii) to demonstrate that there is more widespread agreement than one would think at the outset, and (iii) to show (in §6) that defining “fake news” is

<sup>1</sup> As Brown (2019, 152) shows by refuting Habgood-Coote’s (2018) arguments, “fake news” sometimes functions as a slur, but it need not do so if used with care. When it is, it allows us to raise philosophical questions that could not be discussed if the concept was abandoned.” We agree with this assessment.

<sup>2</sup> For systematic assessments of definitions in the literature, see also Egelhofer & Lecheler (2019), Michaelson et al. (2019) and Fallis & Mathiesen (2019).

not only far from futile, but of vital importance to confront the epistemic threats posed by fake news.

## 2. Defining Fake News

As we have argued elsewhere (Jaster & Lanius 2018), fake news is news that lacks truth and truthfulness. It lacks truth in the sense that it is either literally false or communicates something false. It lacks truthfulness in the sense that it is propagated with the intention to deceive or without concern for the truth. Both conditions will be spelled out in detail below.

In our account, we are employing the *Oxford English Dictionary's* (2018) minimalist notion of “news” as “newly received or noteworthy information, especially about recent events” and, in a more specific usage, as a “broadcast or published report of news.” Accordingly, we use “news” to refer to any report of typically recent events that is broadcast by media or individuals to address a public.<sup>3</sup>

News is nowadays distributed via more channels and by more agents than in the past (Kovach & Rosenstiel 2014). Politicians distribute their content on Twitter or Facebook without professional journalists as intermediaries. Of course, most tweets and Facebook posts do not qualify as news. But reports about recent events that are broadcast to a public are news, no matter on which channel and by whom the content is distributed.<sup>4</sup>

We take reports to be assertions, or *truth warranting* utterances, i.e., utterances that go along with the guarantee of truth on the speaker's part (Carson 2006, 2010).<sup>5</sup> The guarantee of truth need not be given intentionally or knowingly. Instead, it is part of the illocutionary speech act itself. Speakers may know that their utterance comes with a guarantee of truth—but at the same time they may not feel committed to believing their utterances to be true.

<sup>3</sup> Our definition can be modified to account also for a more normative understanding of news. We then define fake news as assertions about recent events that are broadcast by media or individuals to address a public and that lack both truth and truthfulness. Due to their lack of truthfulness, such assertions do not qualify as news (in the normative sense). According to this (slight) modification of our definition, fake news is *not* news (in the normative sense).

<sup>4</sup> Not everyone follows us here. See, for instance, Fallis & Mathiesen (2019) for a view that is more restrictive in this respect.

<sup>5</sup> Truth warrant may be more complicated than this. We usually do not guarantee the truth of what we say, but seem to guarantee only that we believe to know what we say. Moreover, truth warrant is presumably a matter of degrees with full-fledged assertions only on one end of the scale. But these intricacies don't matter for the purposes of this chapter.

## 2.1 Lacking Truth

When things go well, a news report is true—both in its *literal content* (“what is said”) and in its *communicative content* (“what it pragmatically conveys”).<sup>6</sup> In the case of fake news, the literal or communicative content is false. Thus, fake news lacks truth on one of the two dimensions.

An example for fake news with false literal content is the Pizzagate story that circulated in Russian media outlets in 2016. After WikiLeaks published email correspondence from Hillary Clinton’s mail server, conspiracy theorists began to suspect that the emails contained encoded messages having to do with a child pornography ring run by Hillary Clinton and other high profile Democrats in the basement of a pizza joint. Russian websites and smaller news outlets quickly jumped on the story and distributed it widely. After an armed civilian showed up in the pizza joint to self-investigate the issue, the case was reported and discussed in most established media. In reality, there was no child pornography ring in the pizza joint. There was not even a basement. The Pizzagate fake news story is false.<sup>7</sup>

Other cases are more subtle. After turmoil in the German city of Dortmund on New Year’s eve, the American online medium Breitbart reported in January 2017 that “a mob of more than 1,000 men [...] set fire to a historic church” (Hale 2017). What is *said* here is not false. There was a large group of people. Fireworks were launched. There was a fire. Yet, the report is highly misleading because it *conveys* falsities.<sup>8</sup> The report suggests that the fire was started wantonly by Muslims, that the church itself was burning, and that the fire was of considerable (i.e., newsworthy) size. All of that is false. In reality, the fireworks accidentally set a catching net on fire that was attached to scaffolding around parts of the church. According to the fire department, the fire was small and could be contained immediately.<sup>9</sup> This shows that fake news reports need not be literally false. Even

<sup>6</sup> We are using “what is said” in Saul’s (2012) sense. Cf. also Recanati (2004) on the notion of literal content. The claims in this chapter about literal content should be largely uncontroversial. Communicative content, as we understand it, is what a competent speaker, knowing the context and all relevant background information, would assume an utterance to mean (instead of or additional to what it literally says). Communicative content is thus logically independent of any individual speaker’s or audience’s actual intentions. However, since written utterances typically have multiple contexts of interpretation, they can have multiple communicative contents.

<sup>7</sup> Of course, the Pizzagate story could theoretically have turned out to be true. What this entails about the status of the reports will be discussed later on.

<sup>8</sup> Note that we are using the term “misleading” in a technical sense such that an utterance is “misleading” if and only if its communicative content is false. We do not use the term in the sense in which “misleading” necessarily entails the utterance’s actual or potential effect on the audience, as, for instance, Fallis (2015).

<sup>9</sup> See *The Guardian* (2017).

when what is *said* is true, a piece of news is lacking truth if it pragmatically *conveys* something false.

## 2.2 Lacking Truthfulness

Fake news also lacks truthfulness: It is distributed either with an intention to deceive or with no concern for the truth. By all accounts, the Russian news about Pizzagate and the Breitbart story are characterized by an intention to deceive: The news reports were presumably propagated with the goal of inciting false beliefs about the reported events in the respective audiences.

Assertions put forward without *any* concern for truth are what Frankfurt has famously called “bullshit.” In Frankfurt’s view, the bullshitter “does not care whether the things he says describe reality correctly. He just picks them out, or makes them up, to suit his purpose” (Frankfurt 2005, 55). To put it more precisely,

[a person] A is bullshitting relative to a QUD [question under discussion] q if and only if A contributes p as an answer to q and A is not concerned that p be an answer to q that her evidence suggests is true or that p be an answer to q that her evidence suggests is false. (Fallis & Stokke 2017, 295)

We use “bullshit” roughly in Fallis and Stokke’s sense. To bullshit is to contribute an utterance as an answer to a question under discussion without concern that what one says *or communicates* is based on evidence.

The most prominent examples of bullshit fake news are the fabricated news reports circulated by Macedonian teenagers before the US elections in 2016. We know from interviews that the teenagers did not have any interest in the truth or falsity of their reports (Silverman & Alexander 2016). Their goal was not to deceive the audience about the reported content, but to fabricate news that generated as many clicks and thus as much money as possible.<sup>10</sup>

While the teenagers’ news were usually flat-out false, bullshit fake news may as well be misleading. In 2018, Donald Trump tweeted a picture of a wall under construction and added “Great briefing this afternoon on the start of our Southern Border WALL!” The tweet could be literally true: Trump may have actually had a great meeting. But it is misleading: Together with the picture, it insinuates that construction on the southern border wall had already begun. This is not the case; in fact, the picture shows a completely different wall.<sup>11</sup>

<sup>10</sup> They may well want to deceive their audience about their attitude towards the truth. In fact, deceiving their audience in this way seems to be a necessary means for getting clicks.

<sup>11</sup> Of course, we cannot know for sure whether Trump is actually bullshitting in this tweet. But, arguably, he simply tweeted a picture of some wall being constructed without caring about the truth or falsity of the insinuation that goes along with it.

Here is an overview of our definition of fake news:

FAKE NEWS		LACK OF TRUTH	
LACK OF TRUTHFULNESS	Intention to deceive	False utterance	Misleading utterance
	Bullshit (disregard for truth)	Reports on the Pizzagate conspiracy	Breitbart's report on the burning church
		The teenagers' reports on the US elections 2016	Trump's tweet on the "Southern Border WALL"

Importantly, there is a difference between *originating* or *knowingly distributing* fake news on the one hand, and *unknowingly distributing* existing fake news on the other. A report being fake news is tied to (1) features of its content (lack of truth) and (2) the sender's mindset (lack of truthfulness). Like a lie, fake news can spread, even if no one but the originator is untruthful. Let's call someone who originates or knowingly distributes fake news a "fake news distributor."

### 3. Related Phenomena

With a definition of "fake news" at hand, fake news can be systematically differentiated from a variety of related phenomena which are regularly conflated with fake news.<sup>12</sup> Figure 1.1 situates fake news in a broader landscape of related phenomena.

First, fake news is a subspecies of truth-warranting utterances. This sets fake news apart from utterances that are not truth-warranting, among them questions, typical instances of satire or parody, and jokes. Secondly, fake news is a subspecies of news. This sets it apart from claims about history ("The state of Israel was founded in 1948"), scientific facts ("Water is H<sub>2</sub>O"), and assertions uttered in private. Thirdly, fake news lacks truth, which sets it apart from most candid news reports. Fourthly, fake news lacks truthfulness, which sets it apart from typical instances of inadvertently erroneous news reports as well as many cases of conspiracy theories and propaganda. Let's look in more detail at the relation of fake news to satire, conspiracy theories, and propaganda.

Even though the term "fake news" was initially used for *satirical* news shows (Amarasingam & McChesney 2011), these shows do not typically spread fake news. First, speech acts of satire need not be set in a truth warranting context. Usually, no assertion is being made, even if parts of the audience may mistakenly believe so. Secondly, satire only works against the background of a shared

<sup>12</sup> Others have drawn similar distinctions based on their accounts of fake news. See, for instance, Dentith (2017), Gelfert (2018), or Mukerji (2018).

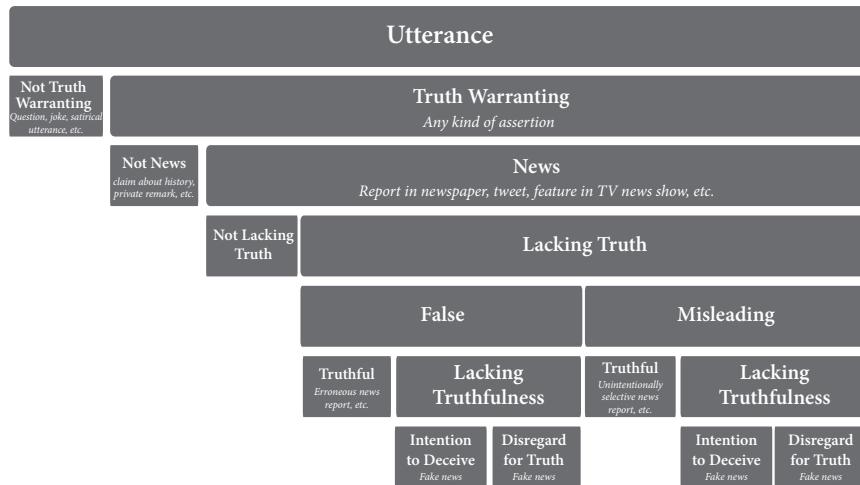


Figure 1.1 The fake news landscape

understanding that truth is not aimed at. The whole point of satirical humor would be lost if the target audience took the satirical content at face value.<sup>13</sup> Satirical fakes of news are not the same thing as fake news.

Likewise for fake news and *propaganda*. Given that propaganda is “[c]ommunication designed to manipulate a target population by affecting beliefs, attitudes, or preferences in order to obtain behavior compliant with the political goals of the propagandist” (Benkler et al. 2018, 29), fake news and propaganda are not the same.<sup>14</sup> The Macedonian teenagers had no political agenda. Thus, their fake news stories are not propaganda. Reversely, propaganda need not be false or misleading. Sometimes, it states truths, sometimes it does not even make statements at all, as in inciting or glorifying pictures.<sup>15</sup> Yet, fake news and propaganda may be co-instantiated. In an attempt to conceal the German Democratic Republic’s inability to prevent the massive potato beetle plague in 1950, national

<sup>13</sup> That is not to say that satire does not work with deception. There is satire that makes a point of deceiving people and revealing the deception afterwards. Note, however, that the humorous element comes with the revelation, not with the deception itself.

<sup>14</sup> Benkler et al. (2018) follow Jowett & O’Donnell’s (2006) influential work on propaganda.

<sup>15</sup> Note that pictures can be used to do mischief with the truth in very analogous ways to fake news: a photomontage depicts an event that has not happened and thus corresponds to a literally false statement. The pictorial analogy to literally true, but misleading verbal statements is a cropped picture which shows an event that has happened, but conveys something false by cutting out relevant aspects of the scenery. Such pictures can be (and often are) used in news reports with an intention to deceive or with disregard for the truth—if they are, the corresponding news reports are fake news. Presumably even more often, the pictures themselves are not even tampered with, but are simply used in a misleading context (by falsely claiming or insinuating that they depict certain events in a certain place at a certain time). This, again, can amount to fake news if the corresponding news report (in combination with the picture) is deliberately or without regard for the truth propagated such that it conveys something false.

newspapers announced that the plague had been caused by the US government by dropping potato beetles over GDR territory. These news reports were both fake news *and* propaganda.

The same holds for fake news and *conspiracy theories*. Roughly, a conspiracy theory is an explanation of some event, according to which certain people or groups share a (typically hidden) interest in the occurrence of an event and conspired to bring it about (Popper 1992, 19).<sup>16</sup> Many fake news reports do not deal with that sort of explanation of events (consider the Breitbart story about the burning church). Reversely, not every conspiracy theory is fake news. Some may turn out to be true (like Watergate), and even when false, their distributors need not be untruthful. Pizzagate is a fake news story because Russian and other outlets distributed it untruthfully, but the conspiracy theory underlying it had been circulating for much longer. A story may start out as a mere conspiracy theory and become the subject of fake news subsequently.<sup>17</sup>

#### 4. The Dimensions of Fake News

In this section, we differentiate seven dimensions that potentially form part of the concept of fake news and are derived from the debate so far.<sup>18</sup>

The *Truth Dimension*: On most accounts, fake news is taken to lack truth in some way. Apart from Mukerji (2018), who analyzes fake news as a form of bullshit, virtually every account in the debate so far commits to the idea that fake news has *something* to do with falsity. Some scholars think that fake news is necessarily false (Allcott & Gentzkow 2017, 213; Horne & Adali 2017, 2; Klein & Wueller 2017, 6; Rini 2017, E-45; Levy 2017, 20; Mustafaraj & Metaxas 2017, 2), while others allow for fake news to be merely misleading (Dentith 2017, 66; Tandoc et al. 2018, 147; Shin et al. 2018; Gelfert 2018, 102).<sup>19</sup>

The *Deception Dimension*: Most definitions of fake news contain the requirement that false content is propagated intentionally. Often, it remains unclear what the intention aims at—whether fake news is intended to deceive about facts in the world or merely about its distributor's state of mind.<sup>20</sup> More often than not,

<sup>16</sup> There are other definitions of conspiracy theories (e.g., Keeley 1999; Mandik 2007; Dentith 2014). Most are in line with the way we set fake news and conspiracy theories apart.

<sup>17</sup> The difference between fake news and fake science is more straightforward. Only news (in the minimal sense) can be fake news, but science is not news. Thus, there are no things that are both fake news and fake science. However, there can be fake news *about* science as well as fake science.

<sup>18</sup> Concepts usually have a number of dimensions, which can be seen as more or less essential (relative to their other dimensions). As a result, the terms expressing such concepts are, as defined by Lanius (2019, 34–7), multi-dimensionally polysemous.

<sup>19</sup> A few scholars also focus on the even narrower feature that fake news is entirely fabricated, such as, for example, Nelson & Taneja (2018).

<sup>20</sup> Of course, any attempt to deceive about facts in the world goes along with the attempt to deceive about one's own state of mind.

however, the intention to deceive about the news's content is taken to be a central criterion of fake news (Horne & Adali 2017, 1; Klein & Wueller 2017, 6; Rini 2017, E-45; Dentith 2017, 66; Egelhofer & Lecheler 2019, 7; Tandoc et al. 2018, 147).

The *Bullshit Dimension*: The intention to deceive about content is not considered an indisputable criterion. Rini (2017, E-44–5), for example, acknowledges that “deception is not always the primary goal of fake news. Often the motives are financial rather than epistemic” (see also Gelfert 2018, 102). Many authors emphasize the political or financial motives of fake news producers (Allcott & Gentzkow 2017, 217; Rini 2017; Levy 2017, 20; Gelfert 2018; McNair 2017, 38; Nelson & Taneja 2018; Tandoc et al. 2018, 138). Someone may knowingly spread fake news without at the same time having the intention to deceive about its truth. What may thus be central to fake news is not an intention to deceive about what is reported, but an indifference to the truth; fake news may turn out to be bullshit.

The *Appearance Dimension*: Sometimes, it is held that fake news is intentionally propagated as “real” news (Levy 2017, 20; Shin et al. 2018; Rini 2017).<sup>21</sup> In fact, a widely undisputed and often explicitly mentioned feature of fake news is that journalistic formats are imitated (Horne & Adali 2017, 1; Rini 2017, E-45; Levy 2017, 20; Mustafaraj & Metaxas 2017, 2; Nelson & Taneja 2018; Tandoc et al. 2018, 138; Gelfert 2018, 103; Mukerji 2018, 929). Thus, some authors understand fake news to falsely pretend to be based on journalistic sources and adhere to journalistic standards (Egelhofer & Lecheler 2019, 10–11).

The *Effect Dimension*: When politics and countermeasures are concerned, the effects of fake news are usually at the center of attention. Some authors (most emphatically Grundmann 2020) stress that fake news actually deceives or is at least likely to deceive parts of the audience. Gelfert (2018, 108), for instance, takes fake news to be “objectively likely to mislead its target audience” and Allcott & Gentzkow (2017, 213) define fake news as “news articles that are intentionally and verifiably false, and could mislead readers.”

The *Virality Dimension*: The public and most academics agree that fake news has become a huge challenge to democratic decision and opinion formation processes. In large part, this is seen to be due to its virality (Allcott & Gentzkow 2017, 217; Shao et al. 2017; Shin et al. 2018; Tandoc et al. 2018, 138). Much like rumors, fake news spreads fast and sometimes uncontrollably. Rini (2017, E-45) thus includes virality as a necessary condition in the definition of fake news. If she is right, fake news necessarily is either intended to be widely propagated or actually virally spread.

The *Media Dimension*: There is broad consensus that the propagation of fake news is driven by digital forms of communication and particularly social media (Tandoc et al. 2018, 138–9). Some scholars take this to be essential to fake news

<sup>21</sup> This dimension is related to the *Deception* and the *Bullshit Dimension*. Make-believe suits both endeavors well.

and argue that fake news can only be propagated online (Bounegru et al. 2018, 8; Klein & Wueller 2017, 6; Mustafaraj & Metaxas 2017, 2). However, only very few authors assume that this is a strictly necessary criterion.

Here is an overview of the seven potential dimensions that emerge from the literature:

<i>Truth Dimension</i>	Fake news is false or misleading.
<i>Deception Dimension</i>	Fake news distributors intend to deceive.
<i>Bullshit Dimension</i>	Fake news distributors are indifferent to the truth.
<i>Appearance Dimension</i>	Fake news mimics “real” news.
<i>Effect Dimension</i>	Fake news entails the actual (or a certain likelihood of) deception on the part of the audience.
<i>Virality Dimension</i>	Fake news is (or is intended to be) widely propagated.
<i>Media Dimension</i>	Fake news is a phenomenon of the Internet or social media.

As can easily be seen, our definition contains the *Truth*, *Deception*, and *Bullshit Dimension*, as it defines fake news as news reports lacking truth in what is said or communicated and being propagated with either an intention to deceive or disregard for the truth.

## 5. The Definitions of Fake News

Let us now look at the following prominent definitions and see which dimensions they emphasize.<sup>22</sup> Rini (2017) has early in the debate proposed a definition of fake news as false stories with an intention to deceive. Gelfert (2018) defines fake news as misleading by design. Dentith (2017, 2018) offers a definition of fake news as misleading or false and intended to deceive. Mukerji (2018) defines it as bullshit asserted in the form of news. Zimmermann & Kohring (2018) give a definition of fake news from the perspective of communication science as recent disinformation. We will discuss them one by one.

We take all of the definitions to be *explications* of the concept of fake news, and thus engage with the project of “transforming a given more or less inexact concept into an exact one” (Carnap 1950, 3). The term “fake news” is used in a variety of ways in ordinary speech. Even upon reflection, there are unclear cases. To explicate the concept of fake news, in our view, we need to find a definition that

<sup>22</sup> For a helpful overview of a number of yet other definitions which are analyzed in terms of the *Truth*, *Deception*, and *Appearance Dimension*, see Egelhofer & Lecheler (2019, 3).

classifies all clear cases of fake news as fake news and all clear counter-instances as things other than fake news, providing guidance in unclear cases.

### 5.1 Rini's Definition: False Stories with Intention to Deceive

Rini (2017, E-45) defines a fake news report as

one that purports to describe events in the real world, typically by mimicking the conventions of traditional media reportage, yet is known by its creators to be significantly false, and is transmitted with the two goals of being widely re-transmitted and of deceiving at least some of its audience.

This definition contains most dimensions we distinguished. It emphasizes the *Truth* ("known [...] to be significantly false"), *Deception* ("deceiving at least some of its audience"), *Virality* and, presumably, the *Media Dimension* ("being widely re-transmitted"). It also seems to emphasize the *Appearance Dimension* (mimicking "real" news), but note that Rini wisely inserts a "typically"-clause here. Many fake news reports certainly mimic "real" news, but many also differ substantially from "real" news in content, form, and style (Horne & Adali 2017). Accordingly, Rini rightly abstains from including the *Appearance Dimension* as a necessary condition.

The differences to our definition lie elsewhere. First, Rini understands fake news to be known to be significantly false. But first, as argued, fake news need not be false, but may merely be misleading. Secondly, fake news need not be *known* to lack truth. The teenagers spreading fake news for profit may not have known about the falsity of many of their clickbaiting news reports—they just did not care. So while we agree that the *Truth Dimension* is crucial, we disagree about its scope and Rini's epistemic twist.

The *Virality Dimension* seems highly plausible at first. Most fake news stories are presumably spread with an intention to reach large numbers of people. But that is because fake news is news and, as such, broadcast to a public. Rini's emphasis on *re-transmission* seems to tacitly presuppose that fake news is a phenomenon of the Internet and social media (*Media Dimension*). Here, we disagree. A print newspaper may just as well spread fake news without the goal of wide re-transmission. To be fair, Rini (2017, E-45) states that "fake news can be spread other ways—email chains, posters on streetlamps, etc." But the paradigm of fake news she targets with the idea of re-transmission is clearly one that locates fake news in social media.<sup>23</sup>

<sup>23</sup> Rini's commitment to this understanding of her definition becomes evident in her discussion of "social media testimony" (2017, E-54).

Whether fake news is transmitted with the goal of “deceiving at least some of its audience” depends on how exactly we understand the condition. Rini (2017, E-45) is quite explicit in this respect. As she points out, not all fake news reports require an intention to deceive about their content. Some are propagated with commercial or personal motivations and put forward with a bullshit attitude. Yet, in her view, there is nevertheless deception involved. On Frankfurt’s original understanding, bullshitters intend to deceive their audiences as well. In contrast to the liar, however, their deception is not about their assertions’ content, but about themselves. In Frankfurt’s view, bullshit requires the (second-order) intention to deceive about the speaker’s indifference to the truth.

While this might not always be the case with respect to the entire audience, it seems true for at least parts of it. Trump may want to fool some of his followers about his intentions, and so may commercial fake news producers. Of course, their motivation to deceive about their indifference to the truth is purely instrumental. If their attitude were all too obvious, there might not be any (or too little) incentive for the audience to click on the fake news story.

Rini clearly understands the intention to deceive to extend to the distributor’s intention to deceive about her own indifference to the truth as well. If “the goal to deceive” is understood this broadly, this amounts to our condition of lack of truthfulness and combines the *Deception* and *Bullshit Dimension*. Note, however, that the bullshitter’s deceitfulness is among the most contested elements of Frankfurt’s definition of bullshit (Fallis & Stokke 2017). We therefore take it to be wise not to commit to it in a definition of fake news and instead spell out explicitly how the lack of truthfulness plays out.

Let’s recapitulate the similarities and differences between Rini’s and our definition. Taking the “typically”-clause seriously, we agree that fake news does not by its nature mimic “real” news (*Appearance Dimension*). We also agree that fake news distributors pursue the goal of deceiving at least some of their audience—provided that this covers both the *Deception* and *Bullshit Dimension*. We disagree that fake news is necessarily distributed with the goal of being widely re-transmitted (*Virality Dimension*) and the underlying idea that fake news is a phenomenon of the Internet (*Media Dimension*). We also disagree that fake news is necessarily (known to be) false.

## 5.2 Gelfert’s Definition: Presentation of Claims as News Misleading by Design

Gelfert (2018, 108) defines fake news as “the deliberate presentation of (typically) false or misleading claims as news, where the claims are misleading by design.” The definition emphasizes the *Truth* (“false or misleading”), *Deception* (“deliberate presentation”), and *Appearance Dimension* (“claims as news”). The *Effect*

*Dimension* enters the picture via the concept of misleadingness Gelfert employs. Fake news is misleading, according to Gelfert, in the sense that it “is likely to result in (and often does cause) false beliefs on the part of its target audience” (105). The *Virality Dimension* is not explicit in the definition itself, but Gelfert clearly thinks that fake news is distributed with “the goal of widespread circulation” (102).

It is unclear whether the *Bullshit Dimension* is instantiated. The core notion in Gelfert’s definition is “misleadingness by design.” Unlike journalistic errors, Gelfert writes, for a report to be fake news “it must be likely to mislead not only in a non-accidental way, but deliberately” (106). In fact, “the spread of false beliefs is not merely a side effect of fake news, but is a direct result of its function” (108) in that

fake news is designed to operate in a way that is unconstrained by the truth, either because it aims to instil falsehoods in its target audience (for example, in order to discredit a political opponent), or because the way it is deliberately operated is objectively likely to mislead its target audience [...]. (108)

It is not easy to pinpoint what exactly Gelfert has in mind when he talks about the way in which fake news is “designed to operate” or “deliberately operated.” The point might be that there is something in the process of spreading fake news that makes it (potentially) misleading. But the process of making up false claims to deceive the target audience into believing whatever is in line with one’s own political agenda is rather different from the process of gathering claims that sell. Moreover, the same processes may turn out to be (even objectively) far less likely to mislead if presented in different ways.

On a slightly different understanding, the idea would be that the way fake news itself operates—the way it diffuses into societies—makes it (potentially) misleading. However, it is unclear how this process of diffusion differs from the way other, reliable, news diffuses into societies. In an attempt to further unpack the idea of misleadingness by design, Gelfert points out that

purveyors of fake news have begun to employ strategies of bringing about belief and ensuring continued propagation of their stories [...], by manipulating their consumers’ preexisting cognitive biases and heuristics. (111)

We are not sure whether Gelfert considers this an essential feature of fake news or whether the exploitation of biases is merely supposed to explain what is new about fake news these days.<sup>24</sup> We agree with the latter idea, but are skeptical concerning

<sup>24</sup> An anonymous referee has suggested to us that, in Gelfert’s view, fake news reports are misleading by design in that their purveyors employ some strategy that is likely to mislead. We are not sure whether this is more illuminating than simply saying that “fake news is likely to mislead” unless one has something to say about the strategy employed.

the former. If a news outlet publishes a false report, according to which Angela Merkel had a friendly meeting with Kim Jong Un in Uganda, and the outlet publishes this report in a deceitful manner, then this should qualify as fake news, even if the report does not manipulate or exploit any pre-existing biases or heuristics on the consumers' part.<sup>25</sup> There are counterexamples to Gelfert's definition as well: When bullshit fake news are badly crafted, they neither aim at instilling false beliefs in the audience (in virtue of being bullshit), nor are they objectively likely to mislead their audience (in virtue of being badly crafted).<sup>26</sup>

We take it, however, that Gelfert's key insights can be accommodated by our definition. Gelfert characterizes the way in which fake news is designed to operate as one that is "unconstrained by the truth." He distinguishes two ways in which this lack of constraint by the truth may manifest: Fake news either

aims to instil falsehoods in its target audience [...], or [...] is objectively likely to mislead its target audience, its real goal being (for example) the generation of clickbait through sensational claims that attract an online audience. (108)

Extensionally, this boils down to the two ways in which news reports may lack truthfulness that we distinguished. In our view, it is the *distributor* of fake news who is not constrained by the truth. The distributor is untruthful, and that just means that she either intends to deceive (*Deception Dimension*) or distributes content without regard for the truth (*Bullshit Dimension*). Thus, even though we agree with Gelfert that "[t]he example of the role of Macedonian clickbait farms [...] suggests that the deliberate nature of fake news does not necessarily consist in the intention to manipulate others by instilling specific false (or malicious) beliefs in them" (107), it would be a mistake, in our view, to shift the attention away from the distributor's attitude toward the truth altogether. Instead of looking for a specific design of fake news such that it is objectively likely to mislead its target audience (*Effect Dimension*), what makes the news reports by the Macedonian teenagers fake news is their disregard for the truth (*Bullshit Dimension*).

Apart from the *Deception Dimension*, we agree with Gelfert that fake news is false or misleading (*Truth Dimension*) and that it is not a phenomenon of the Internet alone (*Media Dimension*). We disagree, however, that fake news aims at widespread circulation (*Virality Dimension*) and is presented as news (*Appearance Dimension*). Moreover, we suggest substituting the *Effect* with the *Bullshit Dimension*.

<sup>25</sup> Of course, any report potentially exploits some sort of bias on some agent's part. But if the condition is understood in this weak sense, this feature cannot be distinctive of fake news.

<sup>26</sup> The latter point also applies to Grundmann's (2020) definition.

### 5.3 Dentith's Definition: Misleading or False and Intended to Deceive

Dentith's (2017) discussion of fake news shifts between an account of what fake news is and what allegations of fake news amount to. He writes (2017, 66) that "fake news is an allegation that some story is misleading – it contains significant omissions – or even false – it is a lie – designed to deceive its intended audience." In a later paper, he does not talk about the allegation anymore and writes: "Fake news is a misleading story intended to deceive some target audience" (Dentith 2018, 24). We will discuss the later version.

Dentith's definition incorporates the *Truth* ("misleading") and the *Deception Dimension* ("intended to deceive"), remaining mostly silent on the other dimensions. It is unclear whether the definition incorporates the *Bullshit Dimension*. As we will see, Dentith's statements about fake news being "intended to deceive" can be interpreted as including or excluding bullshit fake news.

We agree with Dentith that fake news can be misleading (*Truth Dimension*), although we have a somewhat broader understanding of misleadingness. Dentith thinks of misleadingness as being "due to the selective way in which [...] some but not all of the data" is presented (2018, 26). Accordingly, he takes a news report to be misleading if "it contains significant omissions" (2017, 66; 2018, 24). We prefer to allow for many other ways in which statements can be misleading, for instance, by presupposing or implicating something false.<sup>27</sup> We don't take this to be a point of serious disagreement, though. Dentith's reference to omission of facts is perhaps best read as *one way* in which a story can be misleading.

What is more important is that the *Bullshit Dimension* seems to be altogether missing in Dentith's definition. By characterizing fake news as being "intended to deceive some target audience," Dentith seems to exclude the case of the Macedonian teenagers, who clearly did not intend to deceive their audience. Yet, like Rini, he can analyze these cases as bullshit, as long as he agrees with Frankfurt that bullshit comes along with a second-order intention to deceive about one's own indifference to the truth. If understood this way, Dentith's requirement of an "intention to deceive" is close to congruent to the lack of truthfulness postulated in our definition. But as already laid out in connection with Rini, we take it to be advisable not to commit to the deceitfulness of bullshitters in a definition of fake news. Unfortunately, Dentith does not resolve this ambiguity in his papers.

<sup>27</sup> Perhaps one could argue that presupposition is itself an instance of omission because the presupposition vanishes once all the information is provided. Even then, though, cases remain in which something that is irrelevant is made to look relevant by mentioning it. We take it to be wise not to commit on controversial points here and stick with a broad notion of misleadingness.

Summing up, we can say that Dentith's definition resembles ours closely. On the *Truth Dimension*, our definitions both stress the falsity or misleadingness of fake news (although Dentith brushes over some ways in which reports can be misleading). Dentith's requirement that fake news is intended to deceive either excludes important examples of fake news or it commits to Frankfurt's controversial postulate about the bullshitter's deceitfulness regarding her own attitude towards truth. In the latter case, this boils down to subsuming the *Bullshit Dimension* under the *Deception Dimension* instead of keeping them separate as two ways in which lack of truthfulness may manifest.

#### 5.4 Mukerji's Definition: Bullshit Asserted in the Form of News

Mukerji (2018, 929) argues that fake news is “bullshit asserted in the form of a news publication.” This is an elegant definition and an interesting view, since it differs considerably from other definitions. It does not require fake news to be false or misleading (*Truth Dimension*), but focuses instead on the *Bullshit* and *Appearance Dimension*. There are two striking differences to our and most other definitions of fake news.

The first is that, since bullshit can be true, fake news can be true as well, in Mukerji's view. The definition abandons the *Truth Dimension* entirely, treating falsity and misleadingness as merely contingent, albeit highly prevalent features of fake news.

Can fake news be true? Let us consider a thought experiment.

##### The Clinton Report

A magazine fabricates a news story about the Clintons, according to which they are running a child porn ring. The authors take the story to be false and propagate it with the intention to deceive their audience. As it happens, the Clintons are running a child porn ring. The story is true.

Is this report an instance of fake news? If the answer is “yes,” fake news reports are “fake,” not because they lack truth, but because their distributors pretend to be concerned with the truth when they are not. This squares well with a more normatively loaded understanding of news, on which a defining feature of news is that their distributors care about the truth. Fake news distributors fake that concern, thus producing fake news.

Despite the appeal of this line of thought, we are inclined to hold on to the *Truth Dimension*. When it comes to societally relevant phenomena, philosophy should not roam unnecessarily far from the understanding of the phenomenon in other scientific fields—or society, for that matter. A definition according to which

fake news need not lack truth gives up a feature that governs much of the public and scientific thinking about fake news. This considerably diminishes the chances of informing public and scientific discourse.

There is a second pragmatic reason for holding on to the *Truth Dimension*. In view of the upsurge of “post-truth” tendencies in politics and parts of public discourse, there cannot be enough emphasis on the value of truth and reality. There is a difference between facts and what people believe to be facts, and we should stick to truth as the ultimate goal of inquiry. Journalism, in particular, should retain the quest for truth as one of its cornerstones. Mukerji’s move to locate the “fake” in “fake news” entirely in internal features of news distributors shifts the focus away from the ultimate costs of fake news: people holding false beliefs about the world.

This shows clearly that “fake news” cannot be defined without conceptual engineering. We need to balance candidate dimensions carefully against each other and evaluate their advantages and disadvantages. In our view, the trade-off between the elegance of Mukerji’s definition and the normative costs of allowing for true fake news turns out in favor of the *Truth Dimension*. In our view, the Clinton Report is not fake news, but a different phenomenon, which is problematic in its own way and may be called “deceitful propaganda.” Deceitful propaganda can be true and is not the same as fake news. It can (also) be fake news only if lacking truth.

Let’s turn to a more sturdy problem of Mukerji’s view. Mukerji dismisses the *Deception Dimension*. Depending on one’s view on lying, this entails that fake news cannot be lies. In Mukerji’s own view, the problem does not arise. Mukerji takes lies to be assertions that the speaker knows to be false and calls this the “standard definition of lying” (Mukerji 2018, 941). On this understanding of lying, bullshit and lies are not exclusive categories. Someone may state what they know to be false and be indifferent to this feature of their assertion all the same, thus lying and bullshitting simultaneously.

However, what Mukerji calls the “standard definition” is highly contested. Many hold that lies go along with an intention to deceive about what is asserted. On a prominent view, “[a] lie is a statement made by one who does not believe it with the intention that someone else shall be led to believe it” (Isenberg 1973, 248).<sup>28</sup> Since bullshitters at best intend to deceive about their indifference to the truth, but not about what they say, Mukerji’s definition excludes lies from being fake news, on such a deceptionist view of lying.

The problem subsists even if lying does *not* entail a deceptive intention. While lying and bullshitting may not be exclusive categories, many lies *do* go along with an intention to deceive about what is asserted. These lies cannot be bullshit. But in

<sup>28</sup> See also Primoratz (1984, 54n2).

connection with fake news these lies matter. When GDR newspapers reported that the US Air Force had been throwing beetles from the sky, they intended to make GDR citizens believe that very claim. Today, we see news outlets such as Breitbart News, RT, or Sputnik pursue the same strategy. According to our definition, intentionally spreading false news to achieve a political goal (such as changing people's minds about some fact) is a straightforward case of fake news. According to Mukerji's definition of fake news, these cases will have to be assessed differently.

Mukerji can respond by arguing that the GDR example and others like it are bullshit and not lies because the speaker's *ultimate* goal is not to deceive, but to reach some political goal. However, that is implausible, since it would classify virtually any lie as mere bullshit. Only Augustine's (2002) "real lies" would still qualify as lies. But cases in which a person lies just for the sake of deceiving their audience, without *any* further goal, are highly unusual. Virtually anything we call a "lie" would collapse into bullshit.

For these reasons, we disagree with Mukerji about the *Truth*, *Deception*, and *Appearance Dimension*, while agreeing on the others.

### 5.5 Zimmermann and Kohring's Definition: Recent Disinformation

In communication science and journalism, practitioners and scholars have largely shifted to use the term "disinformation" instead of "fake news" (e.g., House of Commons: DCMS 2019; Habgood-Coote 2018; Wardle & Derakhshan 2017; Wardle 2017; Marwick & Lewis 2017). Classically, disinformation is false information spread deliberately to deceive. It differs from misinformation, which is simply false information and does not require a deceptive intention.

It is easy to see that we cannot simply identify fake news with disinformation. "Disinformation" has both a broader and narrower extension than "fake news." It is not confined to news and thus too broad: Ads on billboards or orally told lies about a disliked classmate can be disinformation if false and deceitful. Fake news, in turn, requires no intention to deceive; it may also be bullshit. Thus, disinformation is too narrow to include the fake news produced by the Macedonian teenagers.

Zimmermann & Kohring (2018) offer a definition that is intended to help. Fake news, they say, is "recent disinformation," where this has to be understood in a specific way. Their use of "disinformation" includes a lack of truthfulness in our sense: it requires either an intention to deceive or a disregard for truth. This use of "disinformation" is highly idiosyncratic, as "disinformation" usually does not include cases of bullshit.

The definition moreover relies on an idiosyncratic conception of "recent disinformation," which is interpreted as a "form of journalistic communication"

(Zimmermann & Kohring 2018, 532). For disinformation to be recent, it not only has to concern recent events; it also needs to be addressed to a public. Based on these terminological clarifications, their definition instantiates the *Truth*, *Deception*, and *Bullshit Dimension*, while waiving the others. Their view is therefore roughly equivalent with ours.

Yet, their definition is unfortunate. Of course, “fake news” is regularly used as a discursive weapon and so an alternative term might be helpful. However, Zimmermann and Kohring’s definition only worsens the situation. First, their idiosyncratic use of “recent disinformation” invites misunderstandings even within academia. Secondly, their definition aggravates a general confusion with the use of “information.” Let us explain.

There are many scholars and practitioners in the debate who talk about “false and misleading information.” At first glance, this seems fine because it seems to capture the two ways in which news may lack truth. And, of course, there are understandings of “information” by which talk of misleading information is sensible. Such talk would, for example, make sense if “information” were to be understood as referring to data or Gricean utterances (Grice 1957).

In the context of fake news, however, talk of misleading information turns out conceptually nonsensical. Neither data nor utterances yield the right interpretation of information in this connection. Talking about “information” as data in the context of fake news is particularly unhelpful. Since we are concerned with news, i.e., a specific type of assertions, a semantic understanding is required.<sup>29</sup> But an understanding in terms of (Gricean) utterances is not helpful either. Utterances are actions carried out by someone in some place at some time. Calling those “information” is yet another idiosyncrasy. We thus agree with Zimmermann and Kohring that, in the context of fake news, “information” (and thus “disinformation”) needs to refer to the communicative *content* of utterances. But then information cannot be misleading. *Utterances* are misleading when their communicative content is false. But misleadingness cannot be a property of the content of an utterance itself.<sup>30</sup>

The upshot is that defining fake news as “recent disinformation” risks causing serious misunderstandings. These are partly due to the already existing confusion in the common use of the terms “information” and “disinformation.” They are further aggravated by the idiosyncrasy of the use of words in Zimmermann and Kohring’s definition. Yet, it is surprisingly similar to our definition. In their view, too, fake news is false or misleading and distributed with an intention to deceive or with an indifference to the truth. We seem to disagree whether to locate fake

<sup>29</sup> The understanding relevant in the context of fake news is, arguably, information as a semantic concept. See Lenski (2010) or Floridi (2011).

<sup>30</sup> There is also the possibility to define “misleading” differently than we did: namely, as being likely to bring about false beliefs. This would not solve the problem, however, because it is still the utterances (as events in the world) that are capable of causing false beliefs.

news on the level of utterances or on the level of their communicative content (i.e., information). In contrast to Zimmermann and Kohring, we view fake news not as recent disinformation but as *containing* recent disinformation (in Zimmermann and Kohring's sense of "recent disinformation").

## 5.6 Comparing the Definitions

We can now compare the definitions all at once:

Dimension Definition	Truth	Deception	Bullshit	Appearance	Effect	Virality	Media
Jaster & Lanius: News Reports Lacking Truth and Truthfulness	YES	YES	YES	NO	NO	NO	NO
Rini: False Stories with Intention to Deceive	YES	YES	YES	NO	???	YES	???
Gelfert: Presentation of Claims as News by Design Misleading	YES	YES	???	YES	YES	YES	NO
Dentith: Misleading or False and Intended to Deceive	YES	YES	???	NO	NO	NO	NO
Mukerji: Bullshit Asserted in the Form of News	NO	NO	YES	YES	NO	NO	NO
Zimmermann & Kohring: Recent Disinformation	YES	YES	YES	NO	NO	NO	NO

As the table shows, there is less dispute with respect to some of the potential criteria of fake news than the seeming diversity of definitions suggests. At their core most definitions are more similar to each other than one would expect.

It has become clear that apart from Mukerji's all definitions contain a more or less explicit commitment to the *Deception Dimension*: They all postulate the distributors' intention to deceive, either about the content of the report or about the distributor's mindset. All definitions but Mukerji's contain also a commitment to the *Truth Dimension*.<sup>31</sup> There is broad consensus that a definition will have to contain some reference to falsity and/or misleadingness. When it comes to the

<sup>31</sup> Even on Mukerji's view, fake news will almost always be false because bullshitting is prone to lead to falsities.

*Bullshit Dimension*, Dentith and Gelfert's definitions are the only apparent exceptions to the otherwise prevalent view that fake news can be bullshit.

The *Media Dimension* is not clearly part of any of the definitions discussed. Although many of the discussions of fake news take its *online* distribution as a paradigm, all theorists agree that offline fake news exists. The same holds for the *Effect Dimension*. Only Gelfert characterizes fake news, in part, by its objective likelihood of instilling falsehoods in the audience.

The actual dispute among theorists comes down to the *Appearance* and *Virality Dimension*. We noted in connection with Rini's account that news, as such, is broadcast to a public. Any definition treating fake news as "news" in this minimal sense can therefore dispense with the *Virality Dimension*. The remaining question is whether fake news mimics ordinary news publications, as Gelfert and Mukerji seem to think. A president's tweet or a populist party's Facebook post does not resemble traditional or ordinary news in any meaningful way. Even the websites run by Macedonian teenagers resemble tabloid press news at best. This puts the ball into Gelfert and Mukerji's field. Does Mukerji intend "published in the form of a news publication" to be saying more than published as news in the minimal sense? If so, what is the stronger sense in which fake news has the form of a news publication? The same question arises for Gelfert.

All in all, our discussion has brought to light many points of agreement in the manifold definitions that have been put forward. In the next section, we are going to turn to the merits of having a definition of "fake news" in the first place, using our understanding of fake news as the working definition.

## 6. The Importance of Defining "Fake News": Epistemic Threats, Boundary Work, and Paradigm Repair

Fake news causes severe epistemic problems for societies. First, due to its lack of truth, the spread of fake news tends to generate false beliefs and uninformed decisions. We have seen this effect in the case of Brexit. Moreover, fake news has recently led to violence against innocents in India, Myanmar, and many other countries, because people mistakenly believed innocent people to be dangerous or to have committed crimes (e.g., Mozur 2018).

Secondly, due to its lack of truthfulness, the spread of fake news fosters distrust. Hardly any of our knowledge stems from direct perception. The vast majority of it is gained from the testimony of others, above all journalists. In selecting the sources we accept as knowledgeable, trust is essential. Thus, in undermining the public's trust in news generally, fake news can actively diminish the amount of what people (take themselves to) know.<sup>32</sup>

<sup>32</sup> Whether the phrase in parentheses is needed will depend on whether one accepts an externalist or internalist epistemology.

Third, reports lacking truth and truthfulness undermine societies' capability for deliberation and may thus foster illegitimate collective decisions. As Habermas (1996, 304) famously put it,

deliberative politics acquires its legitimating force from the discursive structure of an opinion- and will-formation that can fulfill its socially integrative function only because citizens expect its results to have a reasonable quality.

Based on a broadly Habermasian view, truthfulness is required for deliberation; deliberation, in turn, is required for the legitimacy of political decisions (Cohen 1997). Only sufficiently informed people are able to deliberate in a way that legitimizes the outcomes of deliberative processes.

Apart from its direct epistemic effects, fake news is also prone to lead to an erosion of the norms of truth and truthfulness. According to Lewis (1979, 347), "the conversational score does tend to evolve in such a way as is required in order to make whatever occurs count as correct play." If many participants in a debate untruthfully make untrue utterances, this may become normal—in a descriptive and ultimately even in a normative sense. If such utterances become a regular part of public debate, they will change the standards by which we evaluate the appropriateness of subsequent contributions to it.

The risk of norm erosion is real for at least two reasons. First, the distributors of fake news frequently employ what has been called the "Firehose of Falsehood" method—they spread rapidly, repetitively, and continuously over multiple channels (Paul & Matthews 2016), thus violating the epistemic standards for contributions to the public debate openly and repeatedly. This pushes the standards of public discourse in a direction where fake news counts as correct play.

Secondly, fake news targets truth and truthfulness in a place where they matter more than in most other areas. Truth and truthfulness are the central values of journalism (Kovach & Rosenstiel 2014), put into practice by operationalizing the norm of "journalistic objectivity" (Godler & Reich 2013). Democratic societies are applying particularly high epistemic standards to the production and propagation of news. By undermining the epistemic standards of news, fake news targets the norms of truth and truthfulness in a realm in which their protection is particularly important and their deterioration leads to an epistemically disastrous situation.

Of course, there is the realistic possibility of effective countermeasures. But a necessary first step is understanding the phenomenon of fake news, pinpointing its epistemic risks, and calling it out. That is why a viable definition of "fake news" is so crucially important.<sup>33</sup>

<sup>33</sup> Thus, we agree with Brown (2019, 144) that it is possible to use "fake news" in a linguistically and politically unproblematic manner and even fruitfully if done so with care. To use the term with care, arguably, requires sufficient clarity about its meaning. Moreover, we provide in this chapter further evidence for Brown's claims that there is significant agreement among academics and other members of the public about its key features and that the concept is not unnecessary.

In journalism studies, two concepts keep coming up when the task of shielding journalism from deterioration is discussed: the concept of *boundary work* and the concept of *paradigm repair*. Boundary work (Carlson 2016; Gieryn 1983; Lewis 2012) is a necessary means for keeping up the epistemic standards for news publications (and also for knowledge more generally). It consists in demarcating what it takes to be an X—a journalist, for instance. Focusing on the scientific community, Gieryn (1983, 4–5) defines boundary work as

the discursive attribution of selected qualities to scientists, scientific method and scientific claims for the purpose of drawing a rhetorical boundary between science and some less authoritative, residual non-science... demarcating, defending, expanding, contesting the limits of legitimate science, the real scientist from the pseudo scientist.

As Carlson points out, this is relevant to journalism as well, since boundaries affect “the allotment of epistemic authority which (...) denotes knowledge practices accepted by others as legitimate” (2016, 316) and “in the case of news, the validity of any story rests on a shared belief that it is a legitimate form of knowledge” (360). Boundary work is therefore essentially important for the protection of epistemic standards in journalism and science alike.

Fake news poses a clear instance of broadcasting against which journalism needs to be set off. Fake news is alien to journalism in much the same way as pseudoscience is alien to science. It is therefore crucial to set the boundaries of journalism in a way that (1) excludes fake news from the set of legitimate practices and (2) draws attention to the features of fake news that make its distribution unacceptable.

The most straightforward means to engage in this task is through explicit categorization and definitions (Carlson 2016). Knowing what fake news is—and knowing that violations of truth and truthfulness are at its core—helps to set the boundaries of legitimate journalistic practices in their proper place: They will include practices that truthful agents employ to deliver true reports of events and exclude practices that do not obey this norm.

The second (and related) concept that is essential in connection with protective measures against the deterioration of journalistic practices is the concept of paradigm repair. Drawing on Kuhn’s (2012) work about science, journalistic paradigm repair is the practice of journalists to sanction violations of norms with the goal of strengthening their validity and compliance (Neuberger 2017). This aims at “reestablishing the authority, validity, and credibility of professional journalism in times of a perceived professional or organizational crisis” (Koliska & Steiner 2019). One important part of paradigm repair is to “consciously articulate (...) taken-for-granted assumptions [of journalism] in explicit terms” (Vos & Moore 2020).

The differences between boundary work and paradigm repair are subtle. While paradigm repair focuses on calling out *internal* attacks on the journalistic paradigm, boundary work focuses on threats from agents at the boundaries of journalism (Carlson & Lewis 2015). Both practices, however, aim at strengthening the norms within journalism. In our view, neither of the two practices should fall exclusively to journalists. Society as a whole needs to sanction violations of crucial paradigms of knowledge generation, assertability, and knowledge transfer. Doing so is crucial to preventing standards from shifting and the contexts of knowledge from blurring. Philosophy, too, has a contribution to make. By drawing attention to the lack of truth and truthfulness in the case of fake news, truth and truthfulness come into sharp focus as paradigms of journalism and norms that need to be upheld and defended against threats.

In summary, neither of the epistemic problems that come along with the spread of fake news is unavoidable. Fake news can be corrected, fake news distributors can be called out for their untruthfulness, and epistemic norms can be defended by publicly criticizing the distribution of fake news. But to do this effectively, we need to explicate the concept of fake news as clearly as possible to better understand the underlying phenomenon and to facilitate appropriate countermeasures. This chapter is intended as a contribution to this endeavor.

## 7. Conclusion

We started out by presenting our definition and differentiating fake news from propaganda, satire and parody, conspiracy theories, and journalistic errors. Then, we introduced seven potential dimensions of the concept of fake news. This allowed us to systematically compare some prevalent definitions with respect to their extensional scope, practical utility, and conceptual transparency. Most definitions can be interpreted in several ways. Often, the best and most charitable interpretations are very similar to each other. One result of this chapter is that the controversy about the definition of “fake news” might ultimately not run deep. Most definitions agree that fake news lacks truth and is published with problematic intentions. Some focus on more epistemic questions; some on its mimicking “proper” news. But, as we have tried to show, our definition is quite congruent with most other definitions—at least being interpreted plausibly. Other interpretations make them not only less similar to ours, but also encounter problems. Finally, we have argued that defining “fake news” (broadly along the lines of our definition) is useful because it lays open the epistemic problems resulting from fake news and enables us to develop effective countermeasures against its distribution and the resulting deterioration of epistemic norms in the public debate.<sup>34</sup>

<sup>34</sup> Both authors contributed equally to this chapter.

## References

Allcott, H. & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2): 211–36.

Amarasingam, A. & McChesney, R. W. (eds.) (2011). *The Stewart/Colbert Effect: Essays on the Real Impacts of Fake News*. Jefferson, NC: McFarland & Company, Inc.

Augustine (2002). *Treatises on Various Subjects (The Fathers of the Church, Vol. 16)*. Washington, DC: Catholic University of America Press.

Benkler, Y., Faris, R., & Roberts, H. (2018). *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. New York: Oxford University Press.

Bounegru, L., Gray, J., Venturini, T., & Mauri, M. (2018). A Field Guide to “Fake News” and Other Information Disorders: A Collection of Recipes for Those Who Love to Cook with Digital Methods. Amsterdam: Public Data Lab.

Brown, E. (2019). “Fake News” and Conceptual Ethics. *Journal of Ethics and Social Philosophy*, 16(2): 144–54.

Carlson, M. (2016). Metajournalistic discourse and the meanings of journalism: definitional control, boundary work, and legitimization. *Communication Theory*, 26(4): 349–68.

Carlson, M. & Lewis, S. C. (2015). *Boundaries of Journalism: Professionalism, Practices and Participation*. London/New York: Routledge.

Carnap, R. (1950). *Logical Foundations of Probability*. Chicago, IL: University of Chicago Press.

Carson, T. L. (2006). The definition of lying. *Noûs*, 40(2): 284–306.

Carson, T. L. (2010). *Lying and Deception: Theory and Practice*. Oxford: Oxford Scholarship Online.

Cohen, J. (1997). Deliberation and democratic legitimacy. In J. Bohman & W. Rehg (eds.), *Deliberative Democracy: Essays on Reason and Politics* (67–92). Cambridge, MA: MIT Press.

Dentith, M. R. X. (2014). *The Philosophy of Conspiracy Theories*. Dordrecht: Springer.

Dentith, M. R. X. (2017). The problem of fake news. *Synthese*, 8(1–2): 65–79.

Dentith, M. R. X. (2018). What is fake news? *University of Bucharest Review: Literary and Cultural Studies Series*, 8(2): 24–34.

Egelhofer, J. L. & Lecheler, S. (2019). Fake news as a two-dimensional phenomenon: a framework and research agenda. *Annals of the International Communication Association*, 43(2): 97–116.

Fallis, D. (2015). What is disinformation? *Library Trends*, 63(3): 401–26.

Fallis, D. & Mathiesen, K. (2019). Fake news is counterfeit news. *Inquiry*, 1–20, DOI: 10.1080/0020174X.2019.1688179.

Fallis, D. & Stokke, A. (2017). Bullshitting, lying, and indifference toward truth. *Ergo*, 4(10): 277–309.

Floridi, Luciano (2011). *The Philosophy of Information*. Oxford: Oxford University Press.

Frankfurt, H. G. (2005). *On Bullshit*. Princeton, NJ: Princeton University Press.

Gelfert, A. (2018). Fake news: a definition. *Informal Logic*, 38(1): 84–117.

Gieryn, T. F. (1983). Boundary-work and the demarcation of science from non-science: strains and interests in professional ideologies of scientists. *American Sociological Review*, 48(6): 781–95.

Godler, Y. & Reich, Z. (2013). How journalists think about facts: theorizing the social conditions behind epistemological beliefs. *Journalism Studies*, 14(1): 94–112.

Grice, H. Paul (1957). Meaning. *The Philosophical Review*, 66(3): S377–88.

Grundmann, T. (2020). Fake News: The Case for a Purely Consumer-Oriented Explication. *Inquiry: An Interdisciplinary Journal of Philosophy*, online first: DOI: 10.1080/0020174X.2020.1813195.

Habermas, J. (1996). *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*. Cambridge, MA: MIT Press.

Habgood-Coote, J. (2018). Stop talking about fake news! *Inquiry: An Interdisciplinary Journal of Philosophy*, 62(9–10): 1–33.

Hale, V. (2017). Revealed: 1,000-man mob attack police, set Germany's oldest church alight on New Year's Eve. Breitbart News. <https://www.breitbart.com/europe/2017/01/03/dortmund-mob-attack-police-church-alight/>.

Horne, B. D. & Adali, S. (2017). This Just in: fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. <https://arxiv.org/abs/1703.09398>.

House of Commons: DCMS (Digital, Culture, Media and Sport Committee) (2019). Disinformation and 'fake news'. Final Report. <https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1791/1791.pdf>.

Isenberg, A. (1973). Deontology and the ethics of lying. In A. Berleant & A. Isenberg (eds.), *Aesthetics and the Theory of Criticism: Selected Essays of Arnold Isenberg* (245–64). Chicago, IL: University of Chicago Press.

Jaster, R. & Lanius, D. (2018). What is fake news? *Versus*, 127(2): 207–24.

Jowett, G. & O'Donnell, V. (2006). *Propaganda and Persuasion* (4th edn). Thousand Oaks, CA: Sage.

Keeley, B. L. (1999). Of conspiracy theories. *The Journal of Philosophy*, 96(3): 109–26.

Klein, D. & Wueller, J. (2017). Fake news: a legal perspective. *Journal of Internet Law*, 20(10): 5–13.

Koliska, M. & Steiner, L. (2019). Paradigm repair. *The International Encyclopedia of Journalism Studies*, 1–7, <https://doi.org/10.1002/9781118841570.iejs0036>.

Kovach, B. & Rosenstiel, T. (2014). *The Elements of Journalism: What Newspeople Should Know and the Public Should Expect* (3rd edn). New York: Three Rivers Press.

Kuhn, Thomas S. (2012). *The Structure of Scientific Revolutions* (4th edn). Chicago, IL: University of Chicago Press.

Lanius, D. (2019). *Strategic Indeterminacy in the Law*. New York: Oxford University Press.

Lenski, Wolfgang (2010). Information: a conceptual investigation. *Information* 1(2): 74–118.

Levy, N. (2017). The bad news about fake news. *Social Epistemology Review and Reply Collective*, 6(8): 20–36.

Lewis, D. K. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic*, 8: 339–59.

Lewis, S. C. (2012). The tension between professional control and open participation: journalism and its boundaries. *Information, Communication & Society*, 15(6): 836–66.

McNair, B. (2017). *Fake News: Falsehood, Fabrication and Fantasy in Journalism*. Abingdon: Routledge.

Mandlik, P. (2007). Shit happens. *Episteme: A Journal of Social Epistemology*, 4(2): 205–18.

Marwick, A. E. & Lewis, R. (2017). Media manipulation and disinformation online. Data & Society. [https://datasociety.net/pubs/oh/DataAndSociety\\_MediaManipulationAndDisinformationOnline.pdf](https://datasociety.net/pubs/oh/DataAndSociety_MediaManipulationAndDisinformationOnline.pdf).

Michaelson, E., Sterken, R., & Pepp, J. (2019). What's new about fake news? *Journal of Ethics and Social Philosophy*, 16(2), 67–92.

Mozur, P. (2018). A genocide incited on Facebook, with posts from Myanmar's military. *New York Times*, 15, <https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html>.

Mukerji, N. S. (2018). What is fake news? *Ergo*, 5: 923–46.

Mustafaraj, E. & Metaxas, P. T. (2017). The fake news spreading plague: was it preventable? In *Proceedings of the 2017 ACM on Web Science Conference* (June) (pp. 235–9), <https://doi.org/10.1145/3091478.3091523>.

Nelson, J. L. & Taneja, H. (2018). The small, disloyal fake news audience: the role of audience availability in fake news consumption. *New Media & Society*, 20(10): 3720–37.

Neuberger, C. (2017). Journalistische Objektivität. Vorschlag für einen pragmatischen Theorierahmen. *Medien & Kommunikationswissenschaft*, 65(2): 406–31.

Paul, C. & Matthews, M. (2016). *The Russian “Firehose of Falsehood” Propaganda Model: Why It Might Work and Options to Counter It*. Santa Monica, CA: RAND Corporation.

Popper, K. (1992). Die offene Gesellschaft und ihre Feinde, Volume II: Falsche Propheten. Hegel, Marx und die Folgen. Tübingen: J.C.B. Mohr.

Primoratz, I. (1984). Lying and the “methods of ethics”. *International Studies in Philosophy*, 16(3): 35–57.

Recanati, F. (2004). *Literal Meaning*. Cambridge: Cambridge University Press.

Rini, R. (2017). Fake news and partisan epistemology. *Kennedy Institute of Ethics Journal*, 27(2S): E-43–64.

Saul, J. M. (2012). *Lying, Misleading, and What Is Said: An Exploration in Philosophy of Language and in Ethics*. Oxford: Oxford University Press.

Shao, C., Ciampaglia, G. L., Varol, O., Flammini A., & Menczer, F. (2017). The spread of fake news by social bots. <https://a51.nl/sites/default/files/pdf/1707.07592.pdf>.

Shin, J., Jian, L., Driscoll, K., & Bar, F. (2018). The diffusion of misinformation on social media: temporal pattern, message, and source. *Computers in Human Behavior*, 83: 278–87.

Silverman, C. & Alexander, L. (2016). How teens in the Balkans are duping Trump supporters with fake news. *Buzzfeed*. <https://www.buzzfeednews.com/article/craigsilverman/how-macedonia-became-a-global-hub-for-pro-trump-misinfo>.

Tandoc, E. C., Lim, Z. W., & Ling, R. (2018). Defining “fake news”. *Digital Journalism*, 6(2): 137–53.

*The Guardian* (2017). German police quash Breitbart story of mob setting fire to Dortmund church. <https://www.theguardian.com/world/2017/jan/07/german-police-quash-breitbart-story-of-mob-setting-fire-to-dortmund-church>.

Vos, T. P. & Moore, J. (2020). Building the journalistic paradigm: beyond paradigm repair. *Journalism*, 21(1): 17–33.

Wardle, C. (2017). Fake news: it's complicated. First Draft.

Wardle, C. & Derakhshan, H. (2017). *Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making*. Strasbourg: Council of Europe report DGI.

Zimmermann, F. & Kohring, M. (2018). “Fake News” als aktuelle Desinformation. Systematische Bestimmung eines heterogenen Begriffs. *Medien & Kommunikationswissenschaft*, 66(4): 526–41.

Zuckerman, E. (2017). Stop saying “fake news”. It’s not helping. <http://www.ethanzuckerman.com/blog/2017/01/30/stop-saying-fake-news-its-not-helping/>.

# 2

## Good News, Bad News, Fake News

*Duncan Pritchard*

### 0. Introductory Remarks

The contemporary terminology of ‘fake news’ is largely a term of art. Pundits, politicians, journalists and such like use this terminology in an undisciplined way, as one would expect. In particular, it is often used as an insult to throw at one’s enemies. As such, it can sometimes be rather misleading, in that something might be labelled ‘fake news’ when its only crime is to provide information that some people find contrary to their interests.<sup>1</sup> Naturally, we wouldn’t want an account of fake news that took this usage at face value (which is not to say, of course, that it is irrelevant to this account why this terminology is used in this way at all), since it would make the concept so broad as to be empty. For one thing, it would entail that the concept could be legitimately applied in contradictory ways, such that the very same phenomenon can be rightly classified as both fake news and not fake news. Accordingly, if we want to make some philosophical headway understanding this phenomenon, it’s important that we first do a little conceptual ground-clearing.

The kind of account of fake news that I want to defend is one on which fake news is to be contrasted with genuine news (including genuine news that has a poor epistemic pedigree), in terms of how it involves deliberately conveying misleading information with an intent to mislead. Fake news is presented as news, but it is not a kind of news at all (not even an epistemically deficient form of news), any more than an excellent forgery is thereby the real thing. Note that on my view fake news needn’t involve the presentation of falsehoods (though it typically will), given that even the literal truth can be misleading (indeed, this is sometimes the most effective way to mislead). In addition, while fake news involves an intent to mislead on my proposal, it doesn’t follow that if it doesn’t mislead then it isn’t fake news; ineffective fake news is still fake news.

Aside from the intrinsic interest in having the right account of fake news in hand, there are also practical implications to this account. In particular, as we will

<sup>1</sup> This is, of course, how Donald Trump famously uses this terminology.

see, this account of fake news has ramifications for how we might go about dealing with the challenges posed by fake news, at both an individual and structural level.

## 1. Understanding Fake News

In order to understand this proposal, it will be helpful to contrast it with the most developed account of fake news in the philosophical literature, due to Axel Gelfert (2018), not least because this account seems very similar to the view I just outlined.<sup>2</sup> After a comprehensive discussion of different kinds of fake news, and different treatments of this notion, both philosophical and non-philosophical, he ends up with the following definition of fake news: “the deliberate presentation of (typically) false or misleading claims *as news*, where the claims are misleading *by design*” (Gelfert 2018, 108, italics in original).<sup>3</sup> On the face of it, Gelfert seems to be endorsing a very similar account to the one that I offer. As we will see, however, there are in fact important divergences that we need to get a handle on.

Let’s begin with what Gelfert gets right. Gelfert’s account of fake news has the virtue of focusing on the intention and systemic goals of the fake news rather than on the veracity of the output. In particular, that the presentation of false or misleading claims as news must be deliberate, with the goal of misleading (i.e., this is ‘by design’), is clearly right. Unintentional errors in a news report don’t make it fake news, even though this would clearly be a case of news that involves falsehoods (and which might thereby mislead). Moreover, a good satirical publication like *The Onion* or the UK’s *Private Eye* is not fake news, even if sometimes gullible people are taken in by their fake headlines, since it is not designed to mislead. Gelfert is also right that it’s important that fake news is presented as news. As he notes (Gelfert 2018, 110), an advertising campaign could contain falsehoods and be designed to mislead, but that wouldn’t thereby make it fake news. Putting the point in terms of a disjunction between false or misleading claims is also helpful, since sometimes the literal truth can be misleading, as when it leaves out important qualifiers or context. (Indeed, the most effective forms of fake news might well involve no literal falsehood at all.)<sup>4</sup> Note, though, that since false claims are also misleading, we can exclusively focus on the second

<sup>2</sup> For a defense of the very different philosophical stance that ‘fake news’ should not be given a philosophical treatment, on account of how it doesn’t pick out a specific phenomenon, see Habgood-Coote (2018). If the philosophical account that I will be offering of fake news is credible, then it is thereby a (partial) response to this particular critical line.

<sup>3</sup> For some closely related philosophical discussions of fake news, see Levy (2017), Rini (2017), and Rose (2019).

<sup>4</sup> This is one of the problems that faces the account of fake news offered by Lazer et al. (2018), as their proposal explicitly defines fake news in terms of ‘fabricated information’. But as Aikin & Talisse (2018) point out, one can in fact use accurate information to mislead, such as by presenting it in a way that triggers a cognitive bias.

disjunction, as I do in the summary of my proposal above, without loss of explanation (even though there might be a wider dialectical benefit to making it explicit that ‘misleading’ here also covers explicitly false claims).<sup>5</sup>

Nonetheless, despite its many merits, Gelfert’s way of thinking about fake news is not quite right, and the reasons why are important, since they reveal deeper conceptual confusions. As we will see, the problem isn’t quite the formulation above, but rather how Gelfert is unpacking this formulation. Moreover, understanding this point will in turn help us to get a handle on what the account of fake news I am proposing involves.

The first issue with Gelfert’s proposal concerns the idea, which we have just endorsed, that the claims at issue in fake news are misleading by design. We have naturally unpacked this idea as meaning that it must be a goal of the fake news to mislead. As we noted, this feature of fake news would explain why a satirical news program or publication wouldn’t qualify as fake news. Note that it is consistent with our characterization of this clause that the agents involved have other motives in play, and even that these other motives are more central to their overall interests (since it is merely *a* goal of the fake news, and not *the* goal, or the overarching goal).<sup>6</sup>

This isn’t the rendering of this clause that Gelfert has in mind, however. Instead, he explicitly argues that while fake news is misleading by design, it needn’t be designed to mislead. That is, what is important to fake news on Gelfert’s account is that it intentionally employs misleading information, but it needn’t be the goal of the fake news that it has the effect of misleading its audience (even though, as we will explain in a moment, on Gelfert’s view fake news must *actually* mislead its audience). The interesting example he gives to illustrate this point is of putatively fake news that is designed as ‘clickbait’. There are websites whose goal (usually for financial reasons) is to encourage visitors, and Gelfert claims that often fake news is created for the purposes of clickbait. Presenting misleading claims as news can make effective clickbait, but since the goal of clickbait is just to generate a high volume of internet traffic, and not specifically

<sup>5</sup> At least, false claims are characteristically misleading, if not universally so. For example, I am here setting aside such unusual cases as when one knows that one is dealing with someone who consistently lies, in which case their false claims might well be reliable ways to discern the truth, and hence not misleading at all.

<sup>6</sup> This is a point that Rini (2017) seems to miss in her discussion of fake news: “[...] I said that fake news requires intentional deception, but this may be too strong. Deception is not always the primary goal of fake news. Often the motive is financial rather than epistemic” (Rini 2017, §2).

Notice how the idea of fake news involving an intent to mislead is converted into this being the “primary goal” of the fake news. But it is entirely consistent with the idea that an intention to mislead is built into one’s account of fake news that the purveyors of fake news also have other motives in play (and indeed that those other motives play a more primary role in their activities). Note, too, that Rini’s account of fake news is also problematic in that she insists that it involves ‘more than mere lying.’ As we have noted, however, while fake news is designed to mislead, one does not need to lie (i.e., assert a falsehood) in order to do that.

to mislead, Gelfert argues that we should not build an intention to mislead into the definition of fake news.

While this is an intriguing proposal, I don't think that it is plausible on closer inspection. For consider what would be involved in developing a website that presented misleading claims as news purely with the goal of generating internet traffic, with no concern at all for whether the intended audience are actually taken in by these claims (i.e., are actually misled). Here is the rub: if the audience don't find the claims even remotely plausible, then why on earth would they be clicking on them under the guise of them being news? The point is that clickbait that wants to generate internet traffic by offering misleading claims *as news* must be in the business of also wanting to actually mislead its clientele, since that is precisely what is going to ensure that this strategy is successful. It is true that those who develop clickbait (qua fake news) will have other motives in play in addition to intentionally misleading people, but the point is that they must at least have this motive in play for their activities to be coherently pursued. Remember that our account of fake news only demands that it should be intentionally misleading; it doesn't require that this should be the only motive in play, or even that it should be the overarching motive in play.

Of course, we should note that there is a variety of clickbait that simply involves presenting outlandish claims, as there is a recognition that a certain kind of person is likely to click on the site as a result. But such sites do not offer any pretense of being news sources, and so are not even in the market for being considered fake news (including by the lights of Gelfert's account). There is thus no plausible sense to the idea that there can be fake news—i.e., claims that are presented as news that are misleading by design—which doesn't also involve being designed to mislead. Either clickbait is presented as news, in which case it fits our rubric for fake news, or it isn't, in which case it is not a plausible candidate for fake news anyway, and so there is no surprise that it fails to satisfy the rubric.

Now one kind of case that one might think would work in Gelfert's favor in this regard is *bullshit*. As Harry Frankfurt (2005) has persuasively argued, what characterizes bullshit is a complete lack of concern for the truth. This sets bullshit apart from lying, as the liar usually does care about the truth, it's just that they want to hide it from the person they are lying to.<sup>7</sup> In this sense the bullshitter will often be the source of misleading information, but not because her overarching goal is to deceive her audience, but rather because her lack of concern for the truth means that she is sanguine about bearing witness to such epistemically problematic reports. Imagine now that the bullshitter presents misleading information as news. Wouldn't this be a natural case of fake news? If so, given that the bullshitter

<sup>7</sup> See also Cassam's (2018) discussion of the related cognitive trait of epistemic insouciance.

doesn't have any overarching goal to deceive her audience, doesn't that speak in favor of Gelfert's account?

Once we examine this case more closely, however, it becomes clear that rather than supporting Gelfert's view, it in fact counts against it. The key thing to note is that the bullshitter's lack of concern for the truth doesn't entail that she is unconcerned with whether her audience believes what she says. On the contrary, the bullshitter wants to be believed. What distinguishes the bullshitter is rather that it doesn't matter to her whether what she asserts is something that she believes, and thus regards as true.<sup>8</sup> Indeed, her unconcern for the truth will mean that she would rather assert something expedient to her interests than something she believes to be true. While she doesn't have any overarching concern to deceive others, that's only because if it is useful to her to assert the truth, and thereby persuade her audience of the truth, then she will pursue that goal with just as much enthusiasm as she would muster if called upon to assert falsehoods. Crucially, however, insofar as she is prompted by her practical interests to assert misleading claims, then she will be aiming for those claims to be convincing, and hence will be trying to mislead her audience. Yet again, then, we find that we are not being presented with a plausible case in which misleading information is deliberately presented as news, and yet there is no desire to actually mislead the target audience.<sup>9</sup> Accordingly, when bullshit involves the presentation of misleading information as news, then it counts as fake news, but once we understand the details of what is involved in this regard, then this is not something that Gelfert's treatment of fake news can accommodate.<sup>10</sup>

<sup>8</sup> I don't think the case of bullshit is unique in this regard, as there are a number of epistemologically interesting cases of assertion where what might initially look like the expression of belief is in fact nothing of the kind, at least in any robust sense of belief of a kind that would be of interest to epistemologists (e.g., which is a constituent part of rationally grounded knowledge). For further discussion of this point, and some of its epistemological ramifications, see Pritchard (2018a).

<sup>9</sup> I think this point also counts against one of the reasons that Pepp et al. (2019) offer for thinking that fake news needn't involve an intention to mislead:

Consider an organization like the *National Enquirer*. Suppose that the motivation of the proprietors of this organization is to maximize profit, and they incentivize their employees in order to carry out this maximization. It is not all that hard to imagine that each employee of the *National Enquirer* might act either (a) to maximize profit, or (b) to maximize personal gain, with no one intentionally putting forward content with the intention to deceive or mislead anyone else. (Pepp et al. 2019, 74)

But how on earth would these actors ensure that this strategy is successful *without* an intention to mislead? After all, they will want their output to be read and believed, and they know that this will result in their readers being misled. That the readers are misled thus doesn't seem to be a merely predictable (but unintended) consequence of their actions, but one that is actively sought. Remember, too, that it is quite consistent with the claim that fake news involves an intention to mislead that it also involves other motivations, and indeed that those other motivations are more central to the activity.

<sup>10</sup> See Mukerji (2018) for an account of fake news as essentially bullshit presented as news. I think Mukerji's analysis overlooks the issue raised here, however. This is because he seems to assume that it follows from Frankfurter's account of bullshit that the bullshitter, by being indifferent to the truth, is thereby also indifferent to whether they are believed. Consider this passage:

Notice, too, that if Gelfert were serious about arguing that the intention to mislead is not a property of fake news, then it would be unclear why he claims that satirical magazines like *The Onion* are not fake news. After all, they clearly employ misleading information by design, and may even be misleading and yet their aim is not to mislead. On Gelfert's view, then, what element of fake news do they fail to satisfy? The proper moral to be drawn regarding these satirical magazines is thus that they are not fake news precisely because their aim is not to mislead. Indeed, the reason Gelfert (2018, 106) gives for why these venues are not trading in fake news is that they do not 'deliberately' mislead. But of course this entails that an intention to mislead is built into the very notion of fake news.

Another way in which Gelfert's explicit unpacking of his account of fake news is problematic concerns his contention, contrary to my proposal outlined earlier, that fake news is a genuine form of news. For example, he writes that "[...] it must be granted that, just as disinformation is a species of *information*, fake news is [...] a form of *news*" (Gelfert 2018, 103, *italics in original*). This is a puzzling claim, however, not least because disinformation doesn't seem to be a species of information. Indeed, Gelfert (2018, 104) notes Fred Dretske's (1981, 57) famous—and to my mind devastating—remark in this regard that disinformation is no more a species of information than a decoy duck is a kind of duck.<sup>11</sup>

So why does Gelfert think it is important to regard fake news as a genuine form of news? His basic reasoning seems to be that anything that reasonably looks like genuine news must be genuine news. Regarding Dretske's point about the decoy duck, for example, he writes that "if something looks like a duck, swims like a duck, and quacks like a duck, it is awfully difficult to recognize that it is not, in fact, a duck" (Gelfert 2018, 104). But this is a very odd claim for an epistemologist to make in support of the idea that decoy ducks are types of duck. Does it really follow from the purely epistemic fact that it is hard to distinguish decoy ducks from real ducks that the former is genuinely a species of a latter? This seems to confuse an ontological question with an epistemological one.

Unlike a liar, however, who seeks to convince us that a given statement he believes to be false is true, a bullshitter does not care about whether his utterances are true.

(Mukerji 2018, §2)

But this is the wrong contrast to draw. Unlike the liar, it isn't essential to the bullshitter's claims that he is seeking to convince us that *what he believes is false* is true, but he is trying to convince us of his claims (it's just that it doesn't matter to him whether they are true or false). Relatedly, bullshit qua fake news also involves an intention to mislead. Since he misses this point, Mukerji falsely supposes that all that is required on this score on the part of the bullshitter is the intention to mislead the audience about his true motives in making his claims (which of course doesn't entail any narrow intention to mislead with regard to the content of the fake news itself). Note, too, that Mukerji's account is also unable to accommodate the phenomenon of fake news that involves straightforward lying, as he himself concedes (Mukerji 2018, §5).

<sup>11</sup> To use the terminology offered by Twardowski (1979), an adjective like 'decoy' when applied to 'duck' is an 'eliminating' adjective, in that it is suggesting that the target item (the duck) is not present. (Another example of an expression using such an adjective is 'false friend'.) This is in contrast to 'determining' adjectives (like 'red' in 'red coat') or 'confirming' adjectives (like 'actual' in 'actual fact') which don't have this implication. I am grateful to an anonymous referee for alerting me to this point.

Gelfert's overarching concern in this regard appears to be the practical one that recipients of the fake news may not be able to tell it apart from the genuine article. As he puts it, “building truth and veracity into the very definitions of ‘information’ and ‘news’—in other words, making them success terms—does little to address the pressing epistemological problem: how to respond to claims presented to us as true by a putative news source, given that, for all we know, they might (or might not) be fake news” (Gelfert 2018, 104). But this is confused, on multiple levels. To begin with, our epistemic difficulty in distinguishing X from Y is not a reason at all to treat X as a species of Y, particularly if we have antecedent reason, as we surely do in this case, to not treat it as a species of Y. (Indeed, as we will see below, it is easy to construct cases where genuine news is hard to distinguish from non-news, but where there is no temptation thereby to treat the latter as a species of the former.) Moreover, opting on this basis to treat fake news as a genuine species of news doesn’t resolve (or make any more tractable) the ‘pressing epistemological problem’ that Gelfert notes anyway but rather leaves it entirely intact. How does treating fake news as genuine news make it any easier to differentiate between the two? Finally, note that Gelfert is in any case arguing against a straw man here, as no one has suggested that news should be understood as a success term—i.e., that it must be true in order to qualify as genuine news (even if that is the suggestion regarding information). (Indeed, I will be arguing below that there can be false news.)

## 2. News and News Sources

At this point it will be useful to say something about what news is. I think we need to understand news explicitly in terms of sources of news, in the sense that news is a kind of information that one gets from a news source (at least *qua* news source—the reason for this qualification will become apparent in a moment). What makes a source of information a news source is that it is designed to convey accurate information to others about recent events, where that information is not already widely known. This is why tree rings are a source of information but not a news source—one can extract accurate information from them, but they are not designed for this purpose. In contrast, a newspaper or media outlet is designed to convey information and that’s what makes it a news source. Moreover, it is important to a news source that it is reporting information that is not already widely known, as otherwise merely stating the obvious—repeating old news ad nauseum, say—would constitute news. This is also why news is concerned with recent events.<sup>12</sup>

<sup>12</sup> Of course, the news could be ultimately concerned with events in the distant past, as when one makes a discovery regarding the ancient world and this becomes news. But it is the discovery which

Note that while paradigm cases of news sources like newspapers transmit a great deal of information, there can be genuine news sources that only convey a relatively rudimentary amount of information. For example, the beacon fires that were once used to alert townsfolk to invasion are clearly news sources, but the accurate information they were designed to transmit was very limited. It is important, however, that a news source has some degree of range in terms of the kind of news that it provides. A one-line message left on the fridge can convey accurate information, and does so by design, but it is hardly a *source* of news (i.e., as opposed to simply being itself a piece of information). Relatedly, it is also important that news sources are relevantly responsive to the world in terms of the information that they convey. A traffic light conveys accurate information about when traffic should proceed and when it should stop, and does so by design, but it isn't a news source as it is determining the facts in question rather than reporting on them. Finally, note that our concern is with the informational output of news sources *qua* news sources. For example, that the website for a particular news source is not well maintained might give us very important information (e.g., about the credibility of the news source), but this is not thereby news.

That news sources are designed to convey accurate information explains why certain apparent sources of information are not genuine news sources. Imagine a computer program that randomly churns out 'reports' on the Internet. Even though these reports might be hard to distinguish from genuine news, and hence might be widely treated as news, this clearly isn't a source of news, since it is not even in the business of trying to convey accurate information (even if, from time to time, it succeeds in this respect regardless). (This example also illustrates the earlier point that, contra Gelfert, just because something might be hard to distinguish from a genuine news source, as the reports made by this computer might be, it doesn't follow that it is a genuine news source.) Indeed, notice how matters change if the computer program were designed to convey accurate information—if, for example, it were constructed in such a way that it provides a digest of reports from other reputable news outlets. Now we would consider it a genuine news source.

Note, too, that a news source on this view is defined in terms of the motivation to convey accurate information, rather than in outcome-oriented terms, such as whether it generates beliefs, accurate or otherwise, in the audience for that news. There are a number of reasons for this. For one thing, a news source doesn't cease

makes it news, and that is a recent phenomenon. Relatedly, news can be regarding the future, as when a newspaper reports scientific claims regarding the anticipated effects of global warming. But, again, what makes this news are the scientific claims, and they are recent. Finally, notice that strictly speaking on this proposal we should be indexing news to a specific time, in that what was previously genuine news will likely not be news in the future (at least assuming that nothing relevant changes in the interim). So old news is no longer a genuine form of news, but it once was (unlike other forms of information which were never forms of news). Since this complication doesn't concern us here, I will be setting it to one side in what follows.

to exist when the audience disappears, much less the intended audience, such that there is no one whose beliefs could be influenced by this news source. This is why it can make sense to discover the news output of a long-lost civilization (e.g., by unearthing scrolls and such like). But a more important reason is that understanding news sources in terms of what doxastic effect its activities have on its audience would generate some counterintuitive results. Suppose that one knows that one's audience will believe the opposite of what one tells them. Would that mean that as a news source one should endeavor to systematically feed one's audience opposing information, and hence misinform rather than inform them? In any case, one's audience might be utterly indifferent to the news source, and hence pay it no heed at all in their deliberations, but that wouldn't entail that it wasn't a genuine news source.

One way of putting this last point is that there can be genuine news sources which are nonetheless ineffective in terms of the outcomes they generate (e.g., no one hears it, or people hear it but ignore it, and so on). The point is that a news source doesn't cease to be genuine just because it is ineffective. Just as we can distinguish between effective and ineffective news sources, we can also distinguish between news sources that are epistemically good and ones that are epistemically bad. Indeed, the goodness of a news source will often be primarily understood epistemically, such that it has a sound epistemic pedigree.<sup>13</sup> In particular, its goodness resides in it having suitable epistemic properties, such as conveying information that has an adequate epistemic basis. This might mean, for example, that good journalistic practices have been followed—such as checking sources, corroborating information, seeking confirmation/disconfirmation of one's information, and so on—as many of these practices have an epistemic import.<sup>14</sup> Relatedly, it is not a coincidence that a good news source is often reliable, since it is designed to be such. News sources are devoted to conveying accurate information, and so when they are well-designed that's just what they will do. Indeed, they will generate not just reliable information, but information that, when believed by someone on the basis of this news source, will itself have a good epistemic pedigree. It is in this sense that a good news source tends to generate beliefs that will themselves have a positive epistemic standing.

Note that epistemically good news sources can sometimes generate reports that are false or misleading, including reports that are intentionally so (as when an errant reporter ignores all protocols and invents a story in order to meet a

<sup>13</sup> As should be clear, when I talk of 'good' news in this context, I am specifically concerned with the epistemic properties of the news source, and not whether the news it generates is to be otherwise welcomed. The same goes, *mutatis mutandis*, with 'bad' news. In what follows I will take this point as granted.

<sup>14</sup> Journalistic practices are not all aimed at enhancing the epistemic pedigree of the news source, of course, as they can also serve other ends, such as principles of fairness. For further discussion of journalistic practices from an epistemic point of view, see Jacquette (2010).

deadline). Since the epistemic goodness of the source relates to its structural epistemic properties and its general reliability, this is compatible with it sometimes not successfully meeting its epistemic goals. This is why it does not follow on this way of thinking about news that it is a success term—the reports made by a genuine news source can be false, and even intentionally so, and thus there can be in this sense false news (which, note, is not thereby fake news, but rather genuine news, albeit of an epistemically problematic variety).

Some news sources are not epistemically good in this sense, however. For example, a school newspaper might not incorporate any of the protocols that are characteristic of epistemically good news sources. It would follow that this news source would tend to deliver reports that are unreliable. But notice that a news source that is bad in this epistemic sense can still be a genuine news source. This is because it is still aiming to convey accurate information of the relevant kind; it is just that it isn't very good at doing so. Remember that we have defined news sources in terms of manifesting a certain kind of intentional activity, and hence one can instantiate this notion well or poorly. That one is not very good at archery, and so regularly misses the target, doesn't mean that what one is doing when one fires one's arrows at the target isn't archery. Of course, if one is consistently hopeless at achieving the relevant goal of that activity, even when to do so would be relatively straightforward, then that starts to call into question whether one is genuinely striving to achieve it in the first place. But the point remains that if we understand what makes something a news source, then it becomes clear that there can be epistemically bad news sources as much as there can be epistemically good news sources, just as there can be ineffective as well as effective news sources.<sup>15</sup>

### 3. Fake News Versus News

As should be plain from the foregoing, fake news is not to be understood as a kind of news, albeit of a problematic variety, as Gelfert proposes. It is, rather, not genuine news at all, since it doesn't meet the requirement of being aimed at conveying accurate information of the relevant kind. Instead, as the 'fake' tag indicates, it is masquerading as real news in order to spread misinformation. Note, too, that fake news is different from other forms of non-news, such as the rings in the tree, the computer program randomly churning out reports, or the satirical magazine (even supposing that these sources deliver new information about

<sup>15</sup> I think this is a point that is overlooked in some of the literature on fake news. For example, Fallis & Mathiesen (2019) and Pepp et al. (2019) both understand news in a way that incorporates good epistemic procedures into the news itself (via appeal to appropriate journalistic practices). But this is to conflate news with a good epistemic pedigree with news *simpliciter*, and thereby to fail to capture the fact that there can be news that has a poor epistemic pedigree but is no less news as a result.

recent events, which of course they usually don't). In the first case, there is no misinformation, let alone misinformation by design. In the second and third cases, even if there is systemic misinformation, there is no intent to misinform.

Moreover, notice that just as fake news is not a form of genuine news, so a fortiori it is not a form of epistemically bad news either. That is, it is not that fake news is genuine news that has epistemically problematic properties, which is the case with epistemically bad news. Even though both phenomena are epistemically problematic, they are distinct in that epistemically bad news is still a variety of news, in virtue of being in the business of trying to convey accurate information (albeit poorly). This point is particularly important to adjudicating whether some cases are fake news or merely epistemically problematic news. For example, we can imagine an agent who is in the grip of various conspiracy theories and hence regularly churning out false stories as news. It is crucial here, however, whether in doing so they are trying to convey accurate information. If they are, then their output isn't fake news, but rather genuine news that is simply epistemically deficient. It is still epistemically problematic, but the problem it poses is distinct from that raised by fake news.<sup>16</sup>

Another problematic feature of Gelfert's account is that although he doesn't demand that fake news involves an intent to mislead, he nonetheless advances a very strong thesis regarding the extent to which fake news must actually mislead. He writes that fake news "must *in fact* mislead a relevant audience" (Gelfert 2018, 103, italics in original), that it is important to fake news that "sufficiently large numbers of people are in fact taken in by it" (Gelfert 2018, 105), and that fake news must be "*objectively likely* to mislead its target audience" (Gelfert 2018, 108, italics in original).<sup>17</sup> These are demanding requirements, and they do not stand up to closer scrutiny. The crux of the matter is that just as we need a distinction between effective news and ineffective news, such that the latter can nonetheless be genuinely news, so we need to distinguish between effective and ineffective fake news, where the latter is nonetheless a bona fide form of fake news. In particular, we need a way of distinguishing between kinds of fake news that are constructed such that they are effective at attaining their epistemic objective of misleading people as opposed to kinds of fake news that are constructed such that they are ineffective at attaining these epistemic objectives.

<sup>16</sup> For a contrasting view, see Pepp et al. (2019) who claim that such a case would be an instance of fake news. As I've just explained, I think this is to confuse fake news with genuine news that is epistemically problematic. As noted in note 15, part of the reason for our divergence is that Pepp et al. (2019) build far more into a news source than I do, as they identify news sources with those that manifest good journalistic practices. But I think this is to equate news sources with good news sources (where the 'goodness' in play is primarily epistemic).

<sup>17</sup> See also Grundmann (2019) and Bernecker (chapter 13 in this volume), who both define fake news in terms of the effects it has on its audience, such that fake news needs to be effective in misleading its intended audience.

Consider, for example, a website that is designed to spread misinformation by embedding false stories within accurate ones, perhaps for propaganda purposes. This would be a paradigmatic case of fake news. Nonetheless, this can be done in an effective way, or in a lousy way. In the former case, this would involve carefully curating the false reports within the accurate information in order to make it more compelling, phrasing the false reports judiciously to ensure that they look as plausible as possible, offering ready-made responses to natural objections that the reader might have to the content of these reports, and so on. The upshot would be to provide an epistemic setting that makes the fake news more effective, in virtue of it being harder to distinguish from genuine news. In the latter case, in contrast, one could imagine the very same website conducting its business in a clumsy way, such that it will be relatively easy for an informed person to spot that the news is fake. In particular, the epistemic stage-setting will be lacking, and hence the alarm bells will sound for most people that this website is not to be trusted.

The point of the matter is that poorly executed fake news is still fake news. In particular, the fact that the lousy fake news website might not in fact mislead anyone, and certainly wouldn't be such that it is 'objectively likely' that it would mislead its intended audience, doesn't entail that it isn't fake news. *Modulo* our earlier point about decoy ducks and real ducks; just as an excellent forgery of a Picasso painting does not become the genuine article just because it is so good as to be hard to distinguish from the real thing, so it doesn't follow that a lousy forgery of a Picasso painting is any less of a forgery for the fact that it is easily spotted. In short, ineffective fake news is still fake news.

More generally, just as we should avoid restrictive accounts of fake news that demand that fake news always be effective, so we should similarly reject proposals that build specific goals or purposes into the account of fake news, such that fake news is identified with a particular form of fake news. Consider, for example, this account of fake news offered by Scott Aikin and Robert Talisse:

Fake news characterizes the activities of institutions that *pose* as journalistic which by design feed and codify the *antecedent* biases of a pre-selected audience by exploiting their vulnerabilities (cognitive and otherwise), all with a view towards facilitating some decidedly political objective.

(Aikin & Talisse 2018, italics in original)

While the proposal captures the point that fake news is not a kind of genuine news (but merely masquerades as such), it offers a much too narrow conception of what fake news involves. On this account, in order for someone to create fake news they must already have a sophisticated conception of cognitive biases and how to effectively exploit them. They must also have political objectives in doing this. But neither seems essential to fake news. On the former front, while it might be a

feature of (a certain form of) *effective* fake news that it is designed to exploit cognitive biases in this way, we have already noted that there can *ineffective* forms of fake news which are no less genuine varieties. Moreover, while many forms of fake news surely do have a political agenda, it's hard to see why this should be an essential feature, as fake news could serve many purposes (including merely being, say, for the amusement of the one propagating it).

Putting all these points together, let's review where we are with our account of fake news. We've argued that it has the following properties. First, unlike genuine news (which is aimed at conveying accurate information), fake news deliberately conveys misleading information. This doesn't mean that fake news necessarily involves the presentation of falsehoods, given that even the literal truth can be misleading (indeed, as we've noted, sometimes this is the most effective way to mislead). Nonetheless, second, fake news is presented as news, just as a forgery of a painting is presented as the real thing. Relatedly, third, fake news involves an intent to mislead, which is why, for example, satirical news magazines are not fake news (even if they in fact do mislead). That said, fourth, fake news need not in fact mislead. While effective fake news typically misleads, ineffective fake news may not be successful on this front, but it is no less a form of fake news as a result (ineffective fake news is still fake news). Finally, fifth, fake news is not itself a type of news, even though it may be hard to distinguish between the two (any more than an excellent forgery is thereby the genuine article). In particular, fake news is to be distinguished from a genuine form of news that has a poor epistemic pedigree.

#### 4. Managing Fake News: Individual

One question we might ask at this juncture is what hangs on our thinking about fake news in just this fashion. I think we can recognize the importance of understanding fake news correctly once we appreciate the kind of expertise needed to spot it. In particular, since fake news is to be distinguished from genuine news, even genuine news that has a poor epistemic pedigree, what is required to spot fake news won't be quite the same as what is required to assess the epistemic credentials of genuine news (though there will be quite a lot of overlap, as we will see).

This ought not to be surprising. Consider someone who works in a treasury whose job is to ensure the quality control of the currency that is being produced. This will require very specific skills to spot the kinds of considerations that are relevant to currency quality, where this might include such things as printing issues, degradation of the (e.g.,) polymer on which the note is printed, errors in what is printed, and so on. Naturally, someone working in this department will also be interested in how the notes were produced, particularly where the quality is lacking, so that they can ensure that future quality is preserved.

Someone with these skills would likely be well-placed to spot a counterfeit note, not least due to their close familiarity with genuine currency, but it remains that the skills involved in spotting counterfeit currency are not quite the same. After all, our treasury currency controller is not dealing with fakes, but rather trying to spot genuine currency that is merely deficient in some way. In particular, there is no element of deceit in play in this quality control process, but merely a sifting of the wheat from the chaff.

In contrast, we can imagine someone who works at the same treasury whose specific job it is to spot counterfeit currency. While this person will have similar expertise to her colleague in quality control, there will clearly be differences to how she goes about her role. The counterfeit currency might not have the kinds of errors or deficiencies that are found in low-quality currency; indeed, one would expect the very best counterfeit currency to be very similar to genuine currency, at least superficially. Our expert in the counterfeit currency department will thus be on the look-out for very specific indications that a counterfeit is in play (an unexpected barcode, a watermark that is just a little too large or not quite in the right place, and so on). Moreover, those working in the counterfeit department of the treasury won't be looking just at the notes, but will also be interested in other information regarding the provenance of the notes—indeed, one would expect that it is information of just this kind that has led to these notes being brought to this department for inspection.

What goes for currency goes for news, or so I claim. While there will be obvious overlap in the kind of skills and knowledge that is required to distinguish epistemically good genuine news from epistemically deficient genuine news, and fake news from genuine news, there is nonetheless a distinct set of skills and knowledge in play in each case. We need to attend to this fact if we wish to train people to spot fake news.

We can capture what is at issue here by comparing what is required to evaluate the epistemic credentials of testimony in a context in which testifiers are generally honest (but might not thereby always have a good epistemic basis for what they assert, much less always assert the truth), and what is required to evaluate the epistemic credentials of testimony in a context where there are bad actors in play who have an intent to deceive (note that in keeping with our account of fake news, this might not involve lying as such). This would be roughly comparable to our quality control person and our counterfeit spotter at the treasury described above, except now the 'currency' under evaluation is testimony. There is one interesting difference between the two cases, however. The reason why people forge currency is very straightforward, as they are obviously seeking financial gain. The reasons why people use testimony to deceive, however, just as the reasons why people propagate fake news, are numerous, and might not be at all straightforward. This is important, since it gives the person evaluating the testimony more to go on when it comes to undertaking their evaluation.

In a testimonial context where one is assured that the testifiers are sincere, one needs only to consider such factors as the reliability of the testifier and the credibility of what they are claiming. In contrast, in testimonial contexts where one lacks such an assurance, then one has to in addition consider the motivations these testifiers might have for asserting what they do. In particular, one needs to consider whether the testifiers have a motivation to mislead by offering this particular testimony. The epistemic burdens on those assessing the testimony are thus higher, and what is required to weed out good testimony from bad is correspondingly more demanding.

What goes here for testimony also applies to the task of identifying fake news. This should be unsurprising, given that being a recipient of (putative) news is itself a kind of testimonial context (albeit one that doesn't typically involve any direct interaction between the testifier and the recipient of the testimony). In particular, if one is on the lookout for fake news, then one needs to take into account the specific motivations someone might have for propagating misleading information. This is different from the case where one is merely distinguishing between genuine news that enjoys a good epistemic pedigree and genuine news that doesn't. In that case, one's concern is entirely with the epistemic source of the news (e.g., does it arise from a reliable news source?), and with the epistemic standing of the news itself (e.g., is it credible, given what else one knows?). Since the sincerity of those offering this news is not in question, one doesn't need to worry about their motivations. When one is in addition on the look-out for fake news, however, then motivations become very salient. In particular, one has to further consider such factors as whether this is 'news' that someone might have a motivation to put out even though it is misleading. For example, where the 'news' benefits a particular political party during an election campaign, then one should be willing to expose it to additional scrutiny, since this fact raises the likelihood that it might be fake news.

We saw earlier that there will be considerable overlap in the expertise required to, on the one hand, differentiate good quality currency from poor quality currency, and, on the other hand, differentiate genuine currency from counterfeit currency. The same will be true in the case of genuine and fake news, especially when it comes to epistemically deficient genuine news and ineffective fake news. For example, a genuine news source might, through error, end up reporting something clearly false, and which an alert recipient of this news can easily spot is false. Similarly, obvious falsehoods might be propagated as fake news, and easily spotted in much the same way.

The interesting contrast, however, will involve the hard cases of differentiating genuine news from effective fake news. This is where the special expertise relevant to spotting fake news—just as in the case of spotting counterfeit currency—becomes important. Effective fake news, after all, will be by its nature hard to distinguish from the genuine article. In particular, it might not be obvious either

that the source of the putative news is epistemically problematic or that what it claims isn't credible. This is where attending to possible motives for putting forward these claims, and relatedly being aware of the wider context of the putative news (political, social, commercial, etc.), becomes important.

We can see some of these points in action by considering some recent empirical work on how people are taken in by fake news. As researchers have noted, the problem is that recipients of the fake news tend to evaluate the putative news source 'vertically' rather than 'horizontally,' where this means that they look for corroborating evidence (or otherwise) for what they are reading from *within* the article itself rather than verifying the putative news source by appeal to independent, and already verified, sources of news.<sup>18</sup> Accordingly, they find themselves becoming more convinced of the fake news the more they read of it, when in fact they ought to be far more suspicious of what they are reading (and would be, had they examined relevant independent, and already corroborated, sources of news).

The mistake our recipients of fake news are making is to fail to realize that they are not in a testimonial context where they can reasonably take it as given that the testimony they are receiving is honest and sincere. In particular, they are failing to realize that they cannot assume that what seems like genuine news is genuine news. Were they to be in such an epistemically friendly testimonial context, then a vertical epistemic evaluation would be far more plausible as a system of epistemic appraisal, just as one might epistemically evaluate someone's testimony in such a testimonial context by considering how credible it is that the claims they are making are true. (Though as we noted above, even then one should also be willing to consider the epistemic standing of the source of the testimony, as even honest and sincere testifiers can be unreliable, particularly about certain subject matter; being responsive to the credibility of what is asserted would be a natural first step, however.) In testimonial contexts where one cannot take it as given that the testimonial actors are honest and sincere, however, then a further layer of scrutiny is required, and that's where the need for a horizontal epistemic evaluation becomes pressing, for how else is one to satisfy oneself that this apparently genuine news is the real deal, and not its fake counterpart? The challenge is thus to have the cognitive skills to know when it is appropriate to adopt the additional levels of scrutiny, and a big part of that, as we noted above, is being sensitive to the possible motives that someone might have to make the target 'news' claims in play.

With all these points in mind, what kinds of cognitive traits should we promote in subjects if we want them to be able to spot fake news? I think the answer lies in

<sup>18</sup> See especially Wineburg & McGrew (2017). Their paper is part of a rich body of recent work on how people evaluate digital news sources that has been produced by the Stanford History Education Group. For some other representative publications arising out of this research, see McGrew et al. (2017; 2018; 2019) and McGrew (2019). I am grateful to Gabe Avakian Orona for alerting me to this work.

the *intellectual virtues*, where by this I mean those distinctive admirable character traits that are guided by a love of the truth, and which are essential components of a life of flourishing, such as intellectual humility, intellectual conscientiousness, honesty, and so on.<sup>19</sup> There is no infallible guide to spotting fake news, but as fallible guides go I believe the intellectual virtues have the most promise.<sup>20</sup>

To begin with, if we are to counter the threats to truthfulness posed by fake news, then it is essential that we foster character traits that involve a love and desire for the truth, as lack of concern for the truth is precisely what is generating the problem in hand. It is thus significant that the intellectual virtues essentially incorporate such a desire for truth as a motivational component. Relatedly, it is also important to the intellectual virtues that these are not cognitive traits that are innate, or of a kind that can be easily acquired and maintained. The intellectual virtues require *cultivation*, and it is the process of such cultivation, given the truth-directed nature of these character traits, that keeps the value of truth and its possession as a target in mind. The proponent of fake news might not care about the truth, but the recipient of it needs to have this epistemic good in her sights.

Even more importantly for our purposes, however, a key facet of the intellectual virtues is the manner in which they are highly context-sensitive in their application. That is, it is part of the very nature of an intellectual virtue that it involves a sensitivity on the part of the subject to relevant circumstances. Indeed, this much follows from the fact that the manifestation of the intellectual virtues (in common with the virtues more generally) lies on a ‘golden mean’ between two vices, one of excess and one of deficiency. So, for example, the intellectual virtue of being intellectually conscientious lies between the vice of excess of pedantically attending to every detail, no matter how irrelevant, and the vice of deficiency of being intellectually unconscientious (e.g., being completely unconcerned about the relevant evidence). To have an intellectual virtue is thus to have the good judgement to know when to manifest this cognitive trait, such that one is neither deficient nor excessive in the relevant respects.

This is significant when it comes to fake news—just as it is relevant to the epistemic assessment of testimony more generally—in that it means that the intellectually virtuous subject will be sensitive to those factors that demand further

<sup>19</sup> I take a broadly Aristotelian line on the intellectual virtues, as my subsequent discussion will reveal. For a key contemporary defense of such an account of the intellectual virtues, see Zagzebski (1996). See also Baehr (2011). For a more general overview of the contemporary literature on intellectual virtue, see Battaly (2014).

<sup>20</sup> At least at the individual level anyway—the reason for this qualification will soon become apparent. Incidentally, elsewhere I have argued at length for the claim that the overarching epistemic goal of education is the cultivation of the intellectual virtues in just this sense—see, especially, Pritchard (2013; 2016; 2018b). See also Baehr (2013) for a proposal in a similar vein. If that’s right, and if it is also correct that the intellectual virtues offer the best way for us to identify fake news (at least at the individual level), then this educational goal will also serve this more specific purpose. For a recent critical discussion of this kind of account of the epistemology of education, see Siegel (2016; 2018) and Carter et al. (2019). See Baehr (2019) for a response to this critical line.

inquiry. Whereas it might often be appropriate to evaluate putative news sources on the assumption that they are genuine (in that subjecting them to additional scrutiny would be excessive), and hence make broadly vertical epistemic assessments of them, the intellectually virtuous subject will also be sensitive to considerations that would demand that one should raise one's level of scrutiny. In particular, the intellectually virtuous subject will be responsive to those factors that would entail that this assumption about putative news sources being genuine should be dropped (such as the motivational considerations that underlie fake news noted above), and hence that a horizontal epistemic assessment would be appropriate instead.<sup>21</sup>

It is also relevant in this regard that the intellectual virtues are by their nature integrated with one another, and more generally with the other (non-intellectual) virtues, such as the moral virtues, with both sets of virtues governed by an overarching virtue of practical wisdom (*phronesis*). This feature of intellectual virtues, coupled with the fact that they are manifested on a golden mean between two corresponding vices of excess and deficiency, ensures that the virtuous agent responds to the threat posed by fake news in a measured manner. In particular, there is a contemporary tendency to deal with this threat by becoming overly skeptical about news sources more generally, and in the process dismissing much that has epistemic merit. Indeed, in the more extreme case this can lead people to be dismissive even of scientific reporting from reputable sources. In virtue-theoretic terms, this is to succumb to a vice of excess, in contrast to a more proportionate, and intellectually virtuous, questioning of sources in light of the presence of fake news in one's epistemic environment.<sup>22</sup>

Moreover, notice that such widespread (and non-virtuous) skepticism is an intellectual strategy that has become detached from one's wider practical and moral goals, something that the integrated nature of the virtues is designed to prevent. In particular, to doubt in this fashion is to cut oneself off from epistemic goods and thereby lose the practical and moral advantages that the possession of these goods can provide. In the case of one's practical goals, the import of such epistemic goods ought to be straightforward. How is one to achieve one's practical goals if one is suspicious of news sources more generally, given how such sources can be vital to informing one's decision-making? I think the relevance of such

<sup>21</sup> This is one reason why I am skeptical about educational attempts to deal with fake news that simply focus on providing students with various kinds of technical expertise, such as knowledge of statistics and how they can be misused. While such technical expertise is undoubtedly useful in this regard, the point is that one also needs to cultivate the good judgement in the students to employ this expertise appropriately, and for that the intellectual virtues are required. This recent national schools project in Finland, as reported in *The Guardian* newspaper, is interesting in this regard: <https://www.theguardian.com/world/2020/jan/28/fact-from-fiction-finlands-new-lessons-in-combating-fake-news>. At least on the surface, the project seems to be primarily about the development of technical expertise, but if one looks closer at how this project is described it begins to look far more about the more general cultivation of intellectual character, and thus intellectual virtue.

<sup>22</sup> I explore this theme in more detail in Pritchard (2019).

epistemic goods to one's moral goals is also relatively clear. My claim here, in common with virtue theoretic accounts more generally, is that knowledge is required for one to be morally virtuous, in that relevant knowledge (e.g., about one's circumstances) is needed to guide the manifestation of the moral virtues. It follows that to deprive oneself of knowledge via such widespread skepticism can prevent one from manifesting these admirable character traits. Given the importance of the virtues to a life of flourishing, to undermine one's manifestation of the moral virtues in this way is effectively a form of self-harm. In any case, possession of the virtues, including the intellectual virtues, acts as an important barrier towards taking such a route.

## 5. Managing Fake News: Structural

In the last section we focused on what consequences our account of fake news has for the individual who needs to be able to spot it. But clearly it cannot only be exclusively the job of the individual to manage fake news; there also needs to be a structural response to this wider social problem. Aside from anything else, even the most intellectually virtuous individual might lack the capacity to differentiate genuine news from its counterfeit counterpart; as we noted, while the intellectual virtues are plausibly the best route to making such a differentiation (at the individual level anyway), they are far from being infallible guides. This is why we need to assist the individual by in addition seeking structural reforms to our epistemic environments that help us to identify fake news.

Our account of fake news is helpful in this respect. This is because one of the challenges facing those who seek structural responses to the problem of fake news is the danger that this leads to putting undue constraints on a free press, something that is widely held to be a cornerstone of a well-functioning democratic society. But notice that this concern betrays a way of thinking about fake news that regards it as being continuous with genuine news, along the lines that we saw Gelfert setting out above. It is only on this supposition that any structural constraints on fake news would *thereby* be an infringement on genuine news. In contrast, insofar as we treat fake news as distinct from genuine news, as we have been urging, then the conceptual space is cleared to allow us to regulate the former without thereby undermining the latter.

Of course, there will undoubtedly be practical hurdles to clear in this regard, as any policy of containing fake news will potentially have a bearing on the practices of those putting forward genuine news, especially since it can be hard, as we have seen, to differentiate effective forms of fake news from the real thing. But such practical hurdles are far from unsurmountable, and at most all they entail is that we should proceed with caution. So, for example, it might be necessary to set the

bar for fake news, of a kind that is to be regulated at a structural level anyway, quite high, such that it is only clear-cut cases of fake news that get curtailed.

It is also obviously relevant here what this ‘curtailment’ involves and who is doing it. In the former case, if all that is happening is that the ‘news’ comes with a warning to the recipient that it might be fake news, then that hardly seems to be infringing on one’s democratic freedoms at all, particularly *modulo* the previous point that we are only picking out clear-cut cases of the phenomenon. In contrast, if ‘curtailment’ means actual removal of the fake news, then that will obviously be more problematic in this regard. This point dovetails with the issue of who is doing the censoring. Where this is itself democratic institutions that are open to challenge from its citizens, then again it will be harder to make the case that this is an infringement of one’s democratic freedoms. In contrast, if this censorship is done by non-governmental bodies, such as corporations (e.g., Facebook, Twitter), especially if this is not subject to outside scrutiny or governmental regulation, then I think we will rightly feel somewhat nervous about how this might lead to a wider constraint of genuine journalistic reporting.

It is not my goal here to navigate through these difficult practical waters, but merely to note that it makes a crucial difference to these debates whether one holds that fake news is itself a kind of genuine news, albeit of a problematic epistemic kind. Once one departs from this supposition—and I have argued that we should do so—then it becomes possible to identify ways of dealing with fake news that need not thereby impose a restriction on genuine news (including genuine news that has a problematic epistemic pedigree). Moreover, although the task of differentiating genuine news from fake news at the individual level may be arduous (even if one grants the subject the intellectual virtues that I have described above), the combination of the intellectually virtuous subject working in concert with a structural approach to fake news should make the task of identifying fake news far more tractable.<sup>23</sup>

## References

Aikin, S. F. & Talisse, R. (2018). ‘On “Fake News”’, 3 *Quarks Daily*, <https://www.3quarksdaily.com/3quarksdaily/2018/05/on-fake-news.html>.

Battaly, H. (2014). ‘Intellectual Virtues’, in S. van Hooft (ed.), *Handbook of Virtue Ethics*, 177–87, London: Acumen.

<sup>23</sup> Thanks to Sven Bernecker, Michel Croce, Thomas Grundman, and Gabe Avakian Orona. Material from this chapter was presented at the ‘Archaeology in a Post-Truth World’ conference, held at UC Irvine in February 2020. I am also grateful to two non-anonymous reviewers for this volume for their detailed comments on an earlier version of this chapter.

Baehr, J. (2011). *The Inquiring Mind: On Intellectual Virtues and Virtue Epistemology*, Oxford: Oxford University Press.

Baehr, J. (2013). 'Educating for Intellectual Virtues: From Theory to Practice', *Journal of Philosophy of Education* 47, 248–62.

Baehr, J. (2019). 'Intellectual Virtues, Critical Thinking, and the Aims of Education', in M. Fricker, P. Graham, D. Henderson, & N. J. L. L. Pedersen (eds.), *Routledge Handbook of Social Epistemology*, ch. 43, London: Routledge.

Bernecker, S. (This Volume). 'An Epistemic Defense of News Abstinence', in S. Bernecker, A. Flowerre, & T. Grundman (eds.), *Epistemology of Fake News*, Oxford: Oxford University Press.

Carter, J. A., Kotzee, B., & Siegel, H. (2019). 'Educating for Intellectual Virtue: A Critique from Action Guidance', *Episteme*, <https://doi.org/10.1017/epi.2019.10>.

Cassam, Q. (2018). 'Epistemic Insouciance', *Journal of Philosophical Research* 43, 1–20.

Dretske, F. (1981). *Knowledge and the Flow of Information*, Cambridge, MA: MIT Press.

Fallis, D. & Mathiesen, K. (2019). 'Fake News is Counterfeit News', *Inquiry*, <https://doi.org/10.1080/0020174X.2019.1688179>.

Frankfurt, H. G. (2005). *On Bullshit*, Princeton, NJ: Princeton University Press.

Gelfert, A. (2018). 'Fake News: A Definition', *Informal Logic* 38, 84–117.

Grundmann, T. (2019). 'Fake News: The Case for a Consumer-Oriented Explication', unpublished ms.

Habgood-Coote, J. (2018). 'Stop Talking about Fake News!', *Inquiry*, <https://doi.org/10.1080/0020174X.2018.1508363>.

Jacquette, D. (2010). 'Journalism Ethics as Truth-Telling in the Public Interest', in S. Allen, (ed.), *Routledge Companion to News and Journalism*, 213–22, London: Routledge.

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). 'The Science of Fake News', *Science* 359, 1094–6.

Levy, N. (2017). 'The Bad News About Fake News', *Social Epistemology Review and Reply Collective* 6, 20–36.

McGrew, S. (2019). 'Learning to Evaluate: An Intervention in Civic Online Reasoning', *Computers & Education*, <https://doi.org/10.1016/j.comedu.2019.103711>.

McGrew, S., Ortega, T., Breakstone, J., & Wineburg, S. (2017). 'The Challenge that's Bigger than Fake News: Civic Reasoning in a Social Media Environment', *American Educator* 41, 4–11.

McGrew, S., Breakstone, J., Ortega, T., Smith, M., & Wineburg, S. (2018). 'Can Students Evaluate Online Sources? Learning from Assessments of Civic Online Reasoning', *Theory & Research in Social Education* 46, 165–93.

McGrew, S., Breakstone, J., Ortega, T., Smith, M., & Wineburg, S. (2019). 'Improving University Students' Web Savvy: An Intervention Study', *British Journal of Educational Psychology* 89, 485–500.

MacKenzie, A. & Bhatt, I. (2018). 'Lies, Bullshit and Fake News: Some Epistemological Concerns', *Postdigital Science and Education*, <https://doi.org/10.1007/s42438-018-0025-4>.

Mukerji, N. (2018). 'What is Fake News?', *Ergo* 5, 923–46.

Pepp, J., Michaelson, E., & Sterken, R. K. (2019). 'What's New About Fake News?', *Journal of Ethics and Social Philosophy* 16, <https://doi.org/10.26556/esp.v16i2.629>.

Pritchard, D. H. (2013). 'Epistemic Virtue and the Epistemology of Education', *Journal of Philosophy of Education* 47, 236–47.

Pritchard, D. H. (2016). 'Intellectual Virtue, Extended Cognition, and the Epistemology of Education', in J. Baehr (ed.), *Intellectual Virtues and Education: Essays in Applied Virtue Epistemology*, 113–27, London: Routledge.

Pritchard, D. H. (2018a). 'Disagreement, of Belief and Otherwise', in C. Johnson (ed.), *Voicing Dissent: The Ethics and Epistemology of Making Disagreement Public*, 22–39, London: Routledge.

Pritchard, D. H. (2018b). 'Neuromedia and the Epistemology of Education', *Metaphilosophy* 49, 328–49.

Pritchard, D. H. (2019). *Scepticism: A Very Short Introduction*, Oxford: Oxford University Press.

Rini, R. (2017). 'Fake News and Partisan Epistemology', *Kennedy Institute of Ethics Journal* 27, 43–64.

Rose, J. (2019). 'To Believe or Not to Believe: An Epistemic Exploration of Fake News, Truth, and the Limits of Knowing', *Postdigital Science and Education*, <https://doi.org/10.1007/s42438-019-00068-5>.

Siegel, H. (2016). 'Critical Thinking and the Intellectual Virtues', in J. Baehr (ed.), *Intellectual Virtues and Education: Essays in Applied Virtue Epistemology*, *Intellectual Virtues and Education*, 95–112, London: Routledge.

Siegel, H. (2018). *Education's Epistemology: Rationality, Diversity, and Critical Thinking*, Oxford: Oxford University Press.

Twardowski, K. (1979). 'Issues in the Logic of Adjectives', in J. Pelc (ed.), *Semiotics in Poland 1984–1969*, 28–30, Dordrecht: Springer.

Wineburg, S. & McGrew, S. (2017). 'Lateral Reading: Reading Less and Learning More When Evaluating Digital Information', *Stanford History Education Group Working Paper No. 2017-A1*, <https://ssrn.com/abstract=3048994>.

Zagzebski, L. T. (1996). *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*, Cambridge: Cambridge University Press.

# 3

## The Fake News about Fake News

*David Coady*

### 1. Introduction

Conventional wisdom holds that the world is facing a new, unprecedented, or at any rate growing, problem that goes by the name ‘fake news’. A recent survey by Gallup/Knight Foundation, for example, found that the majority of Americans consider fake news to be a very serious threat to democracy.<sup>1</sup> Conventional wisdom on this matter is supported by a growing body of academic literature which seeks to define ‘fake news’ and identify the nature of the problem it allegedly presents. I argue that this is all misguided. Conventional wisdom is right that we are confronted with a new problem, and that it is a threat to democracy, but the problem is not fake news; the problem is the term ‘fake news’. I will be offering no definition of this term myself, because I don’t think there is a correct definition of it. My position is that we should refrain from using the term ‘fake news’ altogether.<sup>2</sup> My arguments for this position, which will be partly epistemological and partly ethical, amount to this: the term serves no good purpose, while doing considerable harm.

The current epistemic panic over the putative phenomenon of fake news is very recent. Although the term itself seems to have been around for a while,<sup>3</sup> it only came to be widely used in the aftermath of the 2016 American presidential election. It was the *Collins Dictionary* word of the year in 2017 and though, strictly speaking, it is not a word (it is two words), it will be convenient for me to follow *Collins*’s lead by occasionally calling it a word in this chapter.

Although Donald Trump has claimed that he coined the word,<sup>4</sup> it in fact first gained traction amongst opponents of Trump, as a way of referring to certain pro-Trump news-sites originating in Macedonia. Since then, Trump has used the term repeatedly to refer to virtually any claim that he disagrees with. Many of his critics

<sup>1</sup> See <https://knightfoundation.org/reports/american-views-trust-media-and-democracy> (accessed 31/10/2019).

<sup>2</sup> This position (which I call ‘eliminativism’ about fake news) is also defended by Habgood-Coote (2019).

<sup>3</sup> It was used in an article in *Harper’s Magazine* as early as 1925: <https://harpers.org/archive/1925/10/fake-news-and-the-public/> (accessed 30/10/2019).

<sup>4</sup> See <https://www.youtube.com/watch?v=xN88-pb2dFo&feature=youtu.be> (accessed 30/10/2019).

take exception, not to the term itself, but to his alleged misuse of it, implying that the term has a legitimate meaning which he has distorted. This is sometimes followed by an attempt to specify what that meaning is. Although there is a widespread consensus that Trump has misunderstood the term or is intentionally misusing it, there is no consensus at all about how it should be understood or used. If I am right, the emergence of a new word has been mistaken for the emergence of a new phenomenon: the supposed referent of that word.<sup>5</sup>

The term ‘fake news’ is objectionable, not merely because it is ambiguous; after all much of the language we use is ambiguous. Rather, it is objectionable because although it has no fixed meaning, it does serve a fixed function and this function is objectionable. It can be compared to the somewhat older term, ‘conspiracy theory’ (another term without a fixed meaning), and both terms can be compared with the far older word ‘heresy’. All three terms have functioned to narrow the range of acceptable opinion and restrict the terms of acceptable debate. They are, in effect, policing devices for the enforcement of orthodoxies. They herd opinion, or at least, ‘respectable opinion’, in ways that conform to the agendas of powerful people and institutions. The label ‘heresy’ was once applied to claims that challenged the power of the Church. In a similar way, the label ‘fake news’ is now used of claims that challenge the most powerful institutions of our age: nation states and international corporations.

Before elaborating on the ways in which this has happened, I want to anticipate an objection to my approach, which goes as follows: it’s all very well to point out various ways in which the term ‘fake news’ has been put to pernicious ends, but this doesn’t show that we should abandon the term. Rather, it shows that we need to find an unambiguous, precise, and rationally justifiable definition of the term. In support of this position, the objector might point to terms, such as ‘terrorism’ and ‘propaganda’, both of which have been used to advance objectionable causes, but which arguably can also be used in legitimate ways. The fact that these words have a history of being abused, the objection goes, is no reason for abandoning them. Indeed, the fact that a word has been abused entails that there is a correct way to use it, and that we should rise to the challenge of trying to identify it. This is essentially the line that a prominent group of more than a dozen social scientists and legal scholars took recently in a letter published in the prestigious scientific journal *Science* entitled “The Science of Fake News” (Lazer et al. 2018), in which they claim that the term ‘fake news’ should be retained, despite Trump’s alleged misuse of it. According to them, “we can’t shy away from phrases because they’ve somehow been weaponized. We have to stick to our guns and say there is a real phenomenon here”; they go on to call upon their profession to “help fix democracy by studying the crisis of fake news”.

<sup>5</sup> This is not the first time this has happened. The emergence of the term ‘conspiracy theory’ in the 1960s and the more recent emergence of ‘post-truth’ follow a similar pattern.

I respond to this objection in four ways. First, I'm not convinced that either of the words mentioned above (i.e. 'terrorism' and 'propaganda') have any legitimate use, though it would require too much of a digression to pursue those matters here. Second, because the term 'fake news' has only been in popular use for a short time and has not yet become deeply entrenched, the goal of consigning it to the dustbins of history is much more realistic. Third, none of the existing definitions that I am aware of are satisfactory. And fourth, as I hope to make clear, the unsatisfactory nature of extant definitions is not mere happenstance; there are criteria that any adequate definition would have to meet, which there is good reason to believe no definition of 'fake news' can meet.

In what follows, I will give an all too brief account of some of the ways in which governments and international technology/media corporations have used the term 'fake news' as an instrument of censorship. I will then turn to the ways it has been deployed by academics to legitimate such censorship, and to pathologize views that they disagree with.

## 2. The Term 'Fake News' as an Instrument of State Censorship

Elsewhere I have described the ways in which the governments of numerous countries, including Brazil, Malaysia, Uganda, Tanzania, Kenya, India, Germany, France, the United Kingdom, and Australia have introduced, or are in the process of trying to introduce, censorship regimes justified by the alleged problem of fake news (Coady 2019). The penalties for producing or distributing material that these governments deem to be 'fake news' range from the relatively minor, for example having one's web page taken down, to the draconian, for example being sentenced to lengthy prison terms. Since then, Russia has joined the ranks of countries which have introduced legislation to 'protect' people from 'fake news'. People who fall foul of the Russian law, which defines 'fake news' in such a way that it includes showing "blatant disrespect for the state", face hefty fines.<sup>6</sup> Thailand has started prosecuting people for distributing (or even 'liking' on social media) information that its government deems to be 'fake news'.<sup>7</sup> Singapore has just implemented its own 'fake news law'. People who are judged to have distributed fake news in a way that is "malicious and damaging to Singapore's interests" face up to ten years in prison.<sup>8</sup> Finally, the European Union is in the process of setting up a regulatory

<sup>6</sup> See <https://www.reuters.com/article/us-russia-politics-fakenews/russias-putin-signs-law-banning-fake-news-insulting-the-state-online-idUSKCN1QZ1TZ> (accessed 30/10/2019).

<sup>7</sup> See <https://www.scmp.com/news/asia/southeast-asia/article/2150811/thai-government-steps-efforts-crack-down-fake-news> (accessed 30/10/2019).

<sup>8</sup> See <https://www.aljazeera.com/news/2019/10/singapore-controversial-fake-news-law-effect-191002055650351.html> (accessed 30/10/2019).

framework to prevent citizens from being “exposed” to information it deems to be ‘fake news’.<sup>9</sup>

Unfortunately, we live in increasingly authoritarian and brutal times, and we can no longer assume that everyone will be opposed to state censorship, just as we can no longer assume that everyone will be opposed to state torture. We have learnt to accept state censorship in the West, as we have learnt to accept state torture, because it has been veiled behind euphemisms. Just as ‘they’ torture while ‘we’ only engage in ‘enhanced interrogation’, ‘they’ censor while ‘we’ only ‘regulate the Internet’. For those who don’t object to ‘internet regulation’, so long as it is carried out by ‘good governments’, I have little to say except to remind them of the long history of even the best intentioned censorship regimes leading to harms their authors did not foresee or want, and more specifically of the many cases of people becoming victims of their own censorship legislation. For example, Heiko Mass, Germany’s former justice minister and the man behind the *Netzwerkdurchsetzungsgesetz* (Network Enforcement Act), the first European attempt to censor ‘fake news’, had one of his own tweets deleted as a result of this law.<sup>10</sup> A more alarming example of this phenomenon occurred in Malaysia, where a ‘fake news law’ was introduced to prevent Mahathir Mohamad from becoming prime minister. When, despite this law, he did become prime minister, he promptly announced that he would not be honouring his pledge to repeal the law. Since then he has been using the law introduced by his political opponents to jail his political opponents.<sup>11</sup> Governments do not last forever and even if you are unconcerned by the new powers the French government has just handed President Macron to censor so-called ‘fake news’ on the Internet during French election campaigns,<sup>12</sup> you need to ask yourself how you would feel about Marine Le Pen inheriting those powers.

### 3. The Term ‘Fake News’ as an Instrument of Corporate Censorship

It is not only governments that have moved to control free speech and publication in the name of stopping the alleged problem of fake news. Major corporations, especially media/technology giants have also exploited this emerging epistemic panic, to advance extremely questionable agendas. I have discussed this topic in

<sup>9</sup> See <https://ec.europa.eu/digital-single-market/en/tackling-online-disinformation> (accessed 30/10/2019).

<sup>10</sup> See <https://www.thelocal.de/20180108/justice-minister-falls-victim-to-own-social-media-censorship-law> (accessed 31/10/2019).

<sup>11</sup> See <https://www.cnet.com/news/malaysias-fake-news-law-is-here-to-stay-new-prime-minister-says/> (accessed 31/10/2019).

<sup>12</sup> See <https://www.euronews.com/2018/11/22/france-passes-controversial-fake-news-law> (accessed 31/10/2019).

more depth elsewhere (Coady 2019). Here I will confine myself to briefly discussing some particularly alarming recent developments at Google.

In 2017, Google changed its highly secretive search algorithm. It justified this change in the name of stopping the spread of ‘fake news’, a phenomenon which a senior executive defined, on a blog announcing the change, as “content on the web [that] has contributed to the spread of blatantly misleading, low quality, offensive or downright false information”.<sup>13</sup> Since the algorithm is secret, it’s hard to assess its full impact. However, there is evidence that it favours institutions over individuals and large institutions over smaller institutions. In particular, it appears to have dramatically affected the flow of traffic to independent outlets (i.e. outlets that are independent of nation states and global corporations), such as Alternet,<sup>14</sup> Truthout, Consortium News, and the World Socialist Web Site.<sup>15</sup>

This new algorithm appears to be a clear attack on the democratic and egalitarian ideals of the World Wide Web and of the Internet itself. Some people will baulk at describing this as ‘censorship’, because they think that, by definition, only governments can engage in censorship. I think that would be a mistake. If we restrict the concept of *censorship* in such a way that non-governmental entities can’t engage in it, we lose the moral significance that it should have. Certainly, we can’t justify restricting our concept of *censorship* to *government censorship* on the ground that governments are uniquely powerful. To the extent that Francis Bacon was right that knowledge is power, Google is surely amongst the most powerful organizations in the world, arguably more powerful than any current government. It is true that Google and other media/technology giants that have been exploiting the ‘fake news’ scare<sup>16</sup> lack the direct monopoly on violence (or ‘legitimate violence’ if you prefer) that is often taken to be definitive of states, but through their influence on governments they still have a lot of say over who the targets of state violence will be.

#### 4. The *Washington Post* and ‘Fake News’: A Case Study

Another example of the harms caused by the neologism ‘fake news’ comes from the esteemed *Washington Post*. Although the *Post* is not itself an international technology corporation, it is owned by Jeff Bezos, founder, chairman, and chief executive officer of *Amazon*. Furthermore, for the last few years the *Washington*

<sup>13</sup> See <https://www.theguardian.com/technology/2017/apr/25/google-launches-major-offensive-against-fake-news> (accessed 31/10/2019).

<sup>14</sup> See <https://www.alternet.org/media/editorial-googles-threat-democracy-hits-alternet-hard> (accessed 31/10/2019).

<sup>15</sup> See <https://www.wsws.org/en/articles/2017/07/27/goog-j27.html> (accessed 31/10/2019).

<sup>16</sup> Facebook, which has responded to the ‘fake news’ hysteria by making it harder for its two billion or so users to see any news at all, is a notable example.

*Post* has been routinely cited by those in the grip of ‘fake news’ hysteria as an example, indeed as a paradigmatic example, of the kind of real ‘real news’ with which ‘fake news’ is to be contrasted and it has repeatedly presented itself as the scourge of all that it deems to be ‘fake news’. For all these reasons, the *Washington Post* is a good example of the harms caused by this concept.

In recent years, the *Washington Post* has published a series of sensationalist and false stories about alleged Russian attacks on the American way of life.<sup>17</sup> Do these false stories mean that the *Washington Post* is itself a fake news source? That depends in part on how one defines ‘fake news’. Some definitions, such as that by philosopher Lee McIntyre, require fake news to be not merely false but intentionally false (McIntyre 2018, p. 112). Others, as we shall see, do not.

If, for the sake of argument, we adopt a definition which requires a fake news story to be deliberately false, and we also grant, for the sake of argument, that the *Post* was being *merely* careless with respect to the truth of these news items, rather than actually engaging in deliberate deception, would that really mark an important distinction between this kind of behaviour and outright lies?

Arguably, from a strictly consequentialist perspective, there is no important distinction here. Whatever the motives, the effects of these false stories are largely indistinguishable from those of whatever one regards as fake news. The false claims travel widely across the Internet, and are believed by large numbers of people. The propagators of the falsehoods profit from this, and there is no accountability of a kind that would give them an incentive not to repeat the behaviour. It is true that the *Post* ultimately corrected some of the falsehoods in question, but the retractions were not given anything like the prominence of the articles themselves, and were certainly seen by many fewer people than saw the original stories. What is more, these retractions do not distinguish the *Post* from the kind of news sources it characterizes as ‘fake news’, which also sometimes retract false stories. The *Denver Guardian*, for example, which became one of the paradigms of a ‘fake news’ source after it published a notorious false story entitled “FBI agent Suspected in Hillary Email Leaks Found Dead in Apparent Murder-Suicide”, ended up retracting the story.<sup>18</sup>

Even if you’re not a consequentialist and you see an important distinction between (what we are supposing) is the *Post*’s repeated recklessness with regard to the truth and what is presumably the straight-out lie of the the *Denver Guardian*, it’s not necessarily a distinction that works in the *Post*’s favour. If the *Post* is guilty of bullshitting (i.e being indifferent to the truth of its reports) rather than actual lying, Harry Frankfurt at least would say that is worse:

<sup>17</sup> See <https://theintercept.com/2017/01/04/washpost-is-richly-rewarded-for-false-news-about-russia-threat-while-public-is-deceived/> (accessed 31/10/2019).

<sup>18</sup> See <https://twitter.com/adamjohnsonNYC/status/816364572554698754> (accessed 31/10/2019).

Someone who lies and someone who tells the truth are playing on opposite sides, so to speak, in the same game. Each responds to the facts as he understands them, although the response of the one is guided by the authority of the truth, while the response of the other defies that authority and refuses to meet its demands. The bullshitter ignores these demands altogether. He does not reject the authority of the truth as the liar does, and oppose himself to it. He pays no attention to it at all. By virtue of this, bullshit is a greater enemy of truth than lies are.

(Frankfurt 2005, pp. 40–1)

It may be useful to think of lying (like terrorism) as a weapon of the weak. Those working in the establishment media often don't have to lie, because they can push responsibility for false reporting back on to their sources. Judith Miller, for example, who published numerous false stories about Iraqi Weapons of Mass Destruction in the *New York Times* in 2002 and 2003, significantly bolstering the case for the US invasion, notoriously had this to say about her role morality:

My job isn't to assess the government's information and be an independent intelligence analyst myself. My job is to tell readers of the *New York Times* what the government thought about Iraq's arsenal.<sup>19</sup>

It seems to me that this kind of indifference to the truth is at least as bad as lying. The smaller news outlets that are more likely to be labelled fake news usually don't have access to the sources inside the government and military that Miller and other establishment journalists have, so they don't have the luxury of being able to shift responsibility for false reporting onto those sources.

## 5. Academia and the ‘Science of “Fake News”’

Consider the approach adopted by Lazer et al. in the letter “The Science of Fake News” which I alluded to earlier, of saying that the correct, indeed the ‘scientific’, approach to the issue requires us to define the term more precisely and consistently than is typically done. This is the definition that appears in the letter:

We define “fake news” to be fabricated information that mimics news media content in form but not in organizational process or intent. Fake-news outlets, in turn, lack the news media’s editorial norms and processes for ensuring the accuracy and credibility of information. (Lazer et al. 2018)

<sup>19</sup> See <https://www.smh.com.au/opinion/the-new-york-times-role-in-promoting-war-on-iraq-20040323-gdilbl.html> (accessed 31/10/2010).

This definition is quite different from, and incompatible with, many other definitions in the literature. According to some definitions, fake news must be intentionally false (e.g. McIntyre 2018, p. 112); according to others, it must be false but need not be intentionally so (e.g. Levy 2017, p. 20), while still others don't require fake news to be false at all (e.g. Gelfert 2018 and Meyer 2019). It is a striking feature of the academic literature on so-called 'fake news', especially the literature that presents itself as scientific, that even though they build on each others 'results' and cite each other as if they are talking about the same phenomenon, a glance at their definitions reveals that this is not true.

Putting that concern aside for the moment, what are the "editorial norms and processes" that, according to Lazar et al., characterize the (presumably real) media? The authors deliver a brief history lesson about this, according to which journalistic norms of "objectivity and balance" developed after the First World War as a backlash against the widespread use of propaganda (including by the journalists who later embraced these norms) and the rise of corporate public relations in the 1920s. These norms, they claim, were sustained by the local and national oligopolies that dominated the twentieth-century technologies of information distribution (print and broadcast). Now, in the twenty-first century, we are warned, these norms are being undermined by internet-driven 'fake news'.

What do Lazer et al. mean by the norms of "objectivity and balance" that characterize the real news? They don't tell us. Yet neither of these terms exactly wears its meaning on its sleeve, and the word 'objectivity' is one which has particularly bedeviled philosophy. Sometimes 'objectivity' is simply used as a synonym for 'truth', but that can't be what it means here. It simply wouldn't be credible to maintain that the norm of truth only gained traction in the media after the First World War.

A nice illustration of why the norm of 'objectivity and balance', as it is interpreted by corporate and state-run media, is not a desirable norm and not a means of acquiring truth, can be seen in National Public Radio's (NPR) reaction to a 2016 column by one of its journalists, Cokie Roberts, warning of the dangers of a Trump presidency (Trump was at the time still running for the Republican nomination).<sup>20</sup> NPR vice president Michael Oreskes responded to this column by writing an internal memo to staff warning them not to criticize Trump.<sup>21</sup> In an interview that Oreskes directed Roberts to do with *Morning Edition* about the matter, the host David Greene chastised Roberts for expressing negative views about Trump in the following terms:

<sup>20</sup> See <http://www.cjonline.com/opinion/2016-02-26/steve-and-cokie-roberts-gop-must-stop-trump-now> (accessed 30/10/2019).

<sup>21</sup> See <https://www.npr.org/sections/npr-extra/2016/03/14/470352605/from-mike-oreskes-commentators-and-politics> (accessed 31/10/2019).

Objectivity is so fundamental to what we do. Can you blame people like me for being a little disappointed to hear you come out and take a personal position on something like this in a campaign?<sup>22</sup>

This abdication of the fundamental principle of journalism that one should speak the truth, especially in the face of power, for the sake of ‘objectivity’ understood here as political neutrality between the major parties, is not new (though, as we will see, it is not nearly as old as Lazer et al. think). It was particularly evident during the Bush presidency when every large media outlet in America (and many outside America) suppressed any references to torture and other well-documented war crimes by United States authorities in the name of ‘objectivity’. Particularly infamous were the explicit policies of outlets including NPR, the *Washington Post*, and the *New York Times*, not to use the word ‘torture’ for CIA practices that had long been universally recognized as such, and which they continued to describe as torture when used by governments other than the United States and its close allies.<sup>23</sup> All of this was justified in the name of ‘objectivity’ and ‘balance’, which are understood to mean being non-partisan, which in turn is understood as not taking sides in disputes between leaders of the major political parties. In 2009, Alicia C. Shepard, the NPR ombudsman, defended NPR’s policy of refusing to report that the Bush administration were practicing torture in the following terms:

It’s a no-win case for journalists. If journalists use the words “harsh interrogation techniques,” they can be seen as siding with the White House and the language that some US officials, particularly in the Bush administration, prefer. If journalists use the word “torture,” then they can be accused of siding with those who are particularly and visibly still angry at the previous administration.<sup>24</sup>

This is, indeed, a no-win situation for journalists if winning consists in keeping everyone (or everyone who matters) happy. However, it is not a no-win situation for journalists whose goal is to report the truth.

Contrary to what Lazer et al. and the oligopoly controlled media itself would have us believe, the norm of objectivity or balance, understood as neutrality between the leaders of the major parties is not a long-standing tradition. In the USA, the UK, and Australia it appears to date back to the beginning of the 1980s when media deregulation led to the consolidation of family-owned news outlets into conglomerates owned by major (often international) corporations. Such

<sup>22</sup> See <https://www.npr.org/templates/transcript/transcript.php?storyId=470340825> (accessed 31/10/2019).

<sup>23</sup> See <https://www.theatlantic.com/daily-dish/archive/2010/07/the-nyt-we-changed-reality-because-cheney-wanted-us-to/185229/> (accessed 31/10/2019).

<sup>24</sup> See [https://www.npr.org/sections/ombudsman/2009/06/harsh\\_interrogation\\_techniques.html](https://www.npr.org/sections/ombudsman/2009/06/harsh_interrogation_techniques.html) (accessed 31/10/2019).

corporations dislike taking controversial stands, because doing so alienates customers, and they particularly hate offending those who have (or might soon have) political power, because it's bad for business. As a result, the political journalist's role as a truth-speaker has often been neutered in the name of 'objectivity and balance', and many of those employed in the corporate media have been reduced to little more than stenographers giving equal time or space to the assertions of each side of a political duopoly. This trend was exacerbated by the transformation of the concept of *journalism* itself, at around the same time, from being a *trade* to being a *profession* accompanied by professional codes of ethics that reinforce the idea that political neutrality (i.e. non-partisanship) is part of their role morality. All of this, it needs to be emphasized, pre-dated the Internet.

The fact that 'objectivity' or 'balance', understood as neutrality or non-partisanship, is not the long-standing tradition its advocates like to pretend, is detailed (at least for American journalism) in Serrin and Serrin (2002). That openly and proudly partisan journalism in mainstream media outlets thrived well past the middle of the twentieth century is evident from the fact that the most honoured American journalists of the twentieth century, Edward R. Murrow and Walter Cronkite, are best known for their most openly partisan work: Murrow for denouncing Senator Joseph McCarthy, and Cronkite for denouncing the Vietnam War. There can be little doubt that they would be fired for such acts if they were working in today's corporate or publicly owned media.

So much for the media norms that Lazer et al. favour. What are the media 'practices' that fake news is allegedly undermining? The only practice they mention is fact-checking. Now it is certainly true that the kind of small independent news sources that are most likely to be labelled 'fake news' are usually unable to employ specialized fact checkers. But it would be too hasty to conclude from this their reports are less likely to be true. In the first place, corporate and state media have always exaggerated the role fact checkers play in their reporting. The long-standing "Guidelines on Integrity" for the *New York Times*, for example, state that "writers at the *Times* are their own principal fact checkers and often their only ones".<sup>25</sup> It is certainly true that competition from the Internet has led several large media companies to lay off fact checkers along with other staff. But this does not mean that fact checking has gone away. On the contrary, it is enjoying a remarkable renaissance. A number of fact-checking organizations, such as PolitiFact and Snopes in the United States, have emerged in recent years, and similar organizations have since appeared in the UK and Australia.

<sup>25</sup> See <https://www.nytimes.com/editorial-standards/guidelines-on-integrity.html> (accessed 31/10/2019).

This new form of fact checking differs from the old in at least two ways. First, the old fact checkers evaluated information in their own publications.<sup>26</sup> By contrast, the new fact checkers are assessing the claims of others, usually politicians. Second, the old fact checkers were working behind the scenes prior to publication. The public would only be aware of their work if they made a particularly glaring mistake. The new fact checkers, by contrast, are working in the public arena, and the public has the opportunity to see their work and evaluate it for themselves. What we should expect to find in the new media landscape therefore is what in fact, it seems to me, we do find. There are more false statements in the news (however you define it) than ever before, but this is not having the kind of adverse effects on the public's epistemic states that the promoters of the fake news panic would have us believe. False stories are only a problem to the extent that they are believed, and thanks to the research and outreach afforded to citizens by the Internet, they are in a much better position to evaluate the merits of reports they come across in dialogue with other citizens.

As we have seen, definitions of 'fake news' in the academic literature vary widely. Some stipulate that it be intentionally false (e.g. McIntyre 2018), some that it be false but not necessarily intentionally false (e.g. Levy 2017 and de Ridder 2019), while others don't require it to be false at all (e.g. Lazer et al.). Other definitions go in a somewhat different direction. Marco Meyer (2019) follows Gelfert (2018) in applying the label 'fake news', not to individual items of information (the kind of thing that can be false or intentionally false) but to sources of information and especially websites. On their view, fake news sites are those in which "(typically) false or misleading claims" are presented as "news". Meyer (2019) presents the results of a study he conducted in which he presented articles from 'real news' and 'fake news' to participants in a random order and took records of how credible they found the articles. The most striking thing about Meyer (2019) is that although he finds that many people are unable to distinguish 'fake news' from 'real news', he never questions his own ability to reliably make this distinction. Given that news items can be on any subject whatsoever, it seems that Meyer is implicitly claiming that he has universal expertise: an ability to distinguish truth from falsehood regardless of subject matter. When Meyer (2019) draws the conclusion that those prone to believe 'fake news' are also likely to lack intellectual virtues, including "intellectual humility", he seems oblivious to the irony.

<sup>26</sup> The grand tradition of big media fact checking was actually less about an ethical commitment to the truth than it was about minimizing the dangers of expensive defamation suits and negative publicity.

## 6. Fake News and the Russian Threat

Most of the pernicious uses of the term ‘fake news’ I have been discussing have been driven by fears of the Russian state using the Internet to interfere with Western political systems.<sup>27</sup> These fears are usually driven by people who blame Putin for Trump’s rise to power, and seem to regard Trump as some sort of Manchurian Candidate, even though Trump has demonstrably had a much more anti-Russian foreign policy than Obama. It is worth remembering in the midst of all this hysteria that only a very small percentage of American adults get their news from social media.<sup>28</sup> And most Americans, especially the older ones who are more likely to vote for Trump, get their news from television where they’re influenced by a much older and entirely homegrown brand of disinformation.<sup>29</sup> Since Trump was elected, establishment Democrats in the USA and their allies around the world have used the ‘fake news’ scare as a distraction from the disastrous Democratic election campaign, and as a way of avoiding dealing with the profound degeneration of the American polity that allowed a demagogue like Trump to rise to power.

## 7. Conclusion

The epistemic panic about so-called ‘fake news’ is the latest manifestation of a broader epistemic panic that has been going on, principally amongst older professional men, since the emergence of the Internet. Behind it, there is a misguided nostalgia for the days of broadsheet newspapers with their solemn pronouncements, designed to be read by the head of a household before going off to work. Elsewhere, I have compared this epistemic panic with the one that gripped many ecclesiastical and worldly authority figures in Europe when printing technology first emerged (Coady 2012, pp. 159–60). Suddenly, people had access to a great deal more information (including, of course, false information), and as a result people were less likely to believe what authority figures told them and, as a result, less likely to do what they were told to do.

There is no doubt that people now have access to more false statements presented as news than they ever have in the past. But, for two reasons, I see no need to panic about this. First, as I have already indicated, people now have more resources available to them to evaluate the veracity of information they come across. The merchants of the ‘fake news’ panic talk as if people are entirely passive

<sup>27</sup> Obviously, the Russian government’s own censorship regime justified by the ‘fake news’ threat is an exception.

<sup>28</sup> See <http://web.stanford.edu/~gentzkow/research/fakenews.pdf> (accessed 31/10/2019).

<sup>29</sup> See <http://inthesetimes.com/article/20938/fake-news-russia-meddling-democracy-media-right-wing-indictment> (accessed 31/10/2019).

in the face of what they are told. This is what makes the Lazer et al. (2018) agenda of (in their own words), “making structural changes aimed at preventing exposure of individuals to fake news”, so profoundly authoritarian and sinister. It presupposes that ‘we’ (i.e. whoever is making the structural changes in question) already know which reports are true and which are false, regardless of the subject matter of those reports. If there really were a group of people with this form of universal expertise whom we could trust to determine on our behalf which news is real and which is fake, then we would have no need to rationally inquire into the facts ourselves or debate them amongst ourselves. Indeed, we would have no need to vote ourselves. We could leave all of these activities to these god-like figures.

The second reason I’m not panicking about the indisputable fact there are now more false news reports than ever before is that avoiding falsehood (whether it be false reports or false beliefs resulting from those reports) is not the only value we should be concerned about. Suppose, for the sake of argument, that on average people now believe more falsehoods about politics than in the past. This sounds alarming, but arguably it is an inevitable consequence of them having more information, and in the process gaining more knowledge, than they did in the past. The acquisition of false beliefs is an inevitable consequence of the enterprise of knowledge acquisition. If all we were concerned about were avoiding false beliefs, we’d stay at home with our heads under our pillows trying to avoid acquiring any beliefs at all. William James famously ridiculed obsessive concern for avoiding false beliefs, in the following terms:

It is like a general informing his soldiers that it is better to keep out of battle forever than to risk a single wound. Not so are victories either over enemies or over nature gained. Our errors are surely not such awfully solemn things. In a world where we are so certain to incur them in spite of all our caution, a certain lightness of heart seems healthier than this excessive nervousness on their behalf.

(James 1897/1960, Part VII)

We have more knowledge than our ancestors, both collectively and as individuals. An inevitable corollary of this is that we now (almost certainly) have more false beliefs than they. The expansion of our knowledge is, in part, a result of us not being overly concerned about false beliefs. That is part of the price we pay, and it seems to me that it’s been a price worth paying.

In discussion, some people have interpreted me as claiming that the problem of fake news has been exaggerated. This does not go far enough. I have argued that there is no fake news problem at all (i.e. no problem that deserves the name ‘fake news problem’), on the ground that term ‘fake news’ itself has no legitimate meaning. The term does not correspond to any new phenomenon; nor does it refer to any phenomenon that we cannot talk or think about without resort to neologisms. In short, it serves no valuable function at all. That is not to say that it

serves no function. On the contrary, it serves at least two objectionable functions: it pathologizes views that people using the term disagree with, and it marginalizes the voices of relatively powerless participants in public discourse. In short, the term serves no good purpose while causing considerable harm. Until recently, we managed to get by without it. I'm confident we can learn to do so again.

## References

Coady, David (2012), *What to Believe Now: Applying Epistemology to Contemporary Issues*. Oxford: Wiley-Blackwell.

Coady, David (2019), “The Trouble With ‘Fake News’”, *Social Epistemology Review and Reply Collective* 8(10): 40–52.

De Ridder, Jeroen (2019), “So What if ‘Fake News’ is Fake News?”, *Social Epistemology Review and Reply Collective* 8(10): 111–113.

Frankfurt, Harry (2005), *On Bullshit*, Princeton NJ: Princeton University Press.

Gelfert, Axel (2018), “Fake News: A Definition”, *Informal Logic*, 38(1): 84–117.

Habgood-Coote, Joshua (2019), “Stop Talking about Fake News!”, *Inquiry* 62(9–10): 1033–65.

James, William (1897/1960), *The Will to Believe, Human Immortality, and Other Essays in Popular Philosophy*, Mineola, NY: Dover Publications.

Lazer, David M., Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, Michael Schudson, Steven A. Sloman, Cass R. Sunstein, Emily A. Thorson, Duncan J. Watts, Jonathan L. Zittrain (2018), “The Science of Fake News”, *Science* 359(6380): 1094–6.

Levy, Neil (2017), “The Bad News About Fake News”, *Social Epistemology Review and Reply Collective* 6(8): 20–36.

McIntyre, Lee (2018), *Post-Truth*, Cambridge, MA: MIT Press.

Meyer, Marco (2019), “Fake News, Conspiracy, and Intellectual Vice”, *Social Epistemology Review and Reply Collective* 8(10): 9–19.

Serrin, Judith and William Serrin (2002), *Muckraking: The Journalism that Changed America*, New York: The New Press.

# Conspiracy Theories and Evidential Self-Insulation

*M. Giulia Napolitano*

## 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,<sup>1</sup> I maintain that conspiracy theories are not theories (or explanations) at all.<sup>2</sup> 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.

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

<sup>2</sup> In line with the literature on conspiracy theories, I use ‘theory’ and ‘explanation’ as synonyms, despite the obvious differences between the two.

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.

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. 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.<sup>3</sup> I am thinking of theories such as the fake moon landing, flat earth, or the Illuminati controlling the world. I will not

<sup>3</sup> 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.

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 theorizing.<sup>4</sup>

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.<sup>5</sup> 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.<sup>6</sup> 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.<sup>7</sup>

<sup>4</sup> 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*, see Napolitano & Reuter (ms).

<sup>5</sup> 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.

<sup>6</sup> For an in-depth discussion of this definition, see Dentith (2014).

<sup>7</sup> 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 Napolitano & Reuter (ms). 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.

However, by assuming that every theory involving a conspiracy is a conspiracy theory, 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.<sup>8</sup>

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

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

fruitfulness. Hence, a failure to completely match our intuitions about what conspiracy theories are should not be considered a reason to reject it.

### 3. Conspiracy Theories

It is commonly assumed that conspiracy theories are, at the very least, theories that involve conspiracies.<sup>9</sup> 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 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.<sup>10</sup> 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

<sup>9</sup> 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.

<sup>10</sup> 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).

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 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.<sup>11</sup> 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 is any evidence that they

<sup>11</sup> 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.

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.<sup>12</sup>

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.

#### 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.<sup>13</sup> In other words, I argue that it is irrational to hold conspiracy theories.<sup>14</sup> Even though my account of conspiracy theories is significantly different from traditional accounts, the discussion in this section has

<sup>12</sup> I am grateful to Paul Silva for helpful discussion on this point.

<sup>13</sup> 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).

<sup>14</sup> 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 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 2018). Attributing conspiracism to believers runs the risk of suggesting a stable disposition of the believer to form this type of irrational beliefs. 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.

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.<sup>15</sup> 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 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

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

nature of conspiracies.<sup>16</sup> 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 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.<sup>17</sup>

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.<sup>18</sup> 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.<sup>19</sup> 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.

<sup>16</sup> 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.

<sup>17</sup> For an introduction to Bayesian confirmation theory, see Bovens & Hartmann (2003); Strevens (ms).

<sup>18</sup> 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.

<sup>19</sup> 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.<sup>20</sup>

<sup>20</sup> 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.

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 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.<sup>21</sup>

<sup>21</sup> 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 Section 5.2.

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:

**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.<sup>22</sup> 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 irresponsible 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.<sup>23</sup>

<sup>22</sup> For instance, Dunbar & Reagan (2006).

<sup>23</sup> 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

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.<sup>24</sup> 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 existence 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

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 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.

<sup>24</sup> 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.

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.

## 5. Objections and Replies

### 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.<sup>25</sup> 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 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,<sup>26</sup> 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 and 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 that the reason why such

<sup>25</sup> See, for instance, Cappelen (2018); Haslanger (forthcoming); Nado (2019); Prinzing (2017); Sawyer (2018); Thomasson (forthcoming).

<sup>26</sup> See, for instance, Barkun (2003); Fenster (1999); Hofstadter (1965).

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.<sup>27</sup> 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 our intuitions about what conspiracy theories are.

## 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

<sup>27</sup> This approach to the problem is also compatible with the existence of different revisionary accounts of 'conspiracy theory.'

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,<sup>28</sup> 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

<sup>28</sup> 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.

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.<sup>29</sup>

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, including 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.

<sup>29</sup> 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).

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.<sup>30</sup> 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.<sup>31</sup> 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, 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*.

<sup>30</sup> For a Bayesian analysis of how reflexivity alters confirmation relations, see Kopec (2011).

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

### 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.

## 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.<sup>32</sup>

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.<sup>33</sup> 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.<sup>34</sup>

<sup>32</sup> Such as the role of conspiracy theories as forms of political propaganda (Cassam 2019).

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

<sup>34</sup> 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.

## References

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): 265–80.

Basham, L. (2003). Malevolent global conspiracy. *Journal of Social Philosophy* 34(1): 91–103.

Basham, L. (2018). Social Scientists and Pathologizing Conspiracy Theorizing. In M. R. X. Dentith (ed.), *Taking Conspiracy Theories Seriously*. London: Rowman and Littlefield, pp. 95–108.

Basham, L. & Dentith, M. (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. (forthcoming). Evidential Preemption. *Philosophy and Phenomenological Research*.

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

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.

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

Cassam, Q. (2019). *Conspiracy Theories*. Cambridge: 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. (2012). *What to Believe Now: Applying Epistemology to Contemporary Issues*. Oxford: Wiley-Blackwell.

Coady, D. (2018a). Cass Sunstein and Adrian Vermeule on Conspiracy Theories. *Argumenta* 3(2): 291–302.

Coady, D. (2018b). Conspiracy-Baiting and Anti-Rumour Campaigns as Propaganda. In M. R. X. Dentith (ed.), *Taking Conspiracy Theories Seriously*. London: Rowman and Littlefield, pp. 171–87.

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

Dentith, M. R. X. (2017). Conspiracy Theories on the Basis of the Evidence. *Synthese* 196: 2243–61.

Dentith, M. R. X. (2018). The Problem of Conspiracism. *Argumenta* 3(2): 327–43.

Dieguez, S., Bronner, G., Campion-Vincent, V., Delouvé, 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.

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

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.

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 and Littlefield, pp. 125–39.

Harris, K. (2018). What's Epistemically Wrong with Conspiracy Theorising? *Royal Institute of Philosophy Supplement* 84: 235–57.

Haslanger, S. A. (forthcoming). How Not to Change the Subject. In T. Marques & Asa 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.

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

Keeley, B. L. (1999). Of Conspiracy Theories. *Journal of Philosophy* 96: 109–26.

Keeley, B. L. (2019). The Credulity of Conspiracy Theorists: Conspiratorial, Scientific & Religious Explanation Compared. In J. E. Uscinski (Ed.), *Conspiracy Theories and the People Who Believe Them*. Oxford: Oxford University Press.

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.

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

Muirhead, R. & Rosenblum, N. L. (2019). *A Lot of People Are Saying*. Princeton, NJ: Princeton University Press.

Nado, J. (2019). Conceptual Engineering, Truth, and Efficacy. *Synthese*, <https://doi.org/10.1007/s11229-019-02096-x>.

Napolitano, M. G. & Reuter, K. (ms) What Is a Conspiracy Theory?

Nguyen, C. Thi (2020). Echo Chambers and Epistemic Bubbles. *Episteme* 17(2): 141–61.

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 and Littlefield. pp. 141–53.

Pigden, C. (1995). Popper Revisited, Or What Is Wrong With Conspiracy Theories? *Philosophy of the Social Sciences* 25(1): 3–34.

Prinzing, M. (2017). The Revisionist's Rubric: Conceptual Engineering and the Discontinuity Objection. *Inquiry*, 61(8): 854–80.

Räikkä, J. (2009). On Political Conspiracy Theories. *Journal of Political Philosophy* 17 (2): 185–201.

Räikkä, J. (2014). On the Epistemic Acceptability of Conspiracy Theories. *Social Justice in Practice*, 61–75, [https://link.springer.com/chapter/10.1007/978-3-319-04633-4\\_5#citeas](https://link.springer.com/chapter/10.1007/978-3-319-04633-4_5#citeas).

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

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

Silva, Paul (2020). A Bayesian Explanation of the Irrationality of Sexist and Racist Beliefs Involving Generic Content. *Synthese* 197(4): 1–23.

Strawson, P. F. (1963). Carnap's Views on Conceptual Systems versus Natural Languages in Analytic Philosophy. In Paul Arthur Schilpp (Ed.), *The Philosophy of Rudolf Carnap*. La Salle, IL: Open Court, pp. 503–18.

Strevens, M. (ms). Notes on Bayesian Confirmation Theory.

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

Thomasson, A. L. (forthcoming). A Pragmatic Method for Conceptual Ethics. In H. Cappelen, D. Plunkett, & A. Burgess (Eds.), *Conceptual Engineering and Conceptual Ethics*. Oxford: 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.

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

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.



# Enquiry and Normative Deviance

## The Role of Fake News in Science Denialism

*Filippo Ferrari and Sebastiano Moruzzi*

### 1. Introduction

Science denialism is a widespread and worrying phenomenon. Rejection of standard scientific theories targets several areas of enquiry. For instance: flat-earth theorists deny basic assumptions of physics and astronomy; anti-vaccine supporters oppose compulsory vaccination by casting doubts on the efficacy of vaccines and, sometimes, by linking vaccines to severe pathologies like autism; HIV deniers put into question the very existence of HIV; climate-change deniers downplay the significance of the phenomenon and dispute its anthropogenic causes.<sup>1</sup>

Our main thesis is that science denialism brings about an aberrant form of enquiry—that we shall call *post-enquiry*—in which the epistemic norms governing scientific enquiry are deviated in significant ways.<sup>2</sup> Science denialism doesn’t merely involve a rejection of a scientific theory—otherwise scientists themselves would count as science deniers given that they would reject theories on the basis of their explanatorily inadequacy. Rather, science denialism deeply challenges the practice, common within scientific enquiry, of continuously and, to a certain extent, impartially testing research methods, theories, and evidential sources with the aim of improving the accuracy of scientific theories. In this sense, science denialism brings about a radical deviation of the norms governing the practice of scientific enquiry—a deviation which gives rise to what we shall call a *normative aberration*. We offer an in-depth analysis of the epistemic mechanisms underpinning the normative aberration brought about by science denialism. More specifically, we develop a fine-grained framework to model a variety of normative deviances that can take place in enquiry. In doing so, we focus especially on the kind of normative deviances related to science denialism and, by analysing two case studies, we argue that fake news contributes significantly to shape the

<sup>1</sup> For a collection of essays on pseudoscience and science denialism, see Kaufman & Kaufman 2018. For a detailed discussion of the flat-earth case, see Garwood 2007.

<sup>2</sup> We develop in detail the notion of post-enquiry in Ferrari & Moruzzi 2020.

epistemic norms operating within science denialism. They in fact play two pivotal roles: first, they are used to cast discredit on a variety of (institutional) sources of evidence in relation to a certain set of phenomena (e.g. whether vaccines are safe); second, they also play a part in building the alternative explanation of the targeted phenomena.

This chapter is structured as follows. In Section 2, we characterize enquiry by developing a model of epistemic normativity. In Section 3, we introduce the notions of background assumptions and epistemic filters. In Section 4, we provide two criteria for assessing the good-standing of an enquiry. In Section 5, we illustrate some varieties of normative deviance. In Sections 6 and 7, we discuss two examples of science denialism arguing that they are instances of a specific kind of normatively aberrant enquiry, namely a *post-enquiry*.

## 2. Enquiry and Its Normative Setting

Enquiry is the complex practice of gathering, weighing, and assessing evidence which is aimed at forming, managing, and revising beliefs for the sake of acquiring and sharing true information. We enquire daily into a variety of subject matters—from rather mundane questions, such as whether to take the subway or the bus to reach our workplace, to incredibly advanced and complex questions, such as whether quantum mechanics can be fully reconciled with general relativity. What do all these specific enquiries have to do with the minimal characterization of enquiry? Quite simply, they can be seen as local enquiries which answer to specific sets of questions.

It is evident from the way we typically conduct our enquiries that we conceive of them as normatively constrained practices: there are shared standards and methods concerning how to conduct and arbitrate enquiry in a successful manner. In fact, as Shah and Velleman (2005) have argued, enquiry and its products, primarily beliefs, are regulated by a double alethic standard. The first, *alethic telos*, is that in forming, managing, and revising beliefs enquirers aim at maximizing truth and minimizing falsity.<sup>3</sup> This standard is paramount for assessing the overall good standing of an enquiry on the basis of how it scores on the two axes of maximizing truth and minimizing falsity.<sup>4</sup> The second alethic standard, *alethic*

<sup>3</sup> The view that truth is the chief normative ideal of enquiry is not uncontested. Some take knowledge to play that role (e.g. Williamson 2000; Kelp 2014). Replacing truth with knowledge doesn't pose a threat to our framework since evidence remains central, but it changes some of its predictions in relation to the normative deviances of some enquiries discussed in Section 5 (e.g. while the Cartesian sceptic scores high on the minimizing falsity axis, she doesn't score well on the minimizing lack of knowledge (ignorance) axis). This flexibility of our framework is an advantage rather than a limit. Thanks to an anonymous referee for pointing this out to us.

<sup>4</sup> This idea is elegantly summarized by James's famous expression: "Believe truth! Shun error!"—see James 1897.

*criterion*, provides the chief criterion for assessing the correctness of beliefs. While the alethic telos is common to other cognitive attitudes involved in the enquiry such as conjecturing, hypothesizing, or assuming, the second one is characteristic of beliefs. We can assume something (that we know to be) false, without necessarily doing something incorrect (from an alethic point of view). However, if we believe  $\langle p \rangle$  and  $\langle p \rangle$  is false, then our belief is incorrect.<sup>5</sup> In this sense, truth is the ultimate regulative ideal of enquiry: in pursuing enquiry we are disposed to take truth to be the aim as well as the standard of correctness of our beliefs.

As all regulative ideals, it is not always possible for a cognitively limited and epistemically located subject to follow directly the lead of truth and the best she can do in trying to comply with the ideal of truth is to follow the evidence. We use ‘evidence’ as a broad term including all sorts of epistemic justification and we take evidence to be non-truth-entailing—evidence for believing  $\langle p \rangle$  does not entail the truth of  $\langle p \rangle$ —but nevertheless truth-oriented—evidence for believing  $\langle p \rangle$  is evidence for believing that  $\langle p \rangle$  is true, and evidence against believing  $\langle p \rangle$  is evidence against believing that  $\langle p \rangle$  is true. We can now supplement the normative structure of enquiry by adding the following two evidential norms:

(EN1—Belief): a subject is epistemically permitted to form the belief that  $p$  if and only if she has (strong enough, undefeated) evidence for the belief that  $p$ .

(EN2—Revision): a subject is epistemically required to revise her belief that  $p$  if and only if she has a (undefeated) defeater for the belief that  $p$ .

A piece of evidence is *in the range of an epistemic norm* when it is relevant for the normative assessment issued by that norm. More specifically, a piece of evidence that is in the range of EN1 is a *candidate justification* just in case it is relevant for the normative assessment issued by EN1. By the same token, a piece of evidence that is in the range of EN2 is a *candidate defeater* just in case it is relevant for the normative assessment issued by EN2. A candidate justification may fail to count as a justification (there could be an undefeated defeater of it), and, for the very same reasons, a candidate defeater may fail to count as a defeater. The evidence relevant for the application of EN1, if effective, counts as a *justification* for believing (a certain proposition), whereas the evidence relevant for the application of EN2, if effective, counts as a *defeater* for believing (a certain proposition).

Epistemic and alethic norms together give us the formal normative structure of enquiry. However, when we look at specific enquiries we need to contextualize this normative structure. In particular, we need to understand EN1 and EN2 in relation to the epistemic situation of the enquirer.<sup>6</sup> This is because each enquiry involves a specific enquirer and a specific epistemic situation. Thus, EN1 and EN2

<sup>5</sup> See Ferrari 2018 for a pluralist analysis of truth’s normative functions within enquiry.

<sup>6</sup> Henceforth we make reference to a single enquirer, but all we say applies equally to groups.

give us what we may call the *intension* of the normative concepts *epistemically permissible* and *epistemically required*. This intension is a function from an epistemic situation to extensions where the notion of ‘epistemic situation’ captures the various elements that constitute the perspective of the enquirer on the evidence and by ‘extension’ we mean the set of cognitive actions which, for our purposes, is limited to belief formation and revision. Thus, while the intension of *epistemically permissible* and *epistemically required* are invariant across different enquiries, their extensions may vary from enquiry to enquiry (we explain how this can happen in Section 3). As a result, we have a plurality of enquiries each of which comes with its own verdict on what counts as epistemically permissible and/or required.

Let us be very clear on one important point. By claiming that there is a certain variability in what is epistemically permitted and/or required in a given enquiry we are not suggesting that all enquiries are equally in *good* standing. One thing is to say that what is epistemically permissible and/or required varies in tandem with variations in certain structural features of enquiry; another is to say that all enquiries are in good standing. An overall epistemic assessment of an enquiry crucially depends on how it scores on the *alethic telos*—namely, its degree of success in maximizing truth and minimizing falsity.

It could be claimed that what is *absolutely* epistemically permissible/required is determined on the basis of epistemically virtuous practices of enquiry. This, however, risks concealing the normative intelligibility of certain deviant practices by assessing them as outright irrational. We think that this is a delicate issue. In many cases—and we believe that the anti-vaxxers and flat-earthers are a point in our favour—it would be too hasty to deem these practices as irrational or normatively unintelligible. In fact, we think that our framework allows us to distinguish between practices which are random, and thus unintelligible, from a normative point of view, from deviated enquiries which have nevertheless a normative discipline and are oriented, even if in a limited and aberrant way, towards truth. Avoiding the prejudging of these normatively deviant practices is paramount if we want to understand the epistemic complexity of certain forms of enquiry without underestimating their significance and potential threat. We thus think that part of the explanandum of the phenomenon of scientific denialism is that there is a social practice with its own normative structure that needs to be accounted for. This chapter is primarily devoted to this task.

### 3. The Role of Background Assumptions and Epistemic Filters in Shaping Epistemic Norms

EN1 and EN2 in the formulations above are fully general and unrestricted normative principles. However, as such they are not normatively binding. It’s

only after having specified the epistemic situation of the enquirer that they become normatively binding. In fact, for a norm to be normatively binding we need to fix the extension of what is permitted and what is required—and this may vary, we claim, depending on certain beliefs of the enquirer that we are about to discuss.

There may be various elements that shape an enquirer's epistemic situation. What is crucial for us is what we call an *epistemic filter*. A core component of an epistemic filter is *the (set of) background assumption(s)* that the enquirer implicitly or explicitly takes on board in conducting her enquiry. An epistemic filter is a selection function which takes as inputs the total evidence and the set of the enquirer's background assumptions and outputs a demarcation of the body of evidence over which the epistemic norms range. In other words, an epistemic filter induces a partition in the set of the total evidence on the basis of the enquirer's background assumptions thus determining the set of evidence that falls within (and that which falls without) the range of application of the two norms. Assuming, for simplicity, that the set of total evidence is always invariant, the key element for determining an epistemic filter is given by the set of background assumptions which varies in relation to the enquirer. Let's flag out that epistemic filters impact on the range of the application of the norms without thereby impacting on the status of the evidence: for instance, a piece of evidence can be filtered out without thereby being defeated.

Background assumptions have two key features. The first is that they are beliefs which concern either the epistemic assessments of certain evidential sources (e.g. 'scientific institutions are reliable', 'oracles are unreliable') or considerations concerning the nature or kind of evidence (e.g. whether only certain evidence is trustworthy). The second feature is that background assumptions are held with an extremely high degree of confidence which may not match their epistemic robustness (the belief's evidential status). In some respects, background assumptions are analogous to what Michael Lynch calls *convictions*: "[Convictions] carry authority over what we believe. Once something becomes a real conviction, it is difficult for us to doubt it. [...] The authority that the conviction brings with it—just by virtue of being a conviction—means we may shield ourselves from evidence that may seem to undermine it" (Lynch 2019: 53–62). Our notion of background assumption shares this peculiar epistemological and normative trait with Lynch's notion of conviction, but it's less committal in that it doesn't engender "a commitment that reflects the kind of person we want to be" (Lynch 2019: 57). In this respect, an important aspect of the notion of conviction is that it is emotionally and value laden and it carries not only epistemic but also moral authority. Background assumptions are neutral on the issue of moral authority but they exert epistemic authority over enquirers. In this sense, background assumptions can be characterized, to paraphrase Lynch, in terms of a commitment that reflects the kind of enquirer we take ourselves to be. They thus

contribute to a subject's *epistemic self-conception*—i.e. the conception that a subject has of herself qua enquirer. The way we think of it is as a set of dispositional features, some of which, such as sensitivity to available evidence, are constitutive traits while others, such as the set of background assumptions held by the subject, are contingent, but nevertheless very important. We will return to these two features in Section 4.

What background assumptions are taken on board has consequences on which kind of evidence and/or what sources of evidence the enquirer considers in the formation and revision of beliefs. To use a Quinean metaphor, background assumptions are beliefs that are so central and entrenched in our web of beliefs that changing them would require us to change our views on what counts as admissible kinds and sources of evidence. As a consequence, a great deal of other beliefs would have to be dropped. Abandoning a background assumption would thus be very expensive for an enquirer in that it would require her to reshape her epistemic self-conception. To understand the mechanics of the normative effects that background assumptions have on an enquirer it may be useful to compare them with the effect that pragmatic elements (e.g. high-stakes versus low-stakes scenarios) may have on the extension of knowledge attributions, as suggested by the theorists of the so-called *pragmatic encroachment* (e.g. Stanley 2005). Because of the variability in high/low costs of abandoning a belief on the trustworthiness of certain epistemic sources or what kind of evidence to consider is relative to which background assumption is adopted, there is a variability in the extension of what is epistemically permitted to believe and required to revise across enquiries with different background assumptions.

There are two broad ways in which an epistemic filter may work: by omission and by discrediting. One way of restricting the scope of epistemic norms is by *omitting* certain pieces of evidence due to the specific epistemic locality of an agent—e.g. certain sources are excluded because they are not reachable by the agent.<sup>7</sup> A second way in which epistemic filters restrict the poll of evidence over which the epistemic norms range is by *discrediting* some sources of evidence. The mechanism of epistemic discredit is triggered by background assumptions that assess some kinds of evidence and/or some epistemic sources as unreliable.<sup>8</sup>

We now illustrate how epistemic filters delimit the range of the epistemic norms. Let us call an epistemic filter  $f$ .  $f$  restricts the set of the evidence over which EN1 and EN2 range—which, when unfiltered, are meant to be very general, ranging over any evidence and any propositions whatsoever, without restriction.

<sup>7</sup> The notion of 'reachability' encompasses a variety of aspects such as geographical, cultural, social, or cognitive aspects.

<sup>8</sup> The notions of filtering by omission and filtering by discredit are related to Thi Nguyen's concepts of exclusion by omission and exclusion by discredit—see Nguyen 2020.

Epistemic filters work by restricting the body of admissible and/or relevant evidence in relation to the specific enquiry conducted by the subject on the basis of the background assumptions taken on board by the enquirer. In this respect, the two epistemic norms can be formulated in such a way as to make the role of epistemic filters transparent. The different mechanisms of omitting and discrediting filters may be captured by means of changing the scope of the normative operators occurring in EN1 and EN2.<sup>9</sup> Let's discuss the formulation of the norm with omitting filters first. We use 'o-f' as a label for omitting filters and 'e' as a variable ranging over the total evidence; 'e<sub>of</sub>' is then a sorted variable ranging over the evidence selected by the omitting filter:

(EN1<sub>O-F</sub>): It is epistemically permitted that {subject x forms the belief that p if and only if there is an e<sub>of</sub> (e<sub>of</sub> is strong enough and undefeated for the belief that p and x has e<sub>of</sub>)}.

(EN2<sub>O-F</sub>): It is epistemically required that {subject x revises her belief that p if and only if there is an e<sub>of</sub> (e<sub>of</sub> provides x with an undefeated defeater for the belief that p and x has e<sub>of</sub>)}.

In these formulations of the epistemic norms, the normative operators (permissible and required) take wide scope—i.e. they govern the entire biconditional. EN1<sub>O-F</sub> only asks for an alignment between evidence supporting <p> and the subject forming the belief that p, whereas EN2<sub>O-F</sub> asks for an alignment between the presence of defeaters against <p> and the subject revising the belief that p. However, they do not issue any specific verdict (of impermissibility to believe in case of EN1<sub>O-F</sub>, or permissibility to not revise in the case of EN2<sub>O-F</sub>) in the absence of a piece of evidence selected by the filter. We think that this captures the idea that if some piece of evidence is omitted from the evidence selected by the omitting filter it doesn't give any direct normative guidance to the subject—that's because an omitted piece of evidence is simply left out and not discredited.

Let's now discuss the formulation of the norm with discrediting filters. Let's use 'd-f' as a label for discrediting filters, and, again, 'e' as a variable ranging over the total evidence. We thus take 'e<sub>df</sub>' to be a sorted variable ranging over the evidence selected by the discrediting filter:

(EN1<sub>D-F</sub>): Subject x is epistemically permitted to form the belief that p if and only if there is e<sub>df</sub> (e<sub>df</sub> is strong enough and undefeated for the belief that p and x has e<sub>df</sub>).

<sup>9</sup> See Steinberger 2017: §4 for a discussion of the scope of normative operators.

(EN2<sub>D-F</sub>): Subject x is epistemically required to revise the belief that p if and only if there is  $e_{df}$  ( $e_{df}$  provides x with an undefeated defeater for the belief that p and x has  $e_{df}$ ).

In both formulations, the epistemic operator ‘required’ takes narrow scope—i.e. it only governs the left-hand-side of the biconditional. This means that if a piece of evidence is left out by the discrediting filter it has direct normative consequence on what is impermissible for the subject to believe (in the case of EN1<sub>D-F</sub>) and permitted not to revise (in the case of EN2<sub>D-F</sub>). This is because in the absence of evidence selected by the filter, by contraposition the norms issue a verdict (that it is epistemically impermissible for the subject to believe, in the case of EN1<sub>D-F</sub>, and that the subject is epistemically permitted to not revise, in the case of EN2<sub>D-F</sub>). This normative prediction, we think, captures the idea that what is left out by the discrediting filter receives a negative epistemic assessment.

#### 4. Assessing Background Assumptions

Is it possible to provide an epistemic assessment of background assumptions? And if so, how does such an assessment work? Our view is that any background assumptions can be subject to an epistemic assessment and that some forms of (post-)enquiry whose normative extension is determined by what we would intuitively call bad filters are indeed bad forms of enquiry. We call these kinds of enquiries *normatively aberrant enquiries*.

In Section 3 we have argued that what enquiry-related cognitive acts are epistemically permitted or required for an agent to perform is only determined within an enquiry, namely, only once a (set of) background assumptions(s) is specified and consequently a (set of) epistemic filter(s) is in place. This kind of normative evaluation concerns the set of enquiry-related cognitive acts that enquirers execute in the context of an enquiry—i.e. it targets the forming, managing, and relinquishing of beliefs. Our framework predicts that such an epistemic evaluation is never absolute, but only relative to a specific enquiry (with its background assumptions and epistemic filters). However, there is another kind of epistemic assessment that does not concern the cognitive acts of enquirers within an enquiry but rather the good-standing of enquiries themselves. Let us clarify, at this point, that in the light of what has been said in the previous section we should refine the characterization of an enquiry given at the beginning of Section 2 by adding that any enquiries involve some background assumptions that are held by the enquirers in the way previously specified. Since these background assumptions take part in the determination of the normative extension of enquiry by means of the corresponding epistemic filters, an epistemic assessment of a background assumption carries with it an assessment of any enquiries that involve

these assumptions. Some enquiries (e.g. standard scientific enquiry) score well from the alethic point of view since their normative extension is such that it better fulfils the alethic telos—i.e. the twofold aim of maximizing truth and minimizing falsity.<sup>10</sup> A *normative aberration* indicates a deficient kind of normative deviance from standard scientific enquiry. For an enquiry to be normatively aberrant is to fail to satisfy the alethic telos—namely, to score badly on maximizing truth and minimizing falsity. We shall also add a second element of epistemic assessment that involves the potential of an enquiry to be reformed—i.e. the issue of whether or not the background assumptions of an enquiry are open to revision. The effect of having background assumptions that are particularly resistant to revision is that of worsening the overall epistemic status of a normatively aberrant enquiry.

In order to appreciate how this level of assessment works, let's first clarify the relation between the total evidence and a specific enquiry as conducted by an enquirer. We can represent the set of the total evidence as a set of propositions over which a partition is imposed. The partition distinguishes the effective evidence (strong enough evidence to epistemically support a belief) and the ineffective evidence (evidence that fails to epistemically support a belief).<sup>11</sup> We can represent the different normative extensions of enquiries as providing different sets of rules for managing evidence in order to form and revise beliefs. Epistemic filters play a crucial role in shaping these different sets of rules for managing evidence in that they prompt restrictions on the total evidence over which the epistemic rules range. The thought, then, is that the more an enquiry uses epistemic filters that include effective evidence and exclude ineffective one for forming and revising beliefs, the more its normative extension fulfils the alethic telos.

Fulfilment of the alethic telos involves a good degree (above a specified threshold) of maximization of true beliefs and minimization of false ones on the part of the enquirers. If it turns out that an enquiry employs background assumptions that are false, this enquiry will operate with the evidence selected on the basis of incorrect assessments (recall, background assumptions are assessments of the kind of evidence and the evidential sources) and it will plausibly perform badly in maximizing truth and minimizing falsity. Given that science denialism, as we will shortly explain, involves background assumptions that are unlikely to be true (conspiracy theories based on fake news), the enquiry of science denialists scores badly in fulfilling the alethic telos: epistemic rules determined by false beliefs do not provide a reliable practice for forming true beliefs and avoiding false ones. Notice, further, that the epistemic status of an enquirer who takes part in an enquiry which

<sup>10</sup> We assume that a low score in the fulfilment of the alethic telos is a disvalue. For a discussion of the relation between veritism and the two aims of maximizing truth and minimizing falsity, see Steinberger 2019.

<sup>11</sup> For simplicity, we assume that this partition can be absolutely traced.

scores badly in fulfilling the alethic telos is worse than that of an enquirer who merely has false beliefs. This is because false background assumptions have a structural impact on the enquiry, in fact causing its failing to maximize true beliefs and minimizing false ones since they contribute to determining an extension of epistemic permissions and obligations that systematically fail to effectively guide the enquirer towards truth. On the other hand, merely having some false beliefs does not preclude the enquirer from finding the right path to truth if epistemic permissions and obligations are effectively conducive to truth.

As we have mentioned above, a second element of assessment of an enquiry concerns the degree to which its background assumptions are open to reform. When this degree is not low, we say that an enquiry is *reformable*. When an enquiry is reformed, enquirers abandon their background assumptions. Reforming enquiry in this sense does not amount to a standard case of belief revision. The reason is twofold: first, background assumptions are convictions that take part in the determination of epistemic self-conception of the enquirer; hence enquirers incur high costs (much higher, *ceteris paribus*, than in the case of standard beliefs) in abandoning them; second, a background assumption can be self-sealing since it can give rise to filters that exclude from the range of the epistemic norms all or most of the candidate defeaters for the background assumption. We discuss both points.

Let's start with considering resistance to reformability connected to the fact that background assumptions are held in the manner of convictions. If the subject has a piece of defeating evidence (selected by the filter) for her background assumption, she is nonetheless disposed to hold fast to it since the costs of giving up the background assumption threaten her epistemic self-conception. But what happens if she keeps acquiring more and more defeating evidence? Managing evidence for the purpose of revising beliefs is part of what an enquiry is. This aspect, we take it, is reflected in the epistemic self-conception of an enquirer who is sensitive to how evidence is managed. Given that ignoring defeating evidence involves ignoring an ample part of the evidence available to the subject, her epistemic self-conception is jeopardized. Epistemic self-conception is thus threatened by two factors: by the width of available evidence against background assumptions and by a high degree of insensitivity to evidence for belief revision. Since sensitivity to evidence is a constitutive trait of an enquirer's epistemic self-conception, while which background assumptions are taken on board relative to an enquiry is a contingent fact, the former aspect is more central than the latter for an enquirer's epistemic self-conception. In this respect, a high degree of insensitivity to evidence has a higher cost for the epistemic self-conception of an enquirer than giving up her background assumptions. Thus, if an enquirer systematically ignores a wide body of evidence, the degree of insensitivity to evidence reaches the critical stage in which her epistemic self-conception is pressured to change in relation to her conception of how belief revision must be sensitive to evidence. The more counter-evidence is

made available to the subject, the more the subject is disposed to abandon her background assumptions in order to preserve her epistemic self-conception. This is to avoid reaching a critical point in which the insensitivity to counter-evidence is so high that would jeopardize the epistemic self-conception completely. In this sense, an enquiry is hardly reformable by default unless a vast body of defeating evidence against its background assumptions is selected by the filter.

The second motivation to resist reform is connected to the effect that epistemic filters have on sources and kinds of evidence. An epistemic filter may impact on whether the defeating evidence for background assumptions is made available to the enquirers. Whereas a standard belief can be revised on the basis of a (undefeated) defeater regardless of its provenance, this may not be the case for background assumptions because the epistemic filter it induces may discredit or omit the source to which the defeater belongs. When an enquiry leaves open the possibility of acquiring defeating evidence for its background assumptions, it allows for the possibility of changing the extensions of the norms governing enquiry. Thus, the more an enquiry is open to defeating evidence for its background assumptions, the more it is reformable. The extreme case at the opposite end of the spectrum is an enquiry whose background assumptions rule out the possibility of acquiring defeating evidence against them since this evidence belongs to the sources and the kind of evidence that are excluded by the epistemic filters. Such enquiry is hardly reformable.

Summing up, background assumptions need not be unrevisable and enquiry can be reformable. However, reformability is challenged by a twofold line of resistance: first, we have the enquirer's disposition, *modulo* preserving her epistemic self-conception, to hold on to background assumptions due to their status of convictions within enquiry; second, we have the width of the epistemic filters—i.e. how much evidence they filter out. The narrower the width of the epistemic filters the easier it is for an enquirer to overcome her attachment to background assumptions given the size of the overwhelming defeating evidence she may be subject to. This is because ignoring such evidence would jeopardize her epistemic self-conception in a more daunting way than abandoning background assumptions. As a consequence, the extremely high degree of confidence characteristic of convictions is most stable when epistemic filters are wide, since they filter out a larger body of potentially defeating evidence. Convictions flourish in the self-sealing epistemic environment created by wide epistemic filters. A hardly reformable enquiry is thus a factor that worsens its status of normative aberration.

## 5. Varieties of Normative Deviance

With our framework in place, we are now in a position to illustrate in more detail how the normative extension of an enquiry can be changed by means of the

function of epistemic filters—and, in this sense, it can deviate from what we may call its *standard extension* which, for the purposes of this chapter, we identify with the normative extension proper of scientific enquiry as we know it. This helps us to model the normative extension of (at least some) cases of science denialism and also to distinguish them from other kinds of normative deviances of enquiry.

In order to characterize the various ways in which epistemic filters give rise to different normative extensions we help ourselves with two figures that illustrate the different ways in which the range of an epistemic norm can be shaped by an epistemic filter. For the sake of introducing the kind of normative deviance exemplified by science denialism (to be discussed in Sections 6 and 7), we focus only on the effect of discrediting filters, assuming that all the types of enquiry represented in the figures are cases of enquiries restricted by discrediting filters. We interpret these figures as representing the normative extension for types of enquiries on empirical matters (but the framework wants to be neutral and fully general).

Figure 5.1 illustrates the ways in which the range of the norm for the formation of belief, (namely, EN1<sub>D-F</sub>) can be shaped due to epistemic filtering, whereas Figure 5.2 illustrates the ways in which epistemic filtering can shape the range of the norm for revising beliefs (namely, EN2<sub>D-F</sub>).

The different ways in which an epistemic filter operates in relation to evidential norms is represented by the different frames inside each figure. The frames are labelled after the kind of enquirers ‘inhabiting’ them. Each frame in Figure 5.1 represents a selection of propositions which count as candidate justifications,

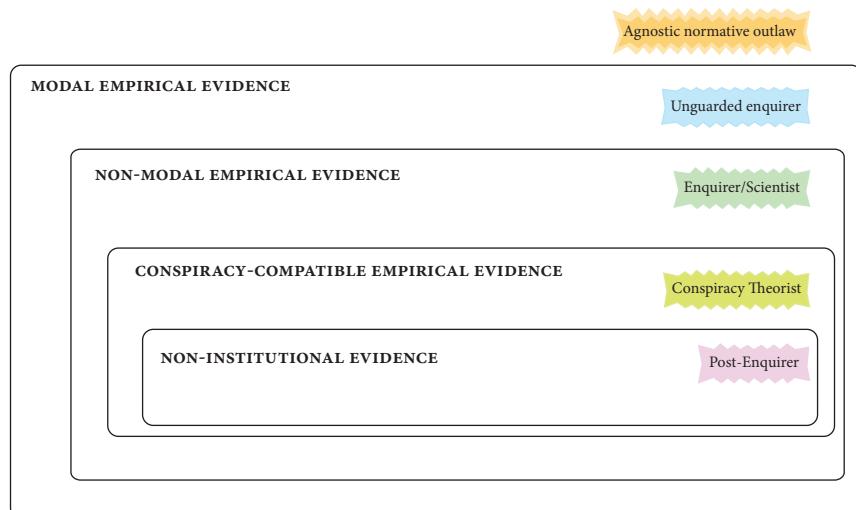


Figure 5.1 Some epistemic filters for EN1<sub>D-F</sub>

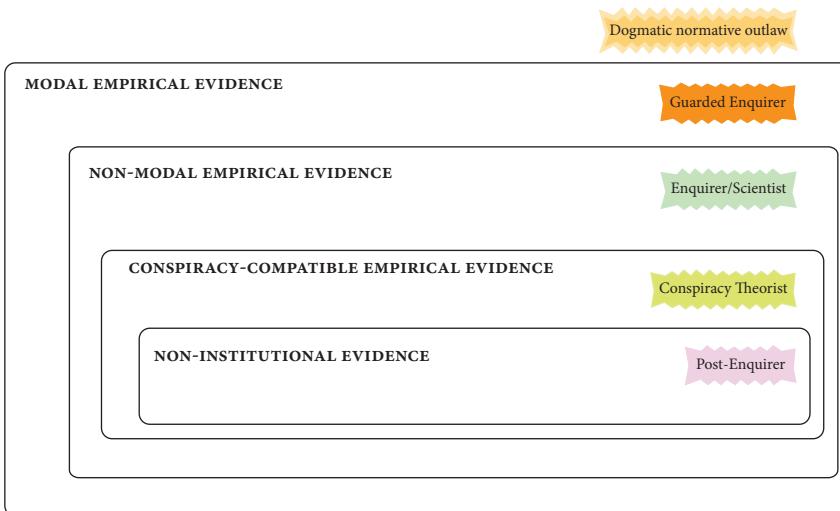


Figure 5.2 Some epistemic filters for EN2<sub>D-F</sub>

whereas each frame in Figure 5.2 represents a selection of propositions which count as candidate defeaters. A change in the set of candidate evidence determines a change in the normative extension and thus a change in what is epistemically permitted and required.

The figures illustrate some choices for delimiting the epistemic space within which enquiry (on a specific empirical subject matter) is conducted. More specifically, each frame is the result of the application of an epistemic filter. Epistemic filters select which propositions are relevant for the enquiry—i.e. which ones are eligible for being candidate evidence. In principle, an infinite number of filters are possible. These figures represent just a selection of these filters which is meant to illustrate a progression of epistemic filters that leads to the filtering mechanism characteristic of science denialism (the innermost frame). The way in which we have devised such a progression should help the reader to appreciate the striking features of science denialism and how fake news play a crucial role in giving rise to such peculiar deviance of enquiry.

Let's first consider what lies *outside* the frames. An agent inhabiting this area is not subject to the epistemic norms, this is 'the outback' of epistemic normativity. Consider Figure 5.1. An agent inhabiting the outback of EN1<sub>D-F</sub> is outside the scope of the norm for belief formation. For this character—the *agnostic normative outlaw*—no proposition is a candidate justification, therefore no proposition can have an epistemic role for forming a belief. Insofar as no belief is present for non-epistemic reasons, the agnostic normative outlaw is thus committed to be agnostic towards all empirical propositions. Let us move now to Figure 5.2. The indigenous

inquirer of this outback area is the *dogmatic normative outlaw*. For her, the revision norm never applies to what she already believes: all candidate defeating evidence is filtered out and thus no defeater is allowed. As a consequence, anything that is in the dogmatist's belief-box is epistemically unassailable—as if they were all articles of faith. If, however, the dogmatist endorses no proposition at all, she is committed to remain agnostic. We can thus summarize the condition of being a normative outlaw (whether dogmatist or agnostic) as a disjunctive condition: either she is committed to hold fast to her beliefs, come what may, or she is committed to fail to form new beliefs, come what may.

Consider, now, the outermost frames of the two figures. The one in Figure 5.1 includes among the set of candidate justifications both modal epistemic propositions and non-modal propositions. The *unguarded enquirer* inhabits this vast epistemic space. The unguarded enquirer forms her beliefs by following  $EN1_{D-F}$  and taking as input this rather big set of candidate evidence. The background assumption operating in her enquiry can be expressed by the proposition that possibility is a reliable guide to truth (the filter thus discredits other kinds of evidence, such as impossibilities, as candidate justification). Whenever the unguarded enquirer does not have the resources to refute such an epistemic possibility, the belief norm is triggered which gives her a permission for forming a belief on the basis of perception, therefore to always believe everything that is suggested by perception regarding issues related to her investigation.

In Figure 5.2, the outermost frame represents the set of candidate defeaters including both epistemic modal and non-modal propositions. These propositions can have a defeating evidential role for enquiry and thus be relevant for the norm of revision  $EN2_{D-F}$ . The character inhabiting this area is the *guarded enquirer* who admits epistemic modal propositions such as *<my perception might be deceptive>* to be defeaters for empirical propositions. The fact that I cannot disprove such an epistemic possibility triggers the revision norm. In other words, the guarded enquirer, unlike its unguarded counterpart, has the widest range of evidence available for revising her beliefs. In this sense, we can say that the guarded enquirer has rather demanding epistemic standards. The background assumption of the guarded enquirer is thus that certainty is the standard for belief and whatever is uncertain is unreliable. The practice of the guarded enquirer is therefore such that the scope of the belief revision norm  $EN1_{D-F}$  includes all epistemic possibilities open to the subject. Whenever the guarded enquirer does not have the resources to refute such an epistemic possibility, the revision rule is triggered. As a result, she is required to revise her own beliefs formed on a perceptual basis. This latter type of guarded enquirer has striking similarities with the Cartesian sceptic and the Cartesian foundationalist who seem to presuppose this kind of normative extension of enquiry.

Moving further inside by one frame, we can find there represented the range of evidential norms relative to enquiry, understood here as empirical scientific

enquiry. The *scientist* is indigenous to this area. Modal evidence<sup>12</sup> is filtered out and the only evidence in the range of epistemic norms is empirical. As mentioned earlier, this is certainly a partial characterization of the filters relevant for scientific enquirer (both descriptively and evaluatively understood). Because of this, it is not easy to make explicit what the background assumptions of scientific inquiry are. For sure, on the descriptive side we can say that scientists do not take certainty as the standard for belief and possibility as a guide for truth, and that scientific enquiry ignores some sources—e.g. fortune tellers. On the evaluative side, we can add that some sources of evidence are reliable (e.g. scientific institutions such as research centres) and some kind of evidence is not required (e.g. certainty, *pace* Descartes). How to best expand this list of background assumptions in order to fully describe and evaluate scientific enquiry is an important, albeit daunting, task which, however, lies outside the scope of this chapter.

The next frame brings us one step closer to the centre. It describes how enquiry is restricted by discrediting filters targeting evidence coming from institutionally recognized sources (e.g. research institutes, universities, journalists). The *conspiracy theorist* conducts her enquiry by implementing this kind of filter on epistemic norms: she discredits an institutional epistemic source (i.e. a certain source of evidence based on a social structure, e.g. medical scientific research) on the basis of the hypothesis that it is a deceiving institution—i.e. an institution has been created (or subsequently sabotaged) for deceiving people. Crucially, a conspiracy theory becomes a background assumption. It is thus important to be clear on what we mean here by ‘conspiracy theory’. For our purposes it is sufficient to provide a minimal characterization. A conspiracy theory is a theory that offers an alternative explanation of an event or phenomenon in opposition to the official explanation.<sup>13</sup> This explanation is based on the hypothesis that there is a conspiracy behind the scenes—often of a political nature—staged by the parties involved in the event in order to manipulate public opinion to their advantage. Often, the conspiracy mechanism is based on the belief that institutional bodies (e.g. universities, research institutes, industries, governments) intentionally produce fictitious and misleading evidence in relation to a certain event or thematic area of discussion (e.g. the shape of the earth, the damage caused by smoking, the relationship between vaccines and autism) in order to divert public attention from a hidden truth that must not leave the secret rooms of institutions. Most of the time, conspiracy theories are associated, correctly, with cases of alternative explanations of known facts that are specious and misleading. To cite a few well-known examples of pernicious conspiracy theories, consider the New World Order

<sup>12</sup> Modal evidence is the evidence expressed by propositions of the form ‘it is epistemically possible that p’.

<sup>13</sup> See Coady 2003.

theory, the Protocols of the Elders of Zion, the 9/11 attacks, etc.<sup>14</sup> However, the hypothesis that some conspiracy is acting in the background of an official ‘institutional’ explanation of a certain event should not be considered always unjustified and untrue. In fact, there have also been conspiracy theories that turned out to be true, such as the case of the multinational tobacco companies.<sup>15</sup> A conspiracy theory can thus be true and justified.<sup>16</sup>

In some (plausibly most) cases, the hypothesis of a conspiracy is unlikely to be true and, moreover, it may also be designed to protect an enquiry from being reformed. When epistemic norms are filtered by background assumptions whose content is such a conspiracy theory we say that we have a case of *post-enquiry*.<sup>17</sup> In a post-enquiry, filters select the candidate justifications in such a way as to create an enquiry with low epistemic standards since the most sophisticated evidence coming from scientific institutions is filtered out. Low epistemic standards decrease the degree of maximization of truth and of minimization of falsity since the enquirers have less chances to get true beliefs and avoid false ones if their available evidence does not point effectively to what is the case. Thus, post-enquiry scores low in the fulfilment of the alethic telos.

An effective way to explain how a conspiracy theory engenders a post-enquiry is to take it to belong to a special class of conspiracy theories identified by Quassim Cassam. These conspiracy theories have the following six features which we call *C-features*: (i) they are speculative, (ii) contrarian, (iii) esoteric, (iv) amateurish, (v) premodern, and (vi) self-sealing (Cassam 2019: 16–31, 92–9). Being speculative means that they are based on conjectures lacking solid evidence—e.g. there is no solid evidence that Stanley Kubrick took part in a conspiracy for faking the moon landing (Hess 2019). They are contrarian because they are against the obvious explanation of events—e.g. although al-Qaeda took responsibility for the 9/11 attacks, the conspiracy theories claim that they were an inside job. They are esoteric because they make use of bizarre explanations such as the flat-earther’s thesis that the boundaries of the earth are guarded by a secret governments’ army. Further, they are amateurish because those who argue for these theories are typically not experts in the relevant field. This last feature is clearly exemplified in the two case studies that we analyse in Sections 5 and 6: anti-vaccine ‘experts’ are not real experts—they are not virologists but, at best, general practitioners; flat earthers are not physicists, geologists, nor astronomers but, at best, amateur scientists. Another feature is being pre-modern which means expressing a world view where “complex events are capable of being controlled by a small number of people acting in secret, and this is what gives these events a

<sup>14</sup> For a list of conspiracy theories, see [https://en.wikipedia.org/wiki/List\\_of\\_conspiracy\\_theories](https://en.wikipedia.org/wiki/List_of_conspiracy_theories).

<sup>15</sup> For details on this case, see Oreskes & Conway 2010.

<sup>16</sup> This is controversial: Coady 2007 argues that conspiracy theories need not be irrational.

<sup>17</sup> Nguyen 2020s definition of echo chamber partly captures our notion of post-enquiry and its resistance to reform.

deeper meaning" (Cassam 2019: 26). The fact that certain conspiracy theories have these features explains why they are unlikely to be true. Therefore, any enquiries whose background assumptions are based on conspiracy theories with *C-features* (i)–(v) is bound to score badly on the degree of fulfilment of the alethic telos. Finally, Cassam adds that these theories are self-sealing—*C*-feature (vi)—in that believers in these theories belong to a "belief bubble" whose assumptions have a special role: "anyone who questions [them]... is excluded from the bubble" (Cassam 2019: 96). This last feature is connected with our notion of reformability in that it indicates a strong resistance of an enquiry to be reformed: many sources that may provide defeating evidence for background assumptions are neutralized because excluded by the epistemic filters.

To sum up, whenever an enquiry is regulated by an epistemic filter based on background assumptions involving a conspiracy theory with the *C*-features, it engenders a normatively aberrant kind of enquiry, which we label *post-enquiry*. More precisely, featuring a background assumption based on a conspiracy theory which satisfies *C*-features (i)–(v) is sufficient for an enquiry to be normatively aberrant and thus a *post-enquiry*. If such a conspiracy theory also satisfies *C*-feature (vi), we have a form of *post-enquiry* which is particularly recalcitrant to being reformed and thus normatively even more aberrant.

Let us briefly discuss, at this point, a potential challenge: couldn't the traditional notion of undermining defeater explain science denialism? Instead of speaking of epistemic filters and normative aberrance, perhaps we can simply say that science denialists unknowably entertain defeated underminers. Normative variance would then be an unnecessarily complex framework that is explanatory disposable. This objection misses an important point related to the impact that our proposal has on epistemic normativity. In our framework we say that a science denier is permitted to hold her own theses (e.g. that earth is flat), no matter how bizarre we take them to be, once she is conducting enquiry based on certain epistemic filters. The undercutting theorists would instead say that there is no objective sense in which science deniers are permitted to hold their views. Of course, in our framework, the permissibility of holding bizarre views within a denialist enquiry does not mean that science denialism is epistemically irreproachable. On the contrary; on our view, science denialism is epistemically criticizable since it uses as background assumptions a typology of epistemically bad conspiracy theories which give rise to normatively aberrant enquiries. However, unlike the undercutting theorist, we allow that science denialists genuinely satisfy some normative requirements for enquiry. The undercutting theorist, on the contrary, holds that, at best, the science denialist thinks to satisfy normative requirements.<sup>18</sup>

<sup>18</sup> We would like to thank Thomas Grundmann for suggesting this critical remark.

That being said, the thesis that we will argue for in the remaining part of the chapter is that two paradigmatic examples of science denialism are clear cases of post-enquiries. More specifically, we advance two connected theses. The first is that those epistemic filters which are characteristic of flat-earth enquiry and anti-vaxx enquiry engender a normatively aberrant form enquiry—i.e. a post-enquiry. The second is that these epistemic filters are based on a conspiracy theory which is conceived and spread following the mechanics and dynamics of fake news. We argue for the first point both with general epistemological considerations and with specific considerations related to the case studies. Concerning the second point, we believe that fake news plays a crucial role in science denialism in making an enquiry aberrant when they have a role in the construction of background assumptions. Since the aim of this chapter is not that of providing an insightful characterization of what counts as fake news, we are fine with Rini's characterization:<sup>19</sup>

A fake news story is one that purports to describe events in the real world, typically by mimicking the conventions of traditional media reportage, yet is known by its creators to be significantly false, and is transmitted with the two goals of being widely re-transmitted and of deceiving at least some of its audience.<sup>20</sup>

The important feature captured by Rini's characterization is that fake news is generally transmitted with the intention of deceiving. In our normative framework for understanding the phenomenon of science denialism, fake news can play a double epistemic and normative role. First, an allegation of promoting fake news made against institutional sources can be taken as a background assumption giving rise to a discrediting filter for any evidence that comes from those sources. Fake news may thus be an important element in the construction of some conspiracy theories that lie at the core of pseudo-scientific theories that are promoted as the (true) alternatives to official scientific theories. Second, fake news can be constructed with the intention of providing direct 'scientific' evidence supporting certain views contrary to what is told by institutional science—as illustrated by Wakefield's paper in the *Lancet*, then retracted by the journal.

## 6. First Case Study: Flat Earth

We now employ our normative framework to analyse the first case study, namely flat-earth enquiry. Flat-earthers believe that our planet is flat. As crazy a view as it

<sup>19</sup> For a critical overview of the current debate, see Jaster & Lanius, Chapter 1 in this volume.

<sup>20</sup> Rini 2017.

may sound to most of us, it has several acolytes—according to a recent survey approximately 2 per cent of US citizens declare belief that the earth is flat.<sup>21</sup> Moreover, flat-earthers conduct a kind of enquiry which is to some extent similar to scientific enquiry: they devise and execute their own experiments, they attempt alternative explanations to those offered by our best scientific theory, and, interestingly, they have developed their own specialized cartography. In their own deviant way, they care about evidence.

However, it is clear that there is something epistemically vicious about the kind of enquiry that flat-earthers conduct. Such an enquiry, we argue, is normatively aberrant and it engenders a post-enquiry. Let us begin by mentioning the fact that flat-earthers assume in the background of their enquiry a global conspiracy theory based on the conviction that governments and scientific institutions (virtually all of them!) have been lying to people about the shape of the earth, in fact promoting the fake news that our planet is approximately spherical in shape. On the contrary—flat-earthers claim—the truth is that the earth is flat and is covered by a giant dome. This conspiracy theory plays the role of a background assumption in flat-earth enquiry. To see why, it is useful to reflect on how flat-earthers react to the evidence provided by scientific institutions. The basic dialectical move they make is to deny the truthfulness of such evidence by making allegations against governments and scientific institutions of producing fake news in order to deceive people and to hide the truth. Let us consider a few examples of this. For instance, when flat-earthers are presented with visual evidence from space that the earth cannot be flat—e.g. the evidence provided by NASA and satellite pictures—they generally react by saying that such evidence is fake—that it has been manufactured by scientific institutions with the help of governments. A second example concerns their views on the Apollo 1969 expedition and the first landing on the moon by Neil Armstrong. According to flat-earthers, the entire expedition was a fake produced by NASA in Hollywood with the help of Stanley Kubrick. Thus, governments and NASA have been producing fake news about the moon landing. A third interesting case concerns a very simple visual experiment that was conducted on 10 June 2018 by activists from the Independent Investigation Group (IIG) at Salton Sea in California, and witnessed by a group of flat-earthers.<sup>22</sup> The purpose of the experiment was to demonstrate the curvature of the earth across the Salton Sea, by rather simple visual evidence. The experiment consisted in watching a boat, moving from one shore to the opposite one, with a 3m x 2m target with horizontal stripes installed on the top of the boat. As predicted by the scientific theory, at a distance of approximately 5km from the departure shore,

<sup>21</sup> See <https://today.yougov.com/topics/philosophy/articles-reports/2018/04/02/most-flat-earthers-consider-themselves-religious>.

<sup>22</sup> See <https://skepticalinquirer.org/2018/11/the-salton-sea-flat-earth-test-when-skeptics-meet-deniers>. An analogous experiment had been performed in a controversy over flat-earthism; see Schadewald 2015, chs 1–3.

one of the horizontal stripes was lost from sight, clearly indicating the curvature of the earth. In reply, the group of flat-earthers simply denied the validity of the whole demonstration—though, only after they fully understood that such a demonstration went against their theory—and their reaction was followed by accusations of promoting fake news and hoaxes.

These three cases show that a certain conspiracy theory is held as a background assumption in conducting flat-earth enquiry. Part of the content of this conspiracy is the allegation made against institutional sources of evidence (like NASA) of producing and promoting fake news. This conspiracy theory has the C-features (i)–(v) that makes it unlikely to be true, thus scoring low in the degree of fulfilment of the alethic telos. This conspiratorial background assumption induces an epistemic filter which excludes all evidence (both supporting and defeating evidence) coming from scientific and governmental institutions from being candidate evidence in the range of the epistemic norms. Moreover, the range of epistemic sources discredited by the filter is so vast—since it discredits all the evidence sourced in scientific and governmental institutions—that it makes flat-earth enquiry hardly reformable, thus satisfying also C-feature (vi). In this respect, the enquiry conducted by flat-earthers is a clear case of post-enquiry which represents a significant normative aberration from standard scientific enquiry.

## 7. Second Case Study: Anti-Vaxxer

The anti-vaccine movement is another, perhaps more representative, case of science denialism where we can clearly identify features of post-enquiry. In what follows, we draw attention to some features of this movement that show how their enquiry is a case of post-enquiry. As with the previous example, fake news plays a key role in this explanation.

The anti-vaccine movement makes use of a conspiracy theory. According to this theory, institutions such as the Centers for Disease Control and Prevention intend to hide the link between vaccines and severe health conditions, such as autism, for reasons connected to a secret agenda.<sup>23</sup> Such a conspiracy theory gives rise to epistemic filters which discredit the reliability of the institutions connected to vaccine research and innovation as sources of evidence. Clearly, the sort of conspiracy theory advocated by anti-vaxxers satisfies C-features (i)–(v). A useful example that highlights the role of this conspiracy theory in the anti-vaxxers' enquiry is provided by the tragic case of a 12-year-old American girl who died after having taken a vaccine. In this case, a single (and, to the best of our knowledge, random) correlation between two events has been postulated, without

<sup>23</sup> Moreover, anti-vaxxers claim that vaccines are part of a (secret) depopulation agenda; see Adams 2015.

any extra testing, as establishing a causal connection between these events. In fact, anti-vax sites have been claiming that the death of the girl was caused by her taking the vaccine (Health Impact News 2014). However, this causal explanation has been proven false after the autopsy revealed that the cause was an overdose of antihistamine (Johnson 2014). Interestingly, at least some anti-vax sites have ignored this further piece of evidence, thus continuing to claim that there was a causal connection between her death and her taking the vaccine. In this case, we can clearly see that fake news is playing the role of epistemic filters, precluding the new evidence provided by the autopsy from being considered relevant evidence—indeed, as evidence defeating the alleged causal explanation. Moreover, in addressing this case, anti-vax sites (Health Impact News 2014) have exploited this piece of news for promoting the vaccines conspiracy theory in order to fuel further discredit on the relevant scientific institution—i.e. the Centers for Disease Control and Prevention. This latter point highlights another interesting feature of the kind of post-enquiry conducted by anti-vaxxers: their post-enquiry uses the conspiracy theory to filter out counterevidence for its own empirical claims thus having an easy way for justifying these claims; at the same time, whenever an anti-vax claim is justified in this way, this fact is used as evidence for the conspiracy theory that constitutes the epistemic filter.

However, it has to be noticed that the range of the epistemic filters regulating the anti-vax enquiry is narrower than the one of flat-earthers since it targets only institutions working on vaccines. As a consequence, the enquiry conducted by anti-vaxxers scores better than the flat-earth enquiry with respect to the possibility of reform. That being said, we explain how the epistemic filters generated by the anti-vaxxers' conspiracy theory has the effect of lowering their epistemic standards, thus making their enquiry score badly on the degree of fulfilment of the alethic telos.

It is clear that anti-vaxxers, in conducting their enquiries, don't use the best available evidence and methodologies. The epistemic filters regulating their enquiry allow them to do so without confronting scientific counterevidence and without being under the requirement of using more sound methodologies. Since anti-vaxxers do not trust scientific institutions, which are deemed unreliable sources of evidence, they must rely on alternative sources and methodologies. Due to the fact that the evidence on vaccines coming from institutional sources is filtered out, scientific methodologies are ignored and alternative defective evidence and methodologies are employed. Consider, first, the effect that this normative aberration has on evidence. Anti-vaxxers typically use anecdotal evidence for justifying general claims. This allows anti-vaxxers to jump from simple correlations to systematic causal explanations: simple correlations count as evidence for supporting the thesis that a causal explanation is instantiated (Howard & Reiss 2018: 200–3).

Another defective methodology typically employed by anti-vaxxers is to criticize a solution to a problem if it is not bulletproof: if the solution is to some extent short of being the perfect solution to the problem at hand, it is worthless (Howard & Reiss 2018: 199–200). So, for example, if vaccines are not 100 per cent safe then we should have a substantive critical attitude towards them—when of course we know that no medicine is 100 per cent safe. This way of opposing vaccines can be understood in our normative framework as lowering the standards for defeaters: less than 100 per cent safe counts as a defeater for the thesis that vaccines are efficacious (Ji 2014). What is particularly bizarre in this context is the application of a sort of epistemic double-standard. When, in fact, the pro-vaccine theory is assessed, standards are raised and enquiry enters the Cartesian area, whereas when the anti-vax theory is assessed, standards are lowered. A related methodological defect is to make a possibility salient—e.g. that vaccines do not cause cancer—and then ask for evidence in favour of it. If not enough evidence is available, then it is claimed that the negation of the possibility—i.e. that vaccines do cause cancer—is deemed to be justified (Wolfson 2014). In this case we can see a very peculiar move made within the normative space of enquiry: when a pro-vaccine claim is considered the anti-vaxxer pretends to use modal evidence for the revision norm raising the standards to the Cartesian level—depicted in Figure 5.2 as the guarded enquirer. Once the challenge to the pro-vaccine thesis is not met, the anti-vaxxer switches to the norm for the formation of belief (EN<sub>1D-F</sub>) using modal evidence as the unguarded enquirer does and thus lowering the standards—to exemplify, until the possibility that vaccines cause cancer is not excluded, the very possibility justifies that proposition.

A further characteristic feature of the methodology of anti-vaxxers is the use of experts with dubious credentials. Often, these pseudo-experts have cynical economic interests since they sell alternative medicines without having enough evidence of their efficacy and reliability, and promote conspiracy theories according to which outbreak of infectious illnesses such as Ebola have been scripted by governments to enforce mass vaccination (Howard & Reiss 2018: 204). Moreover, Howard and Reiss provide a wide range of examples of pseudo-expert anti-vaxxers with no record of scientific publications claiming to have done extensive research (Howard & Reiss 2018: 203–6). This is a clear case of fake news about expertise that plays a crucial epistemic role in promoting the low standards of post-enquiry: the epistemic filters adopted in the post-enquiry conducted by anti-vaxxers discredit the institutional experts and promote these pseudo-experts as reliable epistemic sources.

Finally, let us emphasize the fact that fake news plays a key role for anti-vaxxers in promoting pseudo-scientific evidence. Wakefield's paper is a case in point. Andrew Wakefield published with some collaborators in 1998 a paper in the prestigious journal *Lancet* in which he claims that there is a connection between vaccination and autism in children (Wakefield et al. 1998). After extensive

criticism in the literature, the paper was eventually retracted by the *Lancet* twelve years later. In relation to this case, the role of the double-standard policy is worth noticing. The Wakefield study has been mentioned as providing evidence for the actual unsafety of vaccines whereas at the same time papers against this thesis are discredited by using the conspiracy theory. Moreover, the fake news promoting Wakefield studies by defending their scientific credentials has been circulating among anti-vaxxers even *after Lancet's* retraction. To justify this claim, Andrew Wakefield is presented as a scapegoat, where a conspiracy has made the *Lancet* to retract (Health Impact News 2017).

## 8. Conclusions

We have provided an epistemological model purported to describe some ways in which the normative extension of standard scientific enquiry may be deviated, giving rise to a normatively aberrant form of enquiry which we have called *post-enquiry*. We have put our model to work by analysing two case studies of science denialism, namely flat-earth enquiry and anti-vax enquiry, and we have argued that they are cases of post-enquiry. In so doing, we have illustrated the role that fake news plays in relation to the normative deviances occurring in these examples.

Let us emphasize that our framework of normative deviance captures two features which lie at the core of science denialism. The first is that typical cases of science denialism, as those we have dealt with in this chapter, are cases of enquiry-based practices—and not some normatively arbitrary and rationally unintelligible practices. The second feature is that the kind of enquiry proper of science denialism has specific epistemic features that make it normatively aberrant. Science denialism is a case of post-enquiry: a practice whose epistemic norms have been bent by epistemic filters based on conspiracy theories that are unlikely to be true and, in some cases, hardly reformable. Fake news plays a double role in the mechanics of post-enquiry. First, they are used to implement the discrediting function of the epistemic filter on the institutional sources of evidence; second, they play a constructive part for post-enquiry in taking part in the production of (pseudo-)scientific explanations alternative to the ones of official science.<sup>24</sup>

<sup>24</sup> Parts of this chapter have been presented at several workshops and seminars: "The Vagaries of Normative Deviance in Enquiry", Seoul National University; "Relativism Workshop", Yonsei University; "Philosophical Analysis" research seminar, University of Pavia; "Seric" research seminar, University of Bologna; the research seminar at the Centre for Engaged Philosophy and Crick Centre, University of Sheffield; "Logic and Epistemology" research seminar, University of Bonn; and "Post-truth: Philosophy, Sociology, and Media Studies", University of Bologna. We are especially grateful for comments to Elke Brendel, Massimiliano Carrara, Massimo Dell'Utri, Matti Eklund, Joshua Forstenzer, Steve Fuller, Katherine Furman, Axel Gelfert, Thomas Grundmann, Jinho Kang, Martin Kush, Nikolaj Jan Lee Linding Pedersen, Tommaso Piazza, Alessandra Tanesini, and Giorgio Volpe. Moreover, we

## References

Adams, M. (2015). [https://www.naturalnews.com/049669\\_vaccine\\_injury\\_depulation\\_agenda\\_deadly\\_side\\_effects.html](https://www.naturalnews.com/049669_vaccine_injury_depulation_agenda_deadly_side_effects.html).

Cassam, C. (2019). *Conspiracy Theories*, Cambridge: Polity Press.

Coady, D. (2003). Conspiracy Theories and Official Stories. *International Journal of Applied Philosophy* 17(2): 200–1.

Coady, D. (2007). Are Conspiracy Theorists Irrational? *Episteme* 4(2): 193–204.

Ferrari, F. (2018). Normative Alethic Pluralism. In J. Wyatt et al. (eds), *Pluralism in Truth and Logic*, London and New York: Palgrave Macmillan.

Ferrari, F. & Moruzzi, S. (2020). *Verità e Post-verità. Dall'Indagine alla Post-indagine*, Bologna: 1088 Press.

Garwood, C. (2007). *Flat Earth: A History of an Infamous Idea*, London: Macmillan.

Health Impact News (2014). <https://healthimpactnews.com/2014/gardasil-vaccine-one-more-girl-dead/>.

Health Impact News (2017). <https://healthimpactnews.com/2017/dr-andrew-wakefield-fraud-or-scapegoat/>.

Hess, A. (2019). They Kinda Want to Believe Apollo 11 Was Maybe a Hoax, *The New York Times*, 1 July 2019. <https://www.nytimes.com/2019/07/01/science/moon-landing-hoax-conspiracy-theory.html>.

Howard, J. and Reiss, R. D. (2018). Vaccine Movement: A Litany of Fallacy and Errors. In A. B. Kaufman & J. C. Kaufman (eds), *Pseudoscience: The Conspiracy Against Science*, Cambridge, MA: MIT Press, 195–220.

James, W. (1897). *The Will to Believe and Other Essays in Popular Philosophy*, London: Longmans, Green & Co.

Ji, S. (2014). <https://www.greenmedinfo.com/blog/should-you-trust-daily-beast-about-vaccines>.

Johnson, M. (2014). <http://archive.jsonline.com/news/for-girl-a-day-that-started-with-a-routine-vaccination-ended-in-her-death-b99326702z1-270410121.html>.

Kaufman, A. B. & Kaufman, J. C. (2018). *Pseudoscience: The Conspiracy Against Science*, Cambridge, MA: MIT Press.

Kelp, C. (2014). Two for the Knowledge Goal of Inquiry, *American Philosophical Quarterly*, 51(3): 227–32.

would like to express our gratitude to the editors of this volume for their support, and to two anonymous referees for their extensive and insightful comments which have helped us significantly in improving and clarifying our ideas. Last, we would like to mention that part of the research for this chapter has been carried out within the remit of two projects hosted by the Department of Philosophy, Sociology, Education and Applied Psychology, University of Padua – “Polarization of irrational collective beliefs in post-truth societies. How anti-scientific opinions resist expert advice, with an analysis of the anti-vaccination campaign” funded by CARIPARO and led by Massimiliano Carrara; “Epistemic virtues and knowledge. Investigating and promoting effectiveness in inquiry” funded by a BIRD 2020 departmental grant (SPOL\_BIRD2020\_01) led by Giuseppe Spolaore.

Lynch, M. (2019). *Know-it-all Society*, New York: Liveright.

Nguyen, C. Thi (2020). Echo Chambers and Epistemic Bubbles. *Episteme* 17(2): 141–61.

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

Rini, Regina (2017). Fake News and Partisan Epistemology. *Kennedy Institute of Ethics Journal* 27(S2): 43–64.

Schadewald, B. (2015) The Plane Truth, <https://www.cantab.net/users/michael.behrend/ebooks/PlaneTruth/pages/index.html>.

Shah, N. & Velleman, J. D. (2005). Doxastic Deliberation. *Philosophical Review* 114(4): 497–534.

Stanley, J. (2005). *Knowledge and Practical Interests*, New York: Oxford University Press.

Steinberger, F. (2017). The Normative Status of Logic, *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition), Edward N. Zalta (ed.), <https://plato.stanford.edu/archives/spr2017/entries/logic-normative/>.

Steinberger, F. (2019). Accuracy and Epistemic Conservatism, *Analysis*, 79(4): 658–69.

Wakefield, A. J. et al. (1998). [RETRACTED]: Ileal-Lymphoid-Nodular Hyperplasia, Non-Specific Colitis, and Pervasive Developmental Disorder in Children. *The Lancet*, 351(9103): 637–41.

Williamson, T. (2000). *Knowledge and Its Limits*, Oxford: Oxford University Press.

Wolfson, J. (2014). <https://www.wolfsonintegrativecardiology.com/case-dismissed/>.

# 6

## Facing Epistemic Authorities

### Where Democratic Ideals and Critical Thinking Mislead Cognition

*Thomas Grundmann*

In 2016, “post-truth” was declared to be the word of the year by the *Oxford English Dictionary*. The *OED* defined post-truth as “relating to or denoting circumstances in which objective facts are less influential in shaping public opinion than appeals to emotion and personal belief.” Accordingly, in post-truth circumstances, truth-makers have less influence on public opinion than factors that are irrelevant to truth. Of course, the facts never fully determine public opinion, not even under ideal conditions of a scientifically oriented public discourse. Even rational beliefs are sometimes false, science itself is not immune from errors, and slips can never be ruled out completely. However, in post-truth times, the detachment of public opinion from truth is much more radical. Under these *pathological* circumstances, public opinion is systematically and mainly shaped by truth-unrelated factors. This is an epistemic disaster; and it is clear from the outset that the majority of users of the term “post-truth” have this epistemically bad evaluation in mind. The *OED* definition mentions two truth-unrelated factors that may shape public opinion: emotion and personal belief. Of course, emotions are not always in conflict with what is true. For example, we often fear what is really dangerous to us. Something similar applies to beliefs. However, in post-truth circumstances, emotions and beliefs will shape public opinion, no matter whether they correspond to the truth or not. Even worse, they do so when it is obvious to everyone that emotions and beliefs do not correlate with truth. This becomes transparent when, e.g., right-wing politicians base their arguments on people’s fear of refugees, even in areas where there are literally no such refugees. Bullshit is another phenomenon of this kind (cf., Frankfurt 2005). When people utter bullshit they simply do not care whether what they utter is true or not. Often, they treat their utterance simply as an authentic expression of their personal mental lives, even if what they utter is completely unsupported by or even in conflict with their evidence.

The two truth-unrelated factors mentioned by the *OED* definition are factors that result in *irrational* public opinion. Forming opinions without any evidence of their truth is clearly irrational. However, there seem to be many more factors that

lead public opinion away from the truth in systematic ways.<sup>1</sup> Interestingly, only some of these factors involve epistemic irrationality in the believer. Among the irrational factors, *individual biases* (e.g., overconfidence,<sup>2</sup> confirmation biases,<sup>3</sup> narrative biases,<sup>4</sup> epistemic vices<sup>5</sup> or affective biases<sup>6</sup>) and *social biases* (group think<sup>7</sup> or social identity signaling<sup>8</sup>) play a dominant role. However, when the environment is such that huge bodies of fabricated or misleading evidence are presented to the public, or when only selective evidence is disclosed, this—most likely—will result in a general public opinion that is radically detached from the truth, even if the public uses this evidence in a fully rational way. *Agents of misinformation* use a variety of different strategies to manipulate the available evidence in such a way that the public is radically misguided even when it closely follows the standards of rationality. These agents of misinformation can fabricate fake news;<sup>9</sup> they can hide relevant evidence completely or present only parts of the relevant evidence;<sup>10</sup> they can swamp the publicly accessible body of evidence with counterevidence, such that well-established views seem to no longer be defensible;<sup>11</sup> they can direct public attention exclusively to evidence that induces doubts about official sources of evidence;<sup>12</sup> and they may even raise doubts about the epistemic norm of assertion itself. Then, there are *structural features* of communication that lead—Independently of anyone's bad intentions—to a selective disclosure of evidence that may mislead public opinion radically. What I have in mind here are phenomena such as information cascades<sup>13</sup> (people hiding relevant information because they either wrongly believe that it is irrelevant or fear social pressure), filter bubbles (when internet algorithms confront the user exclusively with information that fits the content of her previous searches),<sup>14</sup> journalistic practices (e.g., the tendency to prefer dramatic stories results in the illusion of general decline; the norm of balanced reporting amplifies the felt

<sup>1</sup> For an excellent survey of these explanatory factors, see McIntyre 2018.

<sup>2</sup> Cf., Dunning 2005; Kahneman 2011: part III. <sup>3</sup> Cf., Kahneman 2011.

<sup>4</sup> Brotherton 2015 argues that conspiracy theories are often the results of our natural tendency to give narrative explanations.

<sup>5</sup> Cf., Cassany 2019. <sup>6</sup> Cf., Kahneman 2011: ch. 13. <sup>7</sup> Cf., Sunstein 2006: ch. 3.

<sup>8</sup> Cf., Lynch 2019.

<sup>9</sup> According to the majority view, fake news results either from a deceptive intention or an attitude of indifference to truth. See Gelfert 2018; Jaster & Lanius 2018.

<sup>10</sup> Jaster & Lanius 2018.

<sup>11</sup> This is how I interpret Kellyanne Conway's reference to "alternative facts" when she defended the claim that Trump had more visitors at his inauguration speech than Obama, on NBC in 2017. I take it that she did not rely here on ontological relativism, but was just claiming that there were pieces of evidence (namely reports from a massively overcrowded subway system) that were in conflict with what the official pictures of the event suggested.

<sup>12</sup> According to Oreskes & Conway 2010, these strategies were used by lobbyists from the tobacco industry and the oil industry to discredit scientific findings about the harmful effects of these industries.

<sup>13</sup> Cf., Sunstein 2006: ch. 3.

<sup>14</sup> For the general worry, see Pariser 2011. Recent empirical studies suggest, however, that these effects are not significant. Cf., Hannack et al. 2013; Haim et al. 2017; Curtois et al. 2018.

significance of fringe views in the public),<sup>15</sup> or specific features of social media (e.g., the option of quick likes for interesting news facilitates the vast distribution of false news<sup>16</sup>). Finally, deeply entrenched *ideologies* may also explain the post-truth condition, at least in part.<sup>17</sup> When postmodernists deny the existence of truth, or relativists defend the idea of alternative facts, or when the *zeitgeist* dictates that assertions should be treated as authentic expressions of the self rather than truth claims that are eligible for criticism, this may lead to an excessive public tolerance of false opinions.

As this list impressively demonstrates, post-truth circumstances have many (often complementary) explanations. Whether some of these sources are completely new phenomena that are closely tied to contemporary communication technologies is controversial.<sup>18</sup> It is also not fully clear whether today's post-truth condition is worse than it was in the past.<sup>19</sup> In this chapter, I will leave these questions open. Here I want to argue that the above list is not exhaustive. One widely neglected factor is missing from the list. I will argue, perhaps surprisingly, that our deeply entrenched Enlightenment principles of unrestricted critical thinking and democratic reasoning also play an important role in explaining our current post-truth condition.

Let me first introduce these Enlightenment principles. I start with the *Principle of Unrestricted Critical Thinking* (PUCT). In his essay "An Answer to the Question: What is Enlightenment?" Kant famously writes:

Have the courage to use your own understanding [...] without another's guidance. [...] It is so comfortable to be a minor. If I have a book that thinks for me, a pastor who acts as my conscience, a physician who prescribes my diet, and so on—then I have no need to exert myself. I have no need to think [...].

In this passage, Kant ironically condemns people who rely on moral or medical experts in their judgments rather than thinking for themselves. Lynch (2016: 38) strongly disagrees, writing:

When, for example, we aren't an expert on something ourselves, we seek advice from those who say they are. But if we are wise, we also get evidence of that person's expertise: references, degrees, or word of mouth. Moreover, we look for them to explain their opinions to us in ways that make sense given what we know.

<sup>15</sup> Cf., McIntyre 2018: ch. 4.

<sup>16</sup> Cf., Vosoughi et al. 2018.

<sup>17</sup> Cf., McIntyre 2018: ch. 6.

<sup>18</sup> For discussion, see, e.g., Duthil Novaes & de Ridder, Chapter 7 in this volume.

<sup>19</sup> Uscinski & Parent 2014: 110 present some evidence from letters to the editors of *The New York Times* (1890–2010) that suggests that conspiratorial talk has recently declined.

In contrast with Kant, Lynch clearly recommends reliance on experts, with the caveat that we should not trust experts *blindly*. However, even Lynch claims that the expert's judgment must make sense from our lay perspective in order for it to be acceptable. Although Kant and Lynch give different epistemic weight to expert judgments (Kant: zero, Lynch: some significant weight), both agree that the weight of one's own judgment should never be reduced to zero. This shared view is expressed by what I call the *Principle of Unrestricted Critical Thinking* (PUCT). Here is what it claims:

(PUCT) Whenever you rationally consider the truth of some proposition  $p$ , you must never stop using your own reasons for  $p$ .

This principle neither excludes the judgments of others from counting as evidence, nor gives equal weight to the judgments of laypeople and experts. But it defends the rational significance of the agent's judgment, no matter whether they are an expert or a layperson.

Compare this with the *Principle of Democratic Reason* (PDR):

(PDR) Whenever you rationally consider the truth of some proposition  $p$ , you must not exclude, marginalize or silence any rational thinker's reasoning (including your own one) about  $p$ .

(PDR) requires that the judgments of all rational thinkers are considered seriously and that nobody should simply defer to the judgments of others. Alongside many other philosophers,<sup>20</sup> Habermas (1991: 132) seems to be an advocate of this principle:

No one can honestly join a discussion, unless she presupposes a context of discourse in which public access, equal participation [...] are—at least in principle—guaranteed. The participants can aim to convince each other only if they pragmatically implicate that their acceptance or denial is determined by nothing but the force of the better argument. [My translation]

What is the relation between (PUCT) and (PDR)? (PDR) entails (PUCT), but not vice versa. This is because (PDR) is a generalization of (PUCT), the latter being a special case of the former. Whereas (PUCT) rules out ignoring the reasoning of the epistemic subject, (PDR) rules out ignoring anyone's reasoning when reasons are rationally aggregated. Since (PDR) is the more general principle, we will focus on its implications. (PDR) prohibits the exclusion of anyone's reasoning, when

<sup>20</sup> Cf., e.g., Landemore 2013.

reasons are aggregated. It does *not* imply that everyone's judgment has equal weight (the Principle of Epistemic Equality), but it does imply that one must not simply defer to experts (thereby giving zero weight to one's own reasoning) and that experts must not ignore laypeople's judgments completely. According to (PDR), expert beliefs must always be critically checked against what rational believers take to be the plausible view. In the extreme case, an expert judgment that appears outrageous or crazy to laypeople may be rejected because it fails this test.

In this chapter, I will defend the following three claims: first, I will argue that (PDR) and (PUCT) are both in conflict with what is required from a purely epistemic point of view; second, I will demonstrate that following the advice of (PDR) can have epistemically dangerous consequences, by facilitating the spread of conspiracy theories and eroding our trust in experts; and, third, I will show in detail that one alternative to (PDR), namely the Preemption View (PV), restricts the scope of critical reasoning and inclusive deliberation in plausible ways.<sup>21</sup>

I will proceed as follows. In Section 1, I will argue that (PDR) is in conflict with the most reasonable view about how one should rationally respond to the judgment of epistemic authorities, i.e., the Preemption View (PV). I will present two arguments for this view and engage with some recent objections from Lackey (2018). In Section 2, I will show how (PDR) can lead to the acceptance of conspiracy theories. In Section 3, I will address three deep worries that Lackey (2018) expresses about the view that results from restricting (PDR) in line with (PV). I will argue that these worries can all be dispelled. I conclude with some general remarks.

## 1. What's Wrong with (PDR) from an Epistemic Point of View

Epistemic abilities are unequally distributed in society. Not everyone has the same cognitive competences. Some people stand out by being much more reliable than average. This epistemic superiority is relative to a reference class of people; and it is always relative to a domain of expertise. No one is an expert on everything. Epistemic superiority is the product of two independent factors: the available body of evidence and one's reasoning competences.<sup>22</sup> Some people have a more reliable judgment than others on a specific topic because they are equipped with more relevant evidence. This is, e.g., true for the judgment of an eyewitness in comparison with someone who was not present at the time and place in question.

<sup>21</sup> Of course, there are other alternatives to (PDR) such as individualism, i.e., the view that testimony does not provide us with any epistemic reasons.

<sup>22</sup> Whether one possesses the relevant evidence or the reasoning competences will, of course, depend on further factors such as training or opportunity.

Some people have a more reliable judgment than others because they are more clever in drawing rational inferences from the shared body of evidence. This is, e.g., the case when Sherlock Holmes outperforms Watson in his judgments about who was the murderer. Of course, a lack of evidence can be compensated, or even overcompensated, by superior reasoning competences, and vice versa.

Given this background, epistemic authorities are people who are not only the epistemic superiors of specific other people, but who also do well enough objectively. And they are known to have these properties.<sup>23</sup> We can turn this into the following definition:

A is an *epistemic authority* for S with respect to domain D iff S is justified in believing that (i) A is an expert about D, and A is epistemically superior to S with respect to domain D (i.e., has superior reasoning competences, and has very likely considered all of S's relevant evidence).<sup>24</sup>

What does it mean that, from S's perspective, the authority must have very likely considered all of S's relevant evidence; and isn't that a very strong, may be unsatisfiable requirement? The idea is that authorities should typically take into consideration types rather than tokens of lay evidence that are relevant to the target proposition and that rely on trustworthy sources. The authority thus need not know every minute detail about the cognitive perspectives of laypeople or all the information they gathered from completely untrustworthy sources. Rather, she must have considered all pieces of noteworthy lay evidence. This is what scientific experts typically do and it does not overexert them.

In my definition, being an expert implies that one does sufficiently well on an absolute scale. Typically, expert scientists are epistemic authorities for laypeople in their domain of expertise. Often, these authorities help laypeople to compensate for their lack of knowledge or understanding about this domain.<sup>25</sup> At other times, laypeople confront the judgments of authorities when they already have beliefs and evidence of their own about the subject matter. So, when laypeople hear the advice of their medical doctors, they often already "know" something about the topic themselves. The interesting question is how laypeople should *rationally* respond in the latter case.

<sup>23</sup> One should keep in mind here that being an epistemic authority is different from being an expert. Whereas the status of an expert does not depend on being recognized as such, an authority has her normatively binding force only in virtue of being recognized as an expert.

<sup>24</sup> Let me flag that this definition is far from being uncontroversial. Some philosophers treat "authority" and "expert" as more or less synonymous terms (Goldman 1999; Lackey 2018), others share my view that authorities must be recognized as superiors (Zagzebski 2012: 103; Jäger 2016: 170). Most closely, my definition resembles Jäger's. There is, however, one crucial difference: Jäger does not require that S takes the authority to have considered all of S's own evidence.

<sup>25</sup> For service-oriented accounts of experts, see Quast 2018 and Croce 2019.

Two extreme responses are obviously inapt. The layperson should not blindly trust the authority, since this reaction would not be *rational*. Blind trust is not based on any reasons. On the other end of the spectrum, the layperson may treat the authority merely as a source of reasons, arguments, and data that must then be assessed by the layperson herself. This response would not give *any* epistemic weight to the authority's judgment, over and above the evidence she discloses. This is what typical individualists such as Plato, Descartes, or Locke would claim. However, the authority's judgment is *additional* higher-order evidence for the proposition in question. This can easily be seen by considering the following argument that is available to the layperson:

- (1) The epistemic authority believes that  $p$  (where  $p$  is some proposition in  $D$ ).
- (2) The authority's beliefs about  $p$  are most likely true.
- (3) Therefore,  $p$  is most likely true.

If the layperson can identify authority judgments, and knows that authorities have reliable judgment in their domain of expertise (which follows from the concept of epistemic authority), then they will be able to justify the conclusion. And this amounts to additional evidence for  $p$ . This evidence is available even in cases where the layperson is not able either to understand or to rationally use the first-order evidence disclosed by the authority. So, it is unreasonable to treat the epistemic authority merely as a source who distributes first-order evidence.

Apart from these two options, there are more moderate and more attractive reactions to the authority's judgment: the Total Evidence View (TEV) and the Preemption View (PV). According to (TEV), we should take into account all of our first-order evidence with respect to the target proposition  $p$ ; and this may include pieces of evidence that were disclosed by the authority. Additionally, we should also take into account the authority's judgment as an extra bit of evidence that has a lot of weight. Then we should base our judgment on the aggregation of all this evidence. As Lackey (2018: 239) puts it: "What I am proposing [...] is that the testimony of experts should always be regarded as a piece of evidence to be weighed with the other relevant evidence we have on the matter." According to (PV), the opposite is true. When we discover that an epistemic authority believes that  $p$ , we should not make any more use of our own reasoning about  $p$  as evidence for or against  $p$ . The use of our own reasoning concerning  $p$  is to be bracketed. As Zagzebski (2012: 107) puts it: "The fact that the authority has a belief  $p$  is a reason for me to believe  $p$  that replaces my other reasons relevant to believing  $p$  and is not simply added to them." Whereas the aggregation model of (TEV) is in line with (PDR), (PV) obviously conflicts with (PDR). This is because (PV) explicitly requires one to ignore the layperson's judgment when forming one's own belief.

In what follows, I will present two arguments in favor of (PV). The first of these is the *Track Record Argument*, which has Raz and Zagzebski as its main advocates. This argument runs as follows. Suppose you regard someone as an epistemic authority with respect to a target proposition  $p$ . Then you take her judgment on this proposition to be more reliable than your own. Now, if you give any weight to your own reasoning about  $p$ , there will be cases in which even this little weight outweighs the authority's judgment. Hence, the resulting track record will be inferior to a general strategy of deference to the authority. But then giving zero weight to your own reasoning about  $p$  will produce the most reliable result, and is therefore instrumentally most rational.<sup>26</sup>

Jennifer Lackey (2018: 238) has objected that even if this strategy of general deference to authority is the most reliable one, its implementation does not require preemption. She thinks that even if our own lay reasoning is not fully preempted, we may regard the authority-based reasons as so powerful that they always outweigh our other reasons. Hence, (PV) would not be required to explain why we should always follow the authority's lead.

However, this move is unconvincing. It seems clear to me that one cannot simply assign to the authority's judgment whatever weight is needed to outweigh one's own reasoning. This would be completely ad hoc and unmotivated. One should rather assign a weight that is proportional to the relative trustworthiness of the authority's judgment. But then there will be cases in which making use of our own reasons is not outweighed by the authority's judgment. Here is a toy case that may help to illustrate my point:

Agent	Credence with respect to $p$	Relative trustworthiness (weighting factor)	Credence $\times$ weighting factor
Epistemic authority	.55 (weak belief)	0.66	.366
Layperson	.1 (strong disbelief)	0.33	.033
Weighted average credence			≈.4 (weak disbelief)

How is this table to be read? The layperson recognizes that whereas she is strongly disbelieving  $p$ , the authority she is confronting weakly believes that  $p$ . Given that she takes the authority to be twice as trustworthy as herself in her judgment about the target proposition she calculates the weighted average credence and revises her credence accordingly. The result of this strategy obviously conflicts with the

<sup>26</sup> Cf., Zagzebski 2012: 114: "[T]here is another problem with treating the authority's belief that  $p$  as just another reason among others to believe. If I do so, I will worsen my track record in getting the truth." See also Raz 1988: 68–9.

strategy of general deference to epistemic authority because the former leads to weak disbelief whereas the latter leads to weak belief.

Lackey has a second objection to the Track Record Argument that is not affected by this criticism. She explicitly claims that always following the authority's advice is not the best strategy:

[I]t is not the case that in order to avoid worsening one's track record for getting the truth, one should always follow the advice of an authority. Here are some alternative policies that would have even better epistemic results: follow the advice of an authority, except when one is certain that the authority is wrong, follow the advice of an authority, except when what the authority says is highly doubtful. If humans adopted any of these policies, they would end up faring better [...] and, moreover, following them relies directly on not screening off the normative force of background evidence [...]. (Lackey 2018: 238)

Lackey is right that *if* the rules she suggests were better than the strategy of straight deference, then this could not be adopted by (PV). The reason is that implementing these rules requires using one's relevant background evidence to assess whether what the authority says is wrong or highly doubtful. However, (PV) strictly prohibits the use of this evidence.

So, the crucial question is: Is Lackey right in claiming that her rules are better than straight deference? Her thought is that in cases in which what the authority says looks crazy or outrageous to the layperson, the latter is more often right than wrong. Hence, rejecting the authority's claim *only in these cases* would improve the general track record. However, I do not think that Lackey is right about this. Often, propositions that look crazy or outrageous to laypeople are in fact true and rational. Lay intuitions are often strongly misleading. Moreover, as Dunning's (2005) empirical work suggests, laypeople are often blissfully unaware of their own incompetence. In particular, even when they are highly confident that their judgment is competent, they often err. This is because their meta-cognitive cues for accurate judgments are highly unreliable.<sup>27</sup> According to Dunning, people are highly confident if (i) they have explicit reasons for their judgment, (ii) their judgment is very fluent, or (iii) this judgment fits well with their pre-existing background views. Now, as Dunning convincingly argues, these cues are often unreliable when people are incompetent. Let us look more closely at each of the cues. First, incompetent people may be ignorant of the most relevant reasons. But if this is so, then having explicit reasons for one's judgment falls far short of establishing that one is right. Second, the fluency of one's judgment is not always determined by competence. Incompetent people judge fluently simply because of

<sup>27</sup> For the following, see Dunning 2005: ch. 3.

repetition effects or recent exposure. Third, fitting one's cognitive background only indicates truth if that background is true itself. However, in incompetent people this background can be grossly misleading. In the end, it is an empirical question whether laypeople would improve their track record by adopting one of the rules that Lackey suggests. But, so far, there is no reason to believe that she is right.

Lackey (2018: 236) raises a further objection to the Track Record Argument that is—in my view—her most important one: the *challenge from the epistemic obscurity of preemptive reasons*. As she correctly observes, the core idea of preemption, namely that the layperson's evidence is properly ignored, does not seem to cope with evidentialism, according to which evidence is always aggregated and never ignored. Preemption thus requires that there are also non-evidential reasons. But then it becomes unclear how these can be integrated into a unified framework of evidential and non-evidential reasons. It seems to me that Zagzebski's account does not have the resources to answer this challenge.

The second argument for (PV) is the *Higher-Order Undercutting Defeat Argument*.<sup>28</sup> I will argue that this argument can answer the challenge from the epistemic obscurity of preemptive reasons, because it can explain preemptive reasons as a special case of undercutting defeaters that any reasonable account of reasons should allow for.

I will introduce this second argument by means of an example. But before I do that, some additional terminology is needed. *Defeaters* turn *prima facie* justified beliefs into unjustified ones; or—more roughly—they remove justification. How they do that is a matter of controversy.<sup>29</sup> What is not controversial, however, is that there are at least two types of defeaters that work differently. A *rebutting defeater* is a new piece of evidence that outweighs one's prior justification (to some extent). Typically, this kind of defeater involves evidence against the truth of the target proposition. In contrast, *undercutting defeaters* involve a new piece of evidence that makes one's prior evidence rationally unusable as evidence for the target proposition.<sup>30</sup> One might think that undercutting defeat can be explained in the same way as rebutting defeat, namely by the aggregation of evidence. For example, whereas red-impressions are evidence of something red in view, red-impressions together with the information that the visible objects are illuminated by red light are no longer evidence of red objects. However, the mechanism of undercutting defeat also seems to work in cases of conclusive evidence. Suppose you have a perfect proof of some mathematical solution. In fact, your evidence entails the conclusion and you draw this conclusion. In this case, adding further

<sup>28</sup> For an earlier version of this argument, see Constantin & Grundmann 2020.

<sup>29</sup> For accounts that are either more internalist or externalist in spirit, see Melis 2014 and Constantin 2020.

<sup>30</sup> The *locus classicus* for this distinction is Pollock 1974.

pieces of evidence cannot turn your conclusive evidence into evidence that no longer supports the conclusion. This kind of reasoning is monotonic. But now suppose that you are informed by a highly trustworthy testifier that you were, without noticing it, exposed to some gas that strongly tends to create the illusion of proper proofs even in cases in which the actual reasoning is grossly invalid. If you share my intuition about this case, you will think that after having received this information you are no longer justified in believing in your mathematical solution.<sup>31</sup> This judgment cannot be explained by the aggregation of evidence. Rather, higher-order evidence seems to possess the normative epistemic power to neutralize the epistemic weight of your conclusive first-order evidence.<sup>32</sup> In this case, it would seem irrational if you still relied on your proof after having received the higher-order information. In my view, this demonstrates that undercutting defeat can be fully explained only if we accept that there are preemptive reasons. So, a proper explanation of undercutting defeaters requires the mechanism of preemption. However, if preemption is required to explain such a mundane phenomenon as the full spectrum of undercutting defeat, then it is hard to see how one can insist that preemptive reasons are obscure.

Let us assume that at least some cases of undercutting defeat are such that the higher-order evidence that a lower-order judgment is irrational makes this judgment irrational. It is still an open question whether and how the identification of an authority's judgment provides us with an undercutting defeater for any judgment that would be based on our own reasoning. To get clearer about this issue, it will be helpful to consider a specific case.

Suppose you meet Bryan, a mental math crack. Someone asks, "What is 175,998 plus 22,453?" It takes Bryan only a few seconds to answer "198,451." Having received this information, should you use your own mental calculation in determining the correct answer? The answer is clearly no. Either the result of your own calculation would correspond to Bryan's or it would deviate from it. A corresponding result would make no difference. In the case of a deviant result, however, it would be highly likely that you made a mistake. After all, you disagree with a superior. Hence, your own reasoning is undercut.

Here is a more abstract characterization of what happens in this case. When the layperson identifies an epistemic authority on some domain, she not only has reasons to believe that the expert's judgment is more competent than her own judgment, but typically she also has reasons to believe that the authority has considered all of the layperson's reasons that are *relevant* to the assessment of the target judgment. This is what we expect of authorities by default.<sup>33</sup> If the layperson

<sup>31</sup> Cf. Christensen 2010.

<sup>32</sup> For opposing views, see Lasonen-Aarnio 2014, 2020, and Weatherson 2019.

<sup>33</sup> In the case of Bryan, the evidence is shared since both, the layperson and Bryan, calculate the same math problem; and even if Bryan does not know all mental math strategies, he knows that they are all deductive and their proper use thus cannot lead to different rational results.

then relied, in making her judgment, on her own reasons *in addition* to the expert's judgment, this would, in all likelihood, lead to a deviation from the expert's judgment, which is from the layperson's own perspective the more competent and, most probably, the rational judgment. However, if this is true, then it is, from the layperson's perspective, very likely that she would deviate from the rational judgment if she relied, in making her judgment, on her own reasons as well. In this way, the layperson acquires reasons to believe that using her own domain-specific reasons would lead to an irrational belief. This in turn generates an undercutting defeater in the technical sense.

We can now turn these considerations into the following argument, which is available to the layperson when she identifies an authority's judgment:

- (1) I am justified in believing that the authority's judgment about p is most likely rational (given that p is a proposition within the authority's domain of expertise).
- (2) I am justified in believing that the authority has most likely considered all of my relevant domain-specific evidence regarding p.
- (3) If (1) and (2) are true, then I am *prima facie* justified in believing that deviating from the authority's judgment on p would most likely render my judgment irrational.
- (4) I am *prima facie* justified in believing that deviating from the authority's judgment on p would most likely render my judgment irrational. [from 1, 2, 3]
- (5) If I made use of my own domain-specific reasons for judging whether p, this would lead to a judgment that either conforms with the authority or deviates from her judgment. [truism]
- (6) When use of my own domain-specific reasons makes no difference to the resulting judgment, it is irrelevant.
- (7) When use of my own domain-specific reasons makes a difference, I am *prima facie* justified in believing that it is most likely irrational. [from 4]
- (8) If I am justified in believing that my use of reasons is irrational, then making use of them is irrational. [undercutting defeat]
- (9) When using my own domain-specific reasons makes a difference, making use of them is irrational. [from 7, 8]
- (10) I cannot rationally use my own domain-specific reasons regarding p. [from 5, 6, 9]

As I have already argued, (1) and (2) are reasonable because the layperson expects the authority to be a rational believer in her domain of expertise, and one who has already considered all the relevant evidence that is accessible to laypeople like the layperson in question. (3) is reasonable because laypeople take *uniqueness*, i.e., the view according to which there is only one rational response to a given body of

evidence, to be the default position. This does not entail that uniqueness is true without any restrictions.<sup>34</sup> (6) seems to be obviously correct: if the use of some evidence makes no difference with respect to the target judgment, then it is irrelevant for this judgment. Finally, (8) articulates a specific interpretation of undercutting defeat, namely one that involves level-connection, according to which the higher-order evidence of first-order irrationality is sufficient to render the corresponding first-order judgment irrational. This is a controversial assumption that I nevertheless rely on here without further justification.<sup>35</sup> I think that cases of undermined judgments that rely on conclusive evidence indicate that this assumption is correct. However, my assumption here would clearly need further defense on another occasion.

In contrast to the Track Record Argument, the Argument from Higher-Order Undercutting Defeat emphasizes that the layperson's identification of an authority judgment makes her own (domain-specific) reasons rationally unusable for her. This is an issue of rationality rather than reliability. Moreover, the second argument uses the initially mysterious-looking mechanism of preemption only insofar as it can be explained by undercutting defeat. Hence, it provides an answer to Lackey's third objection. Finally, the second argument restricts preemption in specific ways. The layperson is prohibited from using her domain-specific reasons when she judges whether *p*. However, she is still permitted to use any of her domain-independent reasons, because there is no reason for the layperson to expect the authority to have superior judgment in these areas as well.

## 2. Why (PDR) Can Have Epistemically Dangerous Consequences

In the previous section, I argued that (PDR), as it stands, is false. According to this principle, we cannot rationally ignore anyone's reasoning, not even if it is the reasoning of laypeople who happen to be ourselves. This directly conflicts with (PV), which requires us to refrain from using our own domain-specific reasons when confronting epistemic authorities. In this section, I will argue that (PDR), in addition to being false, also has very bad epistemic consequences for the formation of public opinion. In particular, it facilitates the spread of conspiracy theories among the public. Or so I will argue.

Defining conspiracy theories is not an easy task. Clearly, not all theories about conspiracies are conspiracy theories. Otherwise, people like me who believe in Nixon's Watergate conspiracy or the NSA conspiracy would correctly be classified as conspiracy theorists. On the other hand, the term "conspiracy theory" is more

<sup>34</sup> For a defense of uniqueness, see White 2005; for a critique, cf., Kelly 2013.

<sup>35</sup> This assumption is disputed by, e.g., Lasonen-Aarnio 2014, 2020 and Weatherson 2019.

substantial than purely pejorative, weaponized concepts that can be applied to basically everything. On my view, this term characterizes theories about conspiracies as having an epistemic basis that exhibits a specific kind of epistemic deficiency. Hence, it is an epistemic vice to believe in such theories on this basis.

Typically, conspiracy theories are radical alternatives to official, scientific, or expert views (cf., Levy 2007). These alternatives are accepted by their believers because the official story seems unacceptable to them. Now, one explanation of why the official stories seem unacceptable is that they do not pass the laypeople's plausibility check. These stories look so outrageous and crazy to laypeople that they simply cannot be accepted as true from their commonsense perspectives. However, true scientific and expert views often are so radically different from laypeople's background beliefs that they must look outrageous to them.<sup>36</sup> This is simply a by-product of these theories' degree of novelty, complexity, and sophistication. So, as long as laypeople have switched *on* their plausibility checks, we can expect them to tend to reject official and scientific theories, and to look for some alternative theory instead.

Here are two familiar examples. The first concerns the collapse of the World Trade Center on 9/11 (cf., Dunbar & Reagan 2006). According to the data on this collapse, it took the Twin Towers between fourteen and sixteen seconds to collapse. In comparison, it would take something nine seconds to free-fall from the top of this building. This difference is so small that laypeople do not understand how the buildings could collapse so quickly without a controlled demolition, i.e., without simultaneous controlled explosions on many of its floors. Keep in mind that the World Trade Center was a very solid construction. The outrageousness of this fact is clearly expressed by Rosie O'Donnell, one of the "Truthers" who characteristically express incredulity here:

Do you know how fast it took those towers to fall? Nine seconds [...]. You know how fast it would have taken something to free-fall from the top of that building? Nine seconds. It's physically impossible.<sup>37</sup>

However, according to the relevant community of experts, simulations have proven that a period like the actual period of the collapse was to be expected under the circumstances. Here we have a clear case of an expert judgment that appears outrageous to some laypeople, even though it is correct.

My second example is the Monty Hall Puzzle. This case has nothing to do with conspiracy theories, but the standard reaction of laypeople in this case is

<sup>36</sup> For a general interpretation of the clash between common sense and science along these lines, see McCauley 2011.

<sup>37</sup> Cited from Dunbar & Reagan 2006: 42-3. Obviously, O'Donnell did not report the exact numbers, but speeded up the collapse a bit to make her point more vivid.

structurally analogous. Suppose you are in a game show. Behind one of three doors is a brand-new car that the player will win if she picks the correct door. The player picks, say, door number 1. Then the host opens one of the other doors, say door number 3, showing that this is not the correct door. The host now offers the player the option to switch from door 1 to 2. Should she accept this offer to raise her chances? A number of scientific studies have shown that an average of 85 percent of people do not switch here (Bruns & Wieth 2004). For them, the chances seem equal and hence there is no reason to switch. However, as expert mathematicians have proven and also simulated, the player's chances are raised from 1/3 to 2/3 if she switches.<sup>38</sup> This is an outrageous but true expert judgment.

Both of these cases have in common that laypeople tend to reject a correct judgment because it looks outrageous on the basis of their own domain-specific reasoning and subsequently take an alternative conspiracy theory to be correct. Only to the extent to which people rely on this reasoning will conspiracy theories or conspiracy-analogous theories be accepted by laypeople and the general public. The public's susceptibility to conspiracy theories thus seems to depend on whether (PDR) is accepted. People who choose (PV) instead seem to be less susceptible to conspiracy theories. (PV) thus protects laypeople from conspiracy theories.

Before moving on, we must address a serious objection to this line of thought. Perhaps we have been misled here by a one-sided selection of cases. Aren't there also famous cases of well-respected authorities who are in fact gurus who lead laypeople away from the truth? Then, giving up on (PDR) would sometimes have the opposite effect, i.e., removing the resources of reasonable lay criticism when authorities are massively wrongheaded. I am thinking here of cases such as the story of Peter Duesberg.<sup>39</sup> He is a famous University of California molecular biologist who does not accept that AIDS is caused by a virus. His political influence on Mbeki's administration in South Africa caused a massive failure to provide the public with antiretroviral drugs, which in turn caused hundreds of thousands of preventable deaths. Now, if we accepted (PV) instead of (PDR), wouldn't that lead to an uncritical acceptance of Duesberg's authority view? You might even speculate that such uncritical acceptance of authority judgments was the main reason for his influence on administrative decisions in South Africa.

This objection, however, underestimates the resources of (PV). It is true that (PV) prohibits the use of *domain-specific* reasons when, e.g., the truth of Duesberg's claim is to be assessed. But laypeople can still rely on *domain-independent* social reasons when they form their judgments. In Duesberg's case, it is a publicly well-known fact that he has published his view in renowned peer-reviewed journals such as *Lancet*, *Science*, *Nature*, and *Journal of AIDS*. But none of his colleagues accepted his view, even after carefully considering his arguments.

<sup>38</sup> For a comprehensive discussion, see Rosenhouse 2009.

<sup>39</sup> For the following, see Oreskes 2019: 146.

So, in this case, Duesberg's authority judgment is defeated as a testimonial reason and as an undercutting defeater by the overwhelming rejection by his colleagues. Hence, (PV) clearly provides laypeople with the resources to reject minority views among epistemic authorities.

### 3. Remaining Deep Worries about Restricted Critical Thinking

So far, I have argued that *unrestricted* democratic reasoning and *unrestricted* critical thinking cannot be accepted. These are suboptimal strategies for achieving true beliefs on the basis of authority reasons; they violate the norms of rationality when one confronts authority judgments; and they can lead to the spreading of unwarranted conspiracy theories. These are good reasons to revise the principles (PDR) and (PUCT) in the light of (PV).

The Principle of Democratic Reason needs to be amended by the following caveat (underlined):

(PDR\*) Whenever you rationally consider the truth of some proposition  $p$ , you must not exclude, marginalize or silence any rational thinker's reasoning (including your own) about  $p$ , unless you receive authority reasons regarding  $p$ . In that case, you are not permitted to use your own domain-specific reasons.

Correspondingly, the Principle of Unrestricted Critical Thinking should also be revised as follows:

(PUCT\*) Whenever you rationally consider the truth of some proposition  $p$ , you must never stop using your own reasons for  $p$ , unless you receive authority reasons regarding  $p$ . In that case, you are not permitted to use your own domain-specific reasons.

In this section, I will address three remaining worries that Lackey (2018) has directed against these restrictions. Her *first worry* is that epistemic preemption leaves us with *blind* and *indefeasible trust* in authorities. More specifically, Lackey (2018: 236) thinks that if one's own reasons are preempted by the authority judgment, then there aren't any reasons left to identify authorities or to reject them as authorities. Moreover, laypeople would lack the required reasons for ranking conflicting authorities.

Lackey's worry here would be substantial, if preemption prohibited the use of any reasons by the layperson. But this is not what preemption in fact requires. It merely prohibits the use of *domain-specific* reasons. In turn, this means that all domain-independent reasons are still usable. These reasons may concern

credentials, social facts, logical facts, etc. And these facts allow for a reasonable identification of authorities as well as their criticism. If, e.g., a layperson recognizes that someone who has so far been accepted as an epistemic authority in biology asserts something that conflicts with what all of her peers claim, or that is based on a fallacy, the layperson is justified in disputing the authority's current assertion. Laypeople can also rely on reputational facts and acceptance by colleagues when they rank epistemic authorities. So, as long as one keeps in mind that preemption does not screen off all reasons across the board, Lackey's first worry can be dismissed.

Lackey's *second worry* (Lackey 2018: 234) concerns the rational inescapability of informationally encapsulated sects or corrupted scientific institutions, when laypeople are not permitted to use their own domain-specific reasons. Suppose you are a layperson living in a totalitarian country in which the scientific acceptance of theories is generally determined by political ideologies (as in Nazi Germany, where race theory dominated biology and German physics disputed relativity theory and quantum mechanics; or as in the Stalinist USSR, where Lysenkoism disputed the relevance of genetics and natural selection for agriculture). Then, (PV) would advise you to follow the grossly misleading majority view of experts. Now, Lackey argues, you might escape from this ideology by using your own reasoning about the subject matter. But this road is closed off if one accepts (PV). Something similar might happen when you are caught in a deceptive echo chamber in which you do not have access to any respectable expert with a diverging view.<sup>40</sup>

What does this worry show exactly? It seems true that (PV) is not a safe escape route from deeply deceptive environments. (PV) cannot guarantee that there is a rational way to avoid deception. It also seems true that it is *possible* for laypeople to break free of such deception when they use their own reasoning, rather than following the authorities' lead. But how realistic is it for this possibility to become actual? I think that this is not very realistic. Of course, experts may be able to escape from such kinds of ideological deception. In Nazi Germany, physicists were able to recognize that the official rejection of relativity theory was a fraud. And in the Stalinist USSR, biologists were able to recognize that Lysenkoism was a fraud. However, (PV) does not apply to experts, even if not all experts are cognitively equal. Inferior experts should not treat superior experts as epistemic authorities, because they cannot generally expect their superiors to have considered all their relevant evidence. Since scientists actively seek new evidence, evidential inequality is the default assumption. But when we consider laypeople, it is very unlikely that they will be able to see the fraud in official science when they use their own reasoning. However, even if an alternative to (PV) were the instrumentally better

<sup>40</sup> For more on this kind of worry, see Nguyen 2020.

strategy for extreme situations, such as encapsulated deceptive groups or corrupted science, it does not follow that this alternative would be generally more successful. So, instrumentally speaking, (PV) is not perfect but, in general, better than any other strategy. Lackey's worry does nothing to show that this is wrong.

Lackey's *third worry* is perhaps her most serious one. She thinks that adopting (PV) may lead to an epistemic catastrophe (2018: 236–7). For, imagine that the layperson always defers to the authority's judgment. Then she does not need to seek new evidence, nor does she have to practice her reasoning skills, nor does she need to derive new knowledge from old knowledge. On (PV), all these required things are outsourced to the authority. What is scary about this situation becomes visible when we consider what happens when we reject someone as an authority for us who had this status before. According to Lackey, we will then lack the resources to form any judgment; and this would indeed be a catastrophe.

I think there is indeed a serious problem with lacking the resources for reasoning oneself. But it is not the one indicated by Lackey. Epistemic dependence on others just seems to be the human condition. If we reject one authority because she turns out to be corrupt or all the other experts disagree with her, then we do not become dependent on our own reasoning, but may also defer to the other authorities who are typically available.

The real problem with lacking the resources for reasoning oneself is twofold. If laypeople always follow the expert's lead, then (i) they will never acquire understanding,<sup>41</sup> and (ii) experts will gradually die out. Experts must be able to reason on their own. Thus, if laypeople did not develop reasoning skills themselves, this would lead to the community's cognitive suicide in the long run. There would not be any new experts to which laypeople could then defer. Moreover, laypeople would lack an appropriate understanding of what they accept, because they would not be able to grasp the reasons for the truths they accept on authority.

If (PV) is correct, then one cannot claim that acquiring new evidence, cultivating reasoning skills and making inferential connections is needed as a proper basis for believing. But there may be other values associated with possessing these things. I have already mentioned two additional epistemic values: acquiring understanding and developing expertise. Since even laypeople want to acquire understanding and gradually develop expertise, it is reasonable to seek new evidence and to practice reasoning, even if these things are not strictly required for forming beliefs.<sup>42</sup>

So far, this is only an abstract argument. How do deference to authorities and developing reasoning skills fit together in practice? Consider the following case.

<sup>41</sup> Here I assume without any further argument that understanding cannot be transmitted through testimony. For dissenting views, see Boyd 2017 and Croce 2018.

<sup>42</sup> A further value of cultivating these skills may be epistemic autonomy. For an approach along these lines, see Faulkner 2016.

A lay mathematician defers, in his mathematical judgments, to an expert mathematician. So, he does not base his judgments on his own proofs. However, the lay mathematician not only wants to know mathematical theorems, he also wants to understand why they are true. In order to acquire this additional understanding, the lay mathematician does two things. First, he follows the expert's lead in his genuine judgment. Second, he forms off-line simulated judgments for himself, following his own assessment of his domain-specific reasons and thereby trying out proofs. If the lay mathematician can use his own reasoning to reach those mathematical solutions that he already believes on authority, then he grasps *why* these propositions are true. He gains understanding without basing his beliefs on his own reasoning. By checking whether his off-line judgments typically correspond with his beliefs on authority, and calibrating his off-line judgment, the lay mathematician may also recognize whether he makes progress in acquiring mathematical expertise himself. We can now see that developing one's own reasoning skills and deferring to authority do not exclude each other, but are in fact complementary if we take into account further epistemic goals beyond just forming epistemically proper beliefs.

#### 4. Conclusion

It turns out that the Principle of Democratic Reason and the Principle of Unrestricted Critical Thinking must both be revised. As they stand, these principles do not reflect what rationality requires, and tend to facilitate the spreading of conspiracy theories among the public. Instead, laypeople must defer to what they identify as the views of epistemic authorities, without giving any weight to the plausibility or implausibility of these views from a lay perspective. In contrast to what critics have claimed, this local restriction of critical thinking does not have any crazy consequences. In particular, it does not involve giving up critical thinking in general. When we confront authorities, critical thinking should have its exclusive role in domain-independent reasoning (e.g., assessing their status as authorities in light of their credentials, reputations, awards, peer reviews, or educations, but also signs of bribery, bias, or being drunk or tired, etc.). If I am right, revising our received intellectual standards is not only a matter for academic discourse, but should also have social and political consequences. That experts play a specific epistemic role in society should be taught in schools and at universities; and institutions should make epistemic authorities and their ranking more visible to the public. It remains to be seen how this can be implemented in the Age of the Internet.<sup>43</sup>

<sup>43</sup> Early versions of this chapter were presented as Zeno Lecture at the University of Utrecht in December 2017, at a workshop on Political Epistemology at the University of Copenhagen in

## References

Boyd, Kenneth, 2017: "Testifying Understanding", *Episteme* 14: 103–27.

Brotherton, Rob, 2015: *Suspicious Minds: Why We Believe Conspiracy Theories*, New York & London: Bloomsbury.

Cassam, Quassim, 2019: *Vices of the Mind: From the Intellectual to the Political*, Oxford: Oxford University Press.

Christensen, David, 2010: "Higher-Order Evidence", *Philosophy and Phenomenological Research* 81: 185–215.

Constantin, Jan, 2020: "Replacement and Reasoning: A Reliabilist Account of Epistemic Defeat", *Synthese* 197: 3437–57.

Constantin, Jan & Grundmann, Thomas, 2020: "Epistemic Authority: Preemption Through Source Sensitive Defeat", *Synthese* 197: 4109–30.

Croce, Michel, 2018: "Expert-Oriented Abilities Vs. Novice-Oriented Abilities: An Alternative Account of Epistemic Authority", *Episteme* 15: 476–98.

Croce, Michel, 2019: "On What It Takes to be an Expert", *Philosophical Quarterly* 69: 1–21.

Curtois, C. et al., 2018: "Challenging Google Search Filter Bubbles in Social and Political Information: Disconfirming Evidence from a Digital Methods Case Study", *Telematics and Information* 35: 2006–15.

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

Dunning, David, 2005: *Self-Insight: Roadblocks and Detours on the Path to Knowing Thyself*, New York and Hove: Taylor and Francis.

Faulkner, Paul, 2016: "Agency and Disagreement", in Patrick Reider (ed.), *Social Epistemology and Epistemic Agency: Decentralizing the Epistemic Agent*, London: Rowman & Littlefield, 75–90.

Frankfurt, Harry G., 2005: *On Bullshit*, Princeton & Oxford: Princeton University Press.

Gelfert, Axel, 2018: "Fake News: A Definition", *Informal Logic* 38: 84–117.

Goldman, Alvin, 1999: *Knowledge in a Social World*, Oxford: Oxford University Press.

December 2017, at the University of Luxembourg in May 2018, at the Conference on Fake Knowledge at the University of Cologne in June 2018, at the Cologne Summer School in Philosophy with Jennifer Lackey in July 2018, at the Conference on Social Epistemology, Santa Maria, Brazil, in November 2018, at the University of Klagenfurt in November 2018, and at the Post-Truth Conference, Bologna, in December 2019. Very helpful comments from and extensive discussions with the following colleagues enabled me to work out the final version of this chapter: Dominik Balg, Sven Bernecker, Elke Brendel, Jan Constantin, Michel Croce, Filippo Ferrari, Axel Gelfert, Sanford Goldberg, Dietmar Heidemann, Frank Hofmann, David Horst, Joachim Horvath, Christoph Jäger, Clemens Kappel, Jens Kipper, Steffen Koch, Jennifer Lackey, Benjamin McCraw, Matthew McGrath, Sebastiano Moruzzi, Tommaso Piazza, Chris Ranalli, Ursula Renz, Luis Rosa, and Sarah Wright. I am extremely grateful to all of them. Finally, I would also like to thank two anonymous reviewers of this chapter.

Habermas, Jürgen, 1991: *Erläuterungen zur Diskursethik*, Frankfurt am Main: Suhrkamp.

Haim, M. et al., 2017: "Abyss or Shelter? On the Relevance of Web Search Engines' Search Results When People Google for Suicide", *Health Communication* 32: 253–8.

Hannack, A. et al., 2013: "Measuring Personalization of Web Search", Proceedings of the 22nd International Conference on World Wide Web. <https://doi.org/10.1145/2488388.2488435>.

Jäger, Christoph, 2016: "Epistemic Authority, Preemptive Reasons and Understanding", *Episteme* 13: 167–85.

Jaster, Romy & Lanius, David, 2018: "What is Fake News?", *Versus* 2: 207–27.

Kahneman, Daniel, 2011: *Thinking, Fast and Slow*, London & New York: Penguin Books.

Kelly, Thomas, 2013: "Evidence Can be Permissive", in M. Steup & J. Turri (eds.), *Contemporary Debates in Epistemology*, Oxford: Wiley Blackwell, 298–312.

Lackey, Jennifer, 2018: "Experts and Peer Disagreement", in M. Benton, J. Hawthorne, & D. Rabinowitz (eds.), *Knowledge, Belief, and God: New Insights in Religious Epistemology*, Oxford: Oxford University Press, 228–45.

Landemore, Helene, 2013: *Democratic Reason: Politics, Collective Intelligence, and the Rule of the Many*, Princeton & Oxford: Princeton University Press.

Lasonen-Aarnio, Maria, 2014: "Higher-Order Evidence and the Limits of Defeat", *Philosophy and Phenomenological Research* 88: 314–45.

Lasonen-Aarnio, Maria, 2020: "Enkrasia or Evidentialism: Learning to Love Mismatch", *Philosophical Studies* 177: 597–632.

Levy, Neil, 2007: "Radically Socialized Knowledge and Conspiracy Theories", *Episteme* 4: 181–92.

Lynch, Michael, 2016: *The Internet of Us: Knowing More and Understanding Less in the Age of Big Data*, New York: Liveright Publishing.

Lynch, Michael, 2019: *Know-It-All Society: Truth and Arrogance in Political Culture*, New York & London: Liveright.

McCauley, Robert, 2011: *Why Religion is Natural and Science is Not*, Oxford: Oxford University Press.

McIntyre, Lee, 2018: *Post-Truth*, Cambridge, MA & London: MIT Press.

Melis, Giacomo, 2014: "Understanding Undermining Defeat", *Philosophical Studies* 170: 433–42.

Nguyen, Thi, 2020: "Echo Chambers and Epistemic Bubbles", *Episteme* 17: 141–61.

Oreskes, Naomi, 2019: *Why Trust Science?*, Princeton & Oxford: Princeton University Press.

Oreskes, Naomi & Conway, Erik, 2010: *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues From Tobacco Smoke to Global Warming*, New York: Bloomsbury.

Pariser, Eli, 2011: *The Filter Bubble: What the Internet is Hiding From You*, London & New York: Penguin Books.

Pollock, John, 1974: *Knowledge and Justification*, Princeton & Oxford: Princeton University Press.

Quast, Christian, 2018: "Expertise: A Practical Explication", *Topoi* 37: 11–27.

Raz, Joseph, 1988: *The Morality of Freedom*, Oxford: Oxford University Press.

Rosenhouse, Jason, 2009: *The Monty Hall Problem*, Oxford: Oxford University Press.

Sunstein, Cass, 2006: *Infotopia: How Many Minds Produce Knowledge*, Oxford: Oxford University Press.

Uscinski, Joseph & Parent, Joseph, 2014: *American Conspiracy Theories*, Oxford: Oxford University Press.

Vosoughi, Soroush et al., 2018: "The Spread of True and False News", *Science* 359: 1146–51.

Weatherson, Brian, 2019: *Normative Externalism*, Oxford: Oxford University Press.

White, Roger, 2005: "Epistemic Permissiveness", *Philosophical Perspectives* 19: 445–59.

Zagzebski, Linda, 2012: *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief*, Oxford: Oxford University Press.

# Is Fake News Old News?

*Catarina Dutilh Novaes and Jeroen de Ridder*

A well-functioning liberal democracy needs a public sphere of rational discourse in which informed citizens participate in reasonable debate about political decisions. To exercise their political responsibilities, citizens must have access to reliable information, at least part of which comes from journalism and the press (sometimes referred to as the ‘fourth power’ or ‘fourth estate’ in a democracy). In light of recent political events in the USA, Europe, and elsewhere, it is understandable that many have expressed concerns about how the (alleged) proliferation of fake news, misinformation, and propaganda threatens to undermine democracy. Indeed, alarmistic complaints that we have now entered a post-truth era have become a staple of recent nonfiction writing (e.g. D’Ancona 2017; Davis 2017; Kakutani 2019; McIntyre 2018).

These concerns suggest that things were substantially different (in fact, *better*) before, and that recent years have witnessed far-reaching changes in political misinformation and disinformation—a flood tide of fake news. In this chapter, we address the question to what extent contemporary fake news really is a novel phenomenon. We begin by delineating the sort of fake news that will be central in our investigation. Section 2 presents three models, or strategies, for the production and proliferation of fake news. All three strategies have been employed during various periods in history and are thus compatible with pre-internet technology. This gives us reason to think that contemporary fake news isn’t radically different or novel, at least in some respects. In Section 3, we look at a number of further features of contemporary fake news which may account for its novelty: (a) content, (b) wide proliferation, circulation, and increased influence, and (c) modes of production, distribution, and consumption. We will show that, as far as the historical and empirical evidence goes, we have little reason to think features (a) and (b) set contemporary fake news apart from its precursors. Aspects of (c), however, do: the Internet and social media have changed the ways in which fake news can be produced, distributed, and consumed; and various actors have exploited these new possibilities to great effect.

Our answer to the question whether fake news is old news is thus carefully qualified. In many ways, contemporary fake news is nothing new, and suggestions that we have entered a post-truth era where alternative facts run rampant are wrong—at least in the sense that there has never been a ‘truth era’; disinformation

and propaganda were widespread phenomena in the (pre-internet) past, too. Nonetheless, contemporary disinformation and propaganda—in the form of fake news—have cleverly adapted to the commercial and technological possibilities of the online media environment, and so we are right to be concerned about the political effects of these new developments.

## 1. Definitions of Fake News and the Scope of this Investigation

The recent history of the term ‘fake news’ can be traced back to journalist Craig Silverman, whose first public use of the term in a 2014 tweet referred to the ‘fake news website’ National Report, which had broadcast a fabricated story of an Ebola outbreak in Texas (Silverman 2017). To his dismay, a few years later the term was appropriated by Trump and his supporters precisely to discredit the legitimacy and reliability of mainstream news organizations whenever they published a news item not to their liking—that is, a complete inversion of Silverman’s original usage. In other words, in a short period the term ‘fake news’ has acquired multiple, even opposite meanings.

In the recent philosophical literature, much attention has been paid to proposed definitions of the term ‘fake news’ (Brown 2019; Fallis & Mathiesen 2019; Gelfert 2018; Levy 2017; Mukerji 2018; Rini 2017; Søe 2019). This is not surprising, as conceptual analysis (in particular in formulations of necessary and sufficient conditions for something to count as X) is widely thought to be one of the main contributions that philosophers can make. But for this exercise to be fruitful, the presupposition must be that there is in fact “a certain fairly coherent social phenomenon” (Pepp et al. 2019, 68) that is picked out by the term ‘fake news’. This presupposition has been contested in particular by Habgood-Coote (2019) and Coady (2019, and Chapter 3 in this volume), who refer to a number of other authors (philosophers and journalists) who question the possibility and/or usefulness of defining the term ‘fake news’.

We side with Habgood-Coote and Coady in questioning whether there really is a sufficiently stable, coherent phenomenon that is picked out by the term ‘fake news’; instead, there seem to be rather heterogeneous uses that are picking out different phenomena. Thus, instead of departing from a general definition of fake news, we will focus on the following phenomenon: coordinated, deliberate efforts to manipulate public opinion by spreading false, misleading, or confusing messages posing as pieces of journalism, in particular but not exclusively in political contexts, also known as propaganda (Stanley 2015).<sup>1</sup> Examples are news-like

<sup>1</sup> Two other phenomena that are frequently associated with the term fake news are: (1) Clickbait, i.e. news-like stories (often, but not necessarily, false) produced primarily for the financial profit of the producer and not to inform consumers. Examples include yellow journalism (Samuel 2016), UK

stories produced and disseminated by websites that masquerade as serious news outlets, often with a domain name mimicking real newspapers,<sup>2</sup> by extreme right-wing or left-wing websites, or by so-called troll factories (Pomerantsev 2019), but also the tobacco industry's organized campaign to discredit scientific findings linking tobacco consumption to cancer (Michaels 2008; O'Connor & Weatherall 2019; Oreskes & Conway 2010).

In what follows, we thus address the question how novel this phenomenon is. An important goal for us is to connect the recent literature in analytic epistemology on fake news to the older, more developed literature on propaganda. We believe that the phenomenon of fake news is an instantiation of a much broader class of phenomena, and that attending to earlier analyses of these phenomena, propaganda in particular, is crucial if we are to attain a better understanding of these recent manifestations that are typically referred to as 'fake news'.<sup>3</sup>

## 2. Manipulation of Public Opinion: Three Models

In order to address the question of the putative novelty of current strategies of manipulating public opinion, in particular by means of purported news items, it will prove useful to distinguish three models of manipulation of public opinion primarily for political purposes. (We set aside advertisement and marketing, which are also well-established forms of manipulation but for the purposes of selling products or services.) These three models are abstractions that may never be instantiated in their pure form in the real world. Moreover, in practice they function as extremities in a spectrum rather than as clear-cut categories; concrete situations will typically instantiate each of them to different degrees. Nevertheless, they offer a convenient vantage point to address the question of how novel the phenomenon of fake news really is, in particular because the view that Models A and B are things of the past, whereas Model C is a recent phenomenon—a real

tabloids, and the fake stories supporting Trump's 2016 candidacy that were concocted by Macedonian teenagers not because they favoured Trump, but because they generated most clicks and profit for them (Silverman & Alexander 2016; Subramanian 2017); and (2) conspiracy theories, such as 'Pizzagate' (Robb 2017) and the still popular theory that the 9/11 terrorist attacks on the World Trade Center were an inside job orchestrated by the US government (Bell 2018). We will not address these phenomena here.

<sup>2</sup> The fact-checking website PolitiFact provides an overview of such websites, many of which are by now defunct: <https://www.politifact.com/article/2017/apr/20/politifacts-guide-fake-news-websites-and-what-they/>.

<sup>3</sup> An anonymous referee objects that our focus is too broad, and that we thus fail to focus on 'real' fake news. We disagree; we believe that a broader focus is precisely what is required to locate and understand fake news within a larger family of related phenomena. A recent instance of an approach similar to ours is Habgood-Coote (2020), who discusses fake news in connection with fascist and authoritarian discourse.

novelty—appears to be widely held, if casual conversation is any indication. We will contest both aspects of this position.

For each of the three models, we will also discuss the extent to which knowers in each of these environments still preserve some degree of *epistemic autonomy*. There is a traditional strand of thinking about autonomy that conceives of it a complete epistemic *self-reliance*. But this is unrealistic for all of us. Knowledge production and circulation is very much a social affair—we depend on others in various and sundry ways for much of what we know. To think that thoroughgoing self-reliance is desirable even as an ideal is untenable (Zagzebski 2012). What is required instead is a *relational* approach to autonomy (Grasswick 2019, 200–2), which is consistent with the fact that much of what we know, we learn from others. From this perspective, epistemic autonomy is not incompatible with epistemic dependence. In fact, in some cases epistemic autonomy will *require* epistemic dependence. The purportedly autonomous agent who trusts no one else's word is an irrational dogmatist (Fricker 2006). Instead, the opposite of epistemic autonomy is *epistemic heteronomy*: the condition of being epistemically under the domination of forces outside of the individual (Grasswick 2019, 202ff.).

Indeed, we submit that it is better to think of epistemic autonomy as a form of *self-governance*: when exercising epistemic autonomy, one engages one's own reasons, thus obtaining suitable justification for one's own beliefs (or disbeliefs or withholdings) (Zagzebski 2012). Self-governance (or at least the sort of self-governance worth pursuing) requires more than merely having options. It also involves having reasonably accurate information or at least justified beliefs about one's options. Thus, there is an epistemic condition on epistemic autonomy: making up one's own mind about what to believe requires access to the evidence for and against one's belief-options.<sup>4</sup> Prima facie, the epistemic autonomy of knowers living in environments saturated with fake news will be compromised, because much of what counts as fake news is intended to manipulate public opinion. However, we will see that the implications for epistemic autonomy are different for each of the three models.

## 2.1 Model A: Pleasing and Seducing the Audience

This model of manipulation of public opinion is as old as democracy itself, at least if Plato is to be believed on the state of Athenian democracy around his lifetime. It simply consists in politicians making assertions with the purpose of pleasing the audience—and thus obtaining their votes—rather than making truthful statements. According to Plato (e.g. in the *Gorgias*), this is exactly what rhetoricians

<sup>4</sup> Thanks to an anonymous referee for alerting us to the importance of making this explicit.

taught their students to do, who then went on to apply the knack they had learned from the rhetoricians in order to manipulate voters in the Assembly. For Plato, the discourse of the rhetorician is no more than a form of flattery, comparable to pastry baking and cosmetics, which may be pleasant at first but is ultimately detrimental. The rhetorician is like a pastry chef, who offers delicious but unhealthy treats, whereas the philosopher is like a true doctor, who restores the health of a sick person by aiming at truth, even if the treatment itself is rather disagreeable. Crucially, when offered the choice between the doctor and the pastry chef, people will often choose the pastry chef (Moss 2007).

In the *Republic* (Book VI), Plato offers an epistemic argument against democracy, which depends on the assumption that voters are easily fooled by seductive but misleading discourse. In a democracy, he argues, those who are experts at garnering votes and nothing else will eventually dominate politics, instead of those who have the required knowledge to govern (the presupposition being that these two classes do not overlap). Most voters do not have sufficient discernment when it comes to issues of governance, but in order to win office or get a piece of legislation passed, politicians must convince the ignorant so as to obtain their support. And so, experts in manipulation and mass appeal will resort to easily digestible messages so as to obtain political power, whereas those with actual knowledge on how to govern but lesser rhetorical skills will not stand a chance.

Recent events in world politics have confirmed that voters are susceptible to false or misleading messages that play into deep-seated sentiments or prejudices (Goodin & Spiekermann 2018, Epilogue). A revealing example is the Brexit referendum in 2016, in particular the infamous Leave slogan: “We send the EU £350 million a week. Let’s fund our NHS instead. Vote Leave.” Leading Leave figures repeated the slogan incessantly, and despite the fact that those with actual knowledge of public finance had repeatedly shown it to be false—thus a quintessential political instance of ‘fake news’—its grip on voters remained powerful.

To what extent voters are indeed susceptible to manipulation by means of seductive but misleading discourse is ultimately an empirical question. A number of well-known results in social and cognitive psychology lend further support to this claim. First, the robust phenomena of confirmation bias, myside bias, and related cognitive tendencies indicate that humans like to hold on to their long-standing beliefs, and thus seek and interpret evidence in ways that minimize revisions, e.g. by avoiding or discrediting strong evidence against their deeply held beliefs (Nickerson 1998). Secondly, the phenomenon of identity-protective cognition refers to the tendency of culturally diverse individuals to selectively credit and dismiss evidence in patterns that reflect the beliefs that predominate in their social group (Achen & Bartels 2017; Kahan 2017), leading to what is sometimes described as ‘tribal politics’. Thirdly, there is much empirical evidence suggesting that we are poorly equipped to detect lies and deception in general, as our default assumption towards other people is that of honest communication; in

a provocative slogan, “we are hardwired to be duped” (Levine 2019; see also Michaelian 2009; Shieber 2012).<sup>5</sup>

But importantly, in Model A, those producing seductive messages to lure voters *do not actively interfere with the messages being sent by others*.<sup>6</sup> Essentially, they still act within the basic tenets of deliberative democracy, which protects free speech (barring outrageous lies in the form of libel or hate speech). In other words, they do not actively interfere with the general structure of the relevant information ecosystem, other than by broadcasting their own messages. These messages, then, can be countered by similarly seductive messages from other political figures reaching the same target audience but with very different content. This should ensure balanced public debates and a certain degree of autonomy<sup>7</sup> for the receiver of these opposing messages in deciding which of them appears more appealing to her, as there is the possibility of discussion and dissent to triangulate (this is one reason why competing voices matter). In this model, knowers can identify credibility markers more or less reliably (even if imperfectly), and can identify and rely on authoritative trustworthy sources (experts, objective journalism, social institutions). This is so even if the circulation of misleading messages in itself constitutes a form of epistemic interference that, in an ideal world, would not occur. (We should perhaps clarify that our investigation here can be described as an exercise in *non-ideal social epistemology*. In particular, the quasi-normative conclusions we draw take the circulation of misleading messages as a given.)

What this model also makes clear is that within liberal democracy there is ample space for political manipulation through the propagation of anti-democratic messages camouflaged as democratic discourse (Stanley 2015). In fact, in the words of historian Robert Moss (1977, 12), “democracy can be destroyed through its own institutions”, as witnessed by the fact that many of the totalitarian regimes of the last century came into power democratically. More generally, the model suggests that persuasiveness rather than truthfulness is the guiding principle for political discourse in a democratic system, which aligns exactly with Plato’s main criticism of democracy.

How does the ideal of epistemic autonomy fare in environments where persuasion rather than truthfulness reigns? Clearly, the presence of misleading or outright false persuasive messages makes it harder to satisfy the epistemic condition on epistemic autonomy. This will be true in all three models under

<sup>5</sup> The extent to which humans are gullible and easily duped, or instead can competently exercise epistemic vigilance (Sperber et al. 2010), remains controversial. For example, Hugo Mercier (2020) argues that we’re quite good at exercising epistemic vigilance. (One of us has an ongoing (friendly) disagreement on these issues with Mercier and Sperber, including in print.)

<sup>6</sup> This is a key difference with Model C, to be discussed below: even if the outcomes might be similar, the details of the mechanisms for epistemic manipulation are different.

<sup>7</sup> Our claim is comparative rather than absolute: there is *more* room for epistemic autonomy in Model A than in Models B and C, even if epistemic autonomy is not fully guaranteed in Model A, given the circulation of misleading messages.

discussion—they are, after all, models that describe strategies for manipulating the public opinion through misinformation. In Model A, however, knowers do retain a fair amount of epistemic autonomy, even though they are bombarded by persuasive discourse coming from different corners. Precisely because diverse messages in support of different positions in the political spectrum are being broadcast, the receiver can, in principle, judge their credibility, assess the reliability of their sources, weigh them against each other, and thus come to her own conclusions as to which of them is more trustworthy. This is not to say that a knower cannot be tricked into believing falsehoods in these circumstances; but at least she is free to draw her own conclusions, within the wide range of messages that she receives. Clearly, careful exercise of judgment about whom to trust becomes more and more important in this model, given the conflicting messages being broadcast by different sources.

## 2.2 Model B: Propaganda and Censorship

In the second model, political figures continue to broadcast messages to promote their own causes and strengthen their position, but they also exercise power to block the production and dissemination of dissenting messages. This model is most clearly instantiated in totalitarian regimes where those in power control the means of production and dissemination of media top-down, such as in the USSR (especially during the Stalin period), Nazi Germany, and more recently in countries such as Turkey, Hungary, Russia, and China (albeit to different degrees). ‘Propaganda’ is the term commonly used to refer to state-produced media content that is intended to bolster support for the leaders and to depict negatively anyone who diverges from the dominant ideology (domestically or internationally).<sup>8</sup> Crucial in this model is the wide use of censorship to prevent the production and dissemination of dissenting, critical messages, represented by bans on books, newspapers, art, etc. with ‘subversive’ content (e.g. the Nazi book burnings). Moreover, in order to keep track of what people think and the content they produce and consume, totalitarian regimes will also typically rely on an extensive spying system.

In both Nazi Germany and the USSR, state-sponsored propaganda was an essential component to ensure compliance and obedience from the masses. The Reich Ministry of Public Enlightenment and Propaganda led by Joseph Goebbels

<sup>8</sup> There are many definitions of the concept ‘propaganda’ available in the literature. Here, we adopt the following working definition: “Communication designed to manipulate a target population by affecting its beliefs attitudes, or preferences in order to obtain behavior compliant with political goals of the propagandist” (Benkler et al. 2018, 29). One important contribution in Stanley (2015) is to show that propaganda thus understood is also pervasive in democratic societies, not only in the totalitarian contexts that are often thought to be the primary loci for propaganda.

(one of the greatest propagandists in world history) was created in 1933, a few months after Hitler came to power, and was tasked with the job of centralizing and controlling all aspects of German cultural and intellectual life. It acted on two main fronts: production of media glorifying the figure of Hitler and Nazi ideology more generally, and censorship of any dissenting message or content (Welch 1993).

Hannah Arendt's *The Origins of Totalitarianism* (1951) is still the philosophical *locus classicus* for an analysis of the two main instances of totalitarianism in the twentieth century, Nazi Germany and the USSR. Arendt emphasizes in particular the power of propaganda to manipulate the masses into compliance, and the fact that truthfulness itself becomes a void concept. The operating principle is rather that of *consistency*; as long as the overall narrative remains consistent and ensures a sense of belonging to something greater than themselves,<sup>9</sup> people are generally unable to resist the allure of propagandistic messages, especially as their overall sense of truth and falsity is already severely undermined.

In an ever-changing, incomprehensible world the masses had reached the point where they would, at the same time, believe everything and nothing, think that everything was possible and that nothing was true.<sup>10</sup> [...] Mass propaganda discovered that its audience was ready at all times to believe the worst, no matter how absurd, and did not particularly object to being deceived because it held every statement to be a lie anyhow. The totalitarian mass leaders based their propaganda on the correct psychological assumption that, under such conditions, one could make people believe the most fantastic statements one day, and trust that if the next day they were given irrefutable proof of their falsehood, they would take refuge in cynicism; instead of deserting the leaders who had lied to them, they would protest that they had known all along that the statement was a lie and would admire the leaders for their superior tactical cleverness.

(Arendt 1968, 80)

Thus understood, fake news propaganda is the bread-and-butter of totalitarian, autocratic regimes, coupled with a situation of scarcity of information and content that results from the active censorship of any message perceived as contrary to the dominant ideology. This involves both curtailing the production of such messages domestically (by silencing opponents; by persecuting, imprisoning, exiling, or even killing dissidents) and controlling the influx of messages from outside. Of

<sup>9</sup> This is a crucial point: the epistemic environment has to be suitably engineered to give rise to the kind of situation described by Arendt. We are thus not claiming that the allure of propagandistic discourse is the same across different epistemic environments.

<sup>10</sup> Tellingly, the phrase "everything was possible and nothing was true" was picked up by Peter Pomerantsev as the title of a book chronicling his experiences working in Russia as a TV-producer in the first decade of the twenty-first century (Pomerantsev 2014).

course, this level of manipulation of the information ecosystem is only fully enforceable in political circumstances of high concentration of power, and is thus incompatible with liberal democracy.

However, somewhat similar situations can also arise in democracies, through the systematic production and maintenance of large-scale media echo chambers: social epistemic structures “in which other relevant voices have been actively discredited” (Nguyen 2018, 2). The right-wing media bubble currently in place in the United States appears to be an example of this (Benkler et al. 2018; see also below). Even so, the model is hard to enforce and maintain in situations where some level of access to uncensored internet and social media is a given. Doing so requires constant efforts of suppression on the part of those in power. Even though their information diet may be just as biased and one-sided, there remains a crucial difference between a Chinese citizen who lacks access to an open internet and an American citizen who spends all his time reading extreme right-wing media outlets. Whereas the latter faces no external impediments to changing his news consumption, the former really has no easy options to do so.

With respect to epistemic autonomy, what Arendt’s description accurately captures is the fact that one of the main goals of totalitarianism is to suppress epistemic autonomy on the part of individuals: those in power attempt to control and manipulate not only what individuals *do*, but also what they *think and believe*. In other words, totalitarian leaders seek to submit individuals to epistemic heteronomy. If Arendt’s analysis is correct, they have sometimes succeeded in doing so.

However, the mere fact that massive resources of time, money, etc. must be diverted to propaganda and censorship in such situations suggests precisely that the pull of epistemic autonomy remains strong. In totalitarian regimes, groups of individuals come together to resist oppression, both political and epistemic. They undertake interventions such as promoting the illegal circulation of forbidden material, at great risk for themselves (as described by Pomerantsev (2019) about his own parents, free thinkers living under the USSR communist regime). Indeed, the idea that, in oppressive situations, epistemic autonomy can be squelched completely seems to underappreciate the potential for epistemic resistance of the oppressed (cf. Medina 2012), which has manifested itself consistently across times and ages.

### 2.3 Model C: Disinformation by Epistemic Pollution

The third model to be considered includes, as the previous two, the production and dissemination of propagandistic discourse, but is chiefly characterized by sustained efforts to pollute the information ecosystem. While in Model B the sender of propagandistic messages also actively works to block the production and dissemination of alternative messages, in this model she instead saturates the

information ecosystem with ‘noise’ so as to interfere with or even prevent the reception of these alternative messages by the target audience.<sup>11</sup> A review of Peter Pomerantsev’s *This is Not Propaganda* (Pomerantsev 2019), a book on the various ‘information wars’ of recent years, sums it up well:

Those rights [to read, to write, to listen to and to say whatever one wishes] now exist almost everywhere, but more information has not necessarily meant more freedom. While autocratic regimes once controlled the narrative by silencing opponents, they now seek to confuse their populations by bombarding them with false information, half truths and competing narratives. It’s a strategy that Pomerantsev describes as “censorship through noise”, or as one of his interviewees, law professor Tim Wu, puts it, states have moved from “an ideology of information scarcity to one of information abundance”. (Bloomfield 2019)

While it is tempting to assume that this is a recent phenomenon intrinsically related to the rise of the Internet, an infamous example from recent but mostly pre-internet history suggests otherwise: the disinformation campaign waged by the tobacco industry to counter and neutralize the dissemination of robust scientific findings linking tobacco consumption with a much higher risk of developing lung cancer (Michaels 2008; Oreskes & Conway 2010). The key principle of the campaign consisted in casting doubt on these scientific findings, and amplifying the reach of the occasional scientific studies (often funded by the tobacco industry itself) that failed to identify a correlation between tobacco and cancer: “to find, fund, and promote research that muddied the waters, made the existing evidence seem less definitive, and gave policy makers and tobacco users just enough cover to ignore the scientific consensus” (O’Connor & Weatherall 2019, 95). (Similar strategies have been deployed by climate change denialists supported by the fossil fuel industry.)

Yet it cannot be denied that the reach of ‘information warfare’ has intensified in the last decade, dovetailing with the increased influence of social media that makes the manipulation of dissemination channels much easier. A number of authors attribute the spread of disinformation campaigns specifically to Russia under Vladimir Putin. As the ‘losers’ in the Cold War, already in the 1990s, new strategies for the manipulation of public opinion were being developed in Russia (Pomerantsev 2019). Pseudo-journalistic content, or fake news, forms an essential component in these strategies. As described by historian Timothy Snyder,

<sup>11</sup> In the words of Steve Bannon, arguably the foremost contemporary propagandist: “The Democrats don’t matter, the real opposition is the media. And the way to deal with them is to flood the zone with shit” (quoted in Illing 2020).

[The term ‘fake news’] sounds like an American invention, and Donald Trump claimed it as his own; but the term was used in Russia and Ukraine long before it began its career in the United States. It meant creating a fictional text that posed as a piece of journalism, both to spread confusion about a particular event and to discredit journalism as such. [These] politicians first spread fake news themselves, then claim that all news is fake, and finally that only their spectacles are real. The Russian campaign to fill the international public sphere with fiction began in Ukraine in 2014, and then spread to the United States in 2015, where it helped to elect a president in 2016. The techniques were everywhere the same, although they grew more sophisticated over time. (Snyder 2018, 11)

One of the main components of these strategies is discrediting traditional sources of information such as mainstream media and scientists and scholars, thus creating an environment of epistemic uncertainty where people feel they can no longer trust those who they hitherto took to be reliable sources. It is precisely at this juncture that the label ‘fake news’ can be weaponized to discredit traditional journalistic reporting, and thus to pave the way for ‘alternative facts’ and narratives that reinforce specific purposes and ideologies. Once the public comes to believe that “nothing is true and everything is possible” (entering a state of epistemic confusion similar to that of those living under totalitarian regimes as described by Arendt), they will limit their attention to information channels that they take to be reliable, which are typically those that confirm their world views. In the United States, for example, it appears that the right-wing information ecosystem is now almost entirely insulated from other media environments, including from the center-right, which has led to radicalization and made its audience more susceptible to foreign and domestic propaganda (Benkler et al. 2018), given the lack of exposure to counter-evidence and different narratives. Indeed, it was often remarked that Fox News effectively functioned as Trump’s own ‘propaganda ministry’.<sup>12</sup>

To what extent can individual knowers remain epistemically autonomous in these circumstances? It’s complicated. On the one hand, because the information ecosystem is flooded with all sorts of information, individuals are forced to make up their own minds and exercise their epistemic autonomy in deciding whom to believe and what to take seriously. In this sense, individuals retain a minimum of autonomy in Model C and it becomes even more crucial to exercise autonomy

<sup>12</sup> Which is not to say that there’s no ‘propagandistic’ journalistic content produced in left-wing media. See, for instance, Frank (2016) for a penetrating analysis of the *Washington Post*’s propagandistic coverage of Bernie Sanders, or Coady (2019), who points out that NPR, the *New York Times*, and the *Washington Post* intentionally refrained from using the term ‘torture’ to describe interrogation techniques that clearly were torture, on the grounds that doing so would amount to choosing political sides. But Benkler et al.’s (2018) data show that the phenomenon is significantly more prevalent and persistent in US right-wing media. Thanks to an anonymous referee for nudging us to consider this point.

responsibly. On the other hand, Model C actively and systematically undermines the resources required to exercise epistemic autonomy responsibly. The trustworthiness of credible sources is attacked, and that of untrustworthy sources artificially inflated; false and misleading messages are introduced as if they come from genuine experts; 'both sides of the issue' are presented as equally worthy of a fair hearing, even when expert opinion clearly comes down on one side. As a result, it becomes harder and harder to use higher-order evidence to allocate trust responsibly and to judge the reliability of messages. In other words, Model C perverts the ideal of epistemic autonomy to undermine itself. Tellingly, *Russia Today*, the Russian government-backed international TV station, has as its slogan 'Question More'; this spurious appeal to epistemic autonomy by a premier purveyor of misinformation embodies the thinking behind Model C.

Consider testimony: the exercise of epistemic autonomy in accepting testimony consists primarily in being discriminating regarding whom we trust (Fricker 2006). But in Model C, propagandists artificially create a perception of untrustworthiness regarding some sources, while artificially inflating the apparent credibility of others. So it seems that, by manipulating allocations of credibility, propagandists really impose a state of epistemic heteronomy onto the agents in question, as they temper precisely with what should ensure the epistemic autonomy of receivers of testimony (their ability to distinguish trustworthy from untrustworthy sources). What's more, since the interventions in question are more veiled than the active censorship in Model B, these processes of epistemic manipulation are more insidious and thus potentially more dangerous than in Model B. Taken to its extreme, then, Model C leaves individual knowers with a phantom of epistemic autonomy: they might feel they are autonomous, whereas in reality they are epistemically heteronomous.

To conclude this section, we have argued that different strategies where fake news (in particular in the form of stories that falsely present themselves as journalistic content) figures prominently may be adopted for the propagation of messages that support the causes and strengthen the position of certain (political) actors. Contrary to what appears to be a popular belief, all three models have been instantiated in the past and continue to be instantiated now. Model C, in particular, which comes closest to descriptions of the current so-called 'post-truth' era, is not an entirely novel phenomenon. But, as we will see in the next section, current technologies seem to enhance the reach and scope of strategies falling under Model C.

### 3. The Putative Novelty of Fake News

As the above three models show, fake news isn't new. That is to say, while the widespread use of the term 'fake news' may be new, the phenomena to which it

refers have been with us for much longer. We can see this even more clearly by considering a few characteristics of contemporary fake news which are sometimes casually put forward as accounting for the novelty of fake news.

### 3.1 Content

The *content* of much contemporary fake news fits seamlessly with historical attempts to manipulate public opinion. Politicians and their spin doctors and enthusiastic supporters cater to voters' preferences and social identities to please them. They present their own views, policies, actions, personalities, and lives in a favourable light—not shunning the use of falsehoods, deception, or 'bullshit' in Frankfurt's (2005) technical sense. At the same time, the views, policies, actions, personalities, and lives of political opponents are discredited through false or one-sided information.<sup>13</sup>

Historians report that already in seventeenth-century England, so-called 'news-books' were published containing sensationalistic false content about war victories and defeats, or even the death of kings, as well as supposedly personal letters with false reports about various events (Young 2016). These venues came and went quickly. The same was true of so-called 'yellow journalism' in the late nineteenth and early twentieth centuries in the USA (Samuel 2016). It was known for its dependence on

the familiar aspects of sensationalism—crime news, scandal and gossip, divorces and sex, and stress upon the reporting of disasters and sports; [...] the lavish use of pictures, many of them without significance, inviting the abuses of picture-stealing and 'faked' pictures; [...] impostures and frauds of various kinds, such as 'faked' interviews and stories. (Mott 1950, 539, quoted in Samuel 2016)

Clearly, all of this sounds eerily familiar, including even the use of misleading pictures and fake quotes, even if current technologies allow for more sophisticated forms of forgery (such as 'deep fakes').<sup>14</sup>

<sup>13</sup> Cf. Robinson et al. (2018) for an overview of common strategies.

<sup>14</sup> For some of these historical examples, it isn't clear whether they fall in the category of politically motivated propaganda or pure profit-seeking 'clickbait'. Most likely, the two have always co-existed and the line between them isn't always sharp. One interesting historical example with tremendous political consequences was the 'Dreyfus affair' at the end of the nineteenth century, where the spread of misinformation by the press played a prominent role.

### 3.2 Proliferation, Circulation, and Influence

A widely shared Buzzfeed piece (Silverman 2016) showed that fake news stories outperformed real news in the final months leading up to the 2016 US election. The Cambridge Analytica scandal has received a lot of attention, and several journalists writing for reputable news outlets (e.g. Cadwalladr 2018; Rosenberg et al. 2018; Wong 2019), in addition to whistleblower Christopher Wylie himself (2019), have floated the suggestion that micro-targeted political ads may have helped Trump win the election. Analysis of Twitter data suggested that falsehoods travelled faster and further than truths (Vosoughi et al. 2018). Perhaps, then, the novelty is that there is much more fake news than before, circulating more widely, and exerting significantly greater influence on the general public.

However, recent empirical research reaches more sobering conclusions. The empirical evidence suggests that fake news is mostly consumed by a small subgroup of heavy internet users (Nelson & Taneja 2018), that it is typically shared by older conservative voters (Grinberg et al. 2019; Guess, Nagler, & Tucker 2019), and that its influence is too small to change election outcomes (Allcott & Gentzkow 2017; Broockman & Green 2014).<sup>15</sup> Moreover, fake news consumption appears to have decreased between 2016 and 2018 (Guess et al. 2019). It's also doubtful that many people truly live in online news bubbles; news diets tend to be relatively diverse (Fletcher & Nielsen 2017; Guess et al. 2018; Nelson & Webster 2017). Of course, it remains extremely difficult to filter out and estimate the exact effects of information consumption on beliefs, attitudes, and behavior; it may well be that further research will ultimately contradict these findings, and that looking at the phenomenon on a more global scale leads to different conclusions. But we can at least say that, at this point, there is no decisive evidence that the circulation and influence of fake news have grown considerably in recent years, at least not in the USA and Europe.<sup>16</sup>

### 3.3 Production, Distribution, and Consumption

Finally, the novelty of fake news might be found in the underlying technologies for producing, distributing, and consuming it. This is independently plausible

<sup>15</sup> Of course, these findings are all specific to the countries studied. In other countries, the spread of fake news seems to have had a considerable influence in election outcomes, such as in the Philippines (Pomerantsev 2019) and Brazil. At the 2018 presidential election in Brazil, for example, the spread of fake news through WhatsApp groups seems to have had considerable impact (Scarabelli 2019).

<sup>16</sup> One caveat: the studies cited do not distinguish between the different uses of 'fake news' that we distinguished above and some combine propaganda, clickbait, and conspiracies in their operationalization of 'fake news'. (Thanks to Hein Duijf for noting this point.) This doesn't threaten our claims, however. If the spread and influence of the combined categories is relatively limited, the same will be true for one subcategory.

anyway, since the rise of the Internet and social media form the most noticeable shift in the news and journalism landscape. (It's not as if human psychology or the general incentives of commercial markets have undergone major changes in recent years.)

A popular account of the influence of the Internet goes as follows. The Internet has made the production of content that is available to a (potentially) worldwide audience much easier. In comparison with the efforts and costs involved in printing a physical newspaper or magazine, the efforts and costs of setting up a website, blog, or social media account are almost negligible. Access, moreover, is often free rather than paid, at least for anyone with an internet connection, consumption is non-exclusive in the sense that no printed copies are necessary, and content can easily remain available for virtually unlimited amounts of time. Features such as these were what led people to express high hopes about the web's potential to enhance freedom, equality, public discourse, and democracy in the early days of the Internet. The Internet was supposed to turn the world into a global village (McLuhan 1964) and act as the great equalizer in the worldwide marketplace of ideas, information, and knowledge, finally realizing the Millian ideal of free exchange of ideas (Mill 1999). Online, everyone's voice would have an equal chance of being heard, everyone could contribute to the conversation, and everyone could simultaneously be a journalist, news consumer, engaged citizen, advocate, and activist.<sup>17</sup>

Very little of this has actually materialized. To be sure, the technical possibilities as such are there—it is still true that anyone can easily put content online and share it with the world, in particular through social media. The point is that the online world has created new forms of inequality, formidable barriers of entry, and virtual monopolies on services and platforms. The influence of tech giants on what gets seen and shared online is hard to overestimate: Google sets the order of search results, YouTube's algorithms throw up recommendations for what to watch next, Facebook's news feed prioritizes your friends' messages, most of our news comes from major outlets rather than local or independent sources, and so on. In his recent book on the online attention economy, the political scientist and media scholar Matthew Hindman paints a gloomy picture:

The number of [news] outlets may have expanded, but the public sphere remains highly concentrated. The number of journalists has plummeted and “fake news” has multiplied, but digital media are just as dependent on a few corporate gatekeepers as ever. Building a consistent news audience remains hugely expensive. The attention economy has doomed most of our civic hopes for the web.

(Hindman 2018, 13–14)

<sup>17</sup> Matthew Hindman (2009, ch. 1) cites a wide range of sources from academia, journalism, politics, public administration, and law who expounded views like the above.

Others concur. Communication scientists Eiri Elvestad and Angela Phillips (2018) list the following items among 'myths of the social media era' debunked by empirical research: 'News personalization will improve plurality, diversity, and ultimately democracy', 'the role of the journalist is merging with the role of the audience', and 'the many are smarter than the few' (meaning that relatively few tech and media giants wield outsized power on what online news gets produced, shared, and consumed). Understanding in what ways the Internet fails to be the democratizing information paradise that people had hoped for does, however, provide insight into what may be distinctive of recent online fake news production and dissemination.

First, contrary to the promises of the early Internet, getting content (including propagandistic fake news) not only out, but actually seen regularly by sizable audiences over longer periods of time—which is required to influence the public opinion in any significant way—requires sustained time and effort and is thus expensive. Russia, as well as the American right, have understood this best, it seems. As both Snyder (2018) and Pomerantsev (2019) document in detail, Russia has invested heavily in information warfare over the past decades by creating and operating fake news websites, fake social media accounts, troll farms, armies of bots, and more. The American right has similarly invested heavily in expanding and transforming its already powerful offline media empire into an online media universe.<sup>18</sup> Benkler et al. (2018) show how this has resulted in a large and mostly isolated echo chamber in which fake news and highly partisan content can be produced by extreme websites of the likes of, e.g. Alex Jones's Infowars, to be then gradually picked up by slightly more legitimate-seeming journalistic outlets such as Breitbart, Daily Caller, and Fox News. All of this content, of course, is shared and promoted on social media platforms. Some of it may eventually make it out of the right-wing universe into more mainstream media, if only to be rebutted or commented on.<sup>19</sup> One novel aspect of recent fake news, then, is not that a lot of money is spent on producing and distributing it; rather, it is that its production and dissemination have adapted to the technological and commercial possibilities of the Internet and social media technology.

Secondly, we've noted above that propagandistic fake news has always catered to the preferences and baser instincts of news audiences. What is novel, however, is how recent purveyors of fake news have exploited opportunities opened up by social media platforms and, specifically, tinkered with big data and the algorithms

<sup>18</sup> Benkler et al. (2018) locate the origins of this media environment in the right-wing talk shows that proliferated in the final decades of the twentieth century. Several journalistic reports suggest a strong influence of the libertarian former hedge-fund manager Robert Mercer behind the scenes (Cadwalladr 2017; Mayer 2017).

<sup>19</sup> Again, our singling out the *right-wing* media universe here isn't a matter of personal political preferences on our part; Benkler et al. (2018) make a point of emphasizing how their research shows that the kind of systematic online network propaganda created on the right side of the American political spectrum simply isn't mirrored on the left.

by which these platforms order, prioritize, distribute, boost, and recommend various kinds of content. As before, this requires understanding of human psychology in order to determine what sort of content will appeal to people. But, now, in addition, it also requires technical knowledge of how the relevant algorithms work in order to game them into boosting specific content.

There are two sides to this. One concerns the human–platform interaction. New platforms do not have clear norms for how users behave on them and, to the extent that norms emerge, they can change quickly. As a result, the barriers for sharing misinformation have lowered. For example, many used to quote the phrase ‘retweet isn’t endorsement’ in their Twitter bio. But if not endorsement, then what? What does it mean when people retweet or subtweet someone else’s message? It can mean many things, ranging from endorsement to rejection, to irony, jokes, virtue signalling, expressing one’s social identity, moral grandstanding, etc. (cf. Rini 2017, 49; Sullivan 2019). In contrast with good old-fashioned lying or deceiving people, online sharing of misinformation comes with inbuilt plausible deniability. One can always back-pedal, saying it was just a joke, irony, cynicism, or ‘something to think about’.

The second side has to do with the workings of the algorithms behind internet platforms. Their technical details are often proprietary information and can change frequently,<sup>20</sup> but some basics should suffice to illustrate the point. Google’s PageRank orders search results based, in part, on the amount of incoming links to a webpage (Page et al. 1999). If you want a website to show up higher, it needs to be linked to by many other sites. Creating a large enough network of mutually referring websites thus helps to promote your content in the order of search results. Facebook’s news feed algorithm takes into account, among many other factors, the amount of engagement posts generate (reactions, shares, comments). To artificially boost content, then, you not only make use of the traditional features that make fake news appealing (recall the discussion of yellow journalism and tabloid journalism above); you can also set up fake accounts and bots to engage disproportionately with certain posts. Most social media platforms prime for engagement based on affective responses: ‘likes’, ‘loves’, up or down votes, or Facebook’s more fine-grained emoticons.<sup>21</sup> Hence, they set users up for ‘hot cognition’ (Lodge & Taber 2005) and ‘emotional contagion’ (Hatfield et al. 1993; Kramer et al. 2014), which are known to produce stronger behavioral effects than mere ‘cool’ analysis and critical thinking and deliberation. YouTube’s algorithms are optimized to maximize ‘watch-through’: to get people to watch videos for as long as possible so that more ads can be shown. This can turn YouTube into a

<sup>20</sup> Hindman (2018, 177–8) suggests that this is why Russia and China in particular have used hacking and espionage to obtain insider knowledge of social media algorithms.

<sup>21</sup> Thanks to Alessandra Tanesini for alerting us to this.

‘radicalization machine’: drawing users to ever more extreme videos because that keeps them hooked (Chaslot 2019; Lewis 2018; Roose 2019).

In conclusion, a second aspect that appears novel about contemporary fake news is how it cleverly exploits the opaque and evolving behavioral norms on the Internet and social media platforms, as well as their technical possibilities. Both the kind of content that is produced<sup>22</sup> and its modes of production and dissemination are aimed at exploiting these norms and possibilities, in order to maximize reach and impact.

#### 4. Conclusion

We started by describing three models for spreading politically charged messages with the purpose of supporting the causes of (political) actors. Initially, we focused on structural features of each of these models rather than on the specifics of *how* each of them is implemented. Model A consists in broadcasting alluring messages while not actively interfering with the dissemination of similar messages by other actors; this stays closest to the ‘standard’ model for liberal democratic discourse. Model B involves active suppression and undercutting of alternative messages, in particular through censorship. Totalitarian regimes such as in Nazi Germany and currently in China are typical (but not the only) instantiations of this model, which operates on the basis of *information scarcity*. Model C goes in the opposite direction by implementing strategies of *information abundance*, or ‘censorship through noise’ (Pomerantsev, 2019). In all three models, fictional stories posing as pieces of journalism—fake news—occupy a prominent position. The models are not mutually exclusive, and specific actors may well engage in mixed strategies. Russia, for example, tends towards Model B domestically but employs Model C internationally.

Various historical and contemporary examples showed that all three models can be, and have been, instantiated with different kinds of technologies for information dissemination. This means that disinformation campaigns and so-called ‘post-truth politics’ as such are not novel phenomena, neither with respect to the kind of content produced, nor, as far as recent empirical evidence shows, in terms of proliferation, circulation, and influence.

However, with respect to production, distribution, and consumption, we argued that digital media and the Internet afford a number of technological possibilities that change the informational landscape: epistemic networks have

<sup>22</sup> To prevent misunderstandings: this does not invalidate the first point made above that the content of contemporary fake news closely resembles that of older forms of misinformation and propaganda. Our point is that sensationalism and catering to people’s preferences and identities is carefully tailored to features of the algorithms that determine what people get to see online.

always existed, but structural differences between networks (e.g. analogue vs. digital) change the ways in which information is produced, distributed, and consumed. These technological possibilities provide access to a virtually infinite range of messages on the Internet, which makes Model C easier to implement than before, whereas Model B becomes more difficult to implement and enforce (though not impossible, as the case of China shows). As for Model A, if everyone else (both domestically and internationally) is engaging in Model C strategies, those who stick to the democratically more acceptable Model A will be disadvantaged. The troubling conclusion seems to be that, in an arms race, everyone will be pushed towards Model C information warfare, whether they want to or not. This poses a dilemma for the supporters of traditional democratic values: should they 'go high where others go low', or should they adopt the same 'dirty tricks' to overturn or at least counter political actors with anti-democratic tendencies?

In sum, we propose a cautious and qualified conclusion concerning what's new about contemporary fake news: much of its features resemble older forms of fake news and propaganda very closely. But two aspects seem novel: (1) the adaptation to the commercial and technological possibilities of the Internet and social media for distribution and consumption and (2) the clever use of big data, algorithmic boosting, and troll farms or fully automated social media bots.

We think this qualified account offers a number of advantages in comparison with other recent discussions of fake news. In particular, it allows us to both reject alarmism about fake news *and* to support calls for investigating and monitoring these developments closely (as the European Union, for example, is already doing).<sup>23</sup> On the one hand, we can side with those who reject the recent alarmism over fake news and 'post-truth politics' by emphasizing that propaganda, misinformation, and other phenomena closely resembling contemporary fake news have been with us for quite some time before the advent of the Internet (at the very least since the twentieth century and arguably even well before). On the other hand, though, we can also support those who call for more scholarly attention, public scrutiny, and perhaps even regulatory policy and legal measures pertaining to disinformation campaigns and the role of fake news therein. A number of democracies in the world are under threat (e.g. Brazil, India, Poland, Hungary), and there's quite some evidence supporting the claim that online disinformation campaigns play a significant role in these developments. The digital media environment has created distinctively new possibilities for fake news and other forms of disinformation to influence the public opinion; it is crucially important for the health of democracy that we grapple with these novel developments.<sup>24</sup>

<sup>23</sup> Report 'Tackling online disinformation': <https://ec.europa.eu/digital-single-market/en/tackling-online-disinformation>.

<sup>24</sup> Thanks to Elias Antilla, Heijn Duijf, Thirza Lagewaard, Chris Ranalli, Merel Talbi, and Alessandra Tanesini for discussion about the chapter and to the editors of this volume and two anonymous referees for valuable comments on an earlier version. Jeroen de Ridder's research for this

## References

Achen, Christopher and Larry Bartels. 2017. *Democracy for Realists*. Princeton, NJ: Princeton University Press.

Allcott, Hunt and Matthew Gentzkow. 2017. “Social media and fake news in the 2016 election”. *Journal of Economic Perspectives* 31(2): 211–36.

Arendt, Hannah. 1968. *Totalitarianism: Part Three of The Origins of Totalitarianism*. San Diego, CA: Harvest.

Bell, Chris. 2018. “The people who think 9/11 may have been an ‘inside job’”. BBC News, 1 Feb. 2018, <https://www.bbc.com/news/blogs-trending-42195513>.

Benkler, Y., R. Faris, and H. Roberts. 2018. *Network Propaganda*. New York: Oxford University Press.

Bloomfield, Steve. 2019. “*This Is Not Propaganda* by Peter Pomerantsev review – quietly frightening”. *The Guardian*, 10 Aug. 2019, <https://www.theguardian.com/books/2019/aug/10/this-is-not-propaganda-peter-pomerantsev-review>.

Broockman, David and Donald Green. 2014. “Do online advertisements increase political candidates’ name recognition or favorability? Evidence from randomized field experiments”. *Political Behavior* 36(2): 263–89.

Brown, Étienne. 2019. “‘Fake news’ and conceptual ethics”. *Journal of Ethics & Social Philosophy* 16(2): 1–11.

Cadwalladr, Carole. 2017. “Robert Mercer: the big data billionaire waging war on mainstream media”. *The Guardian*, 26 Feb. 2017, <https://www.theguardian.com/politics/2017/feb/26/robert-mercier-breitbart-war-on-media-steve-bannon-donald-trump-nigel-farage>.

Cadwalladr, Carole. 2018. “I made Steve Bannon’s psychological warfare tool’: meet the data war whistleblower”. *The Guardian*, 18 Mar. 2018, <https://www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-facebook-nix-bannon-trump>.

Chaslot, Guillaume. 2019. “The toxic potential of YouTube’s feedback loop”. *Wired*, 13 July 2019, <https://www.wired.com/story/the-toxic-potential-of-youtubes-feedback-loop/>.

Coady, David. 2019. “The trouble with ‘fake news’”. *Social Epistemology Review and Reply Collective* 8(10): 40–52.

D’Ancona, Matthew. 2017. *Post-Truth: The New War on Truth and How to Fight Back*. New York: Random House.

Davis, Evan. 2017. *Post-Truth: Why We Have Reached Peak Bullshit and What We Can Do About It*. London: Little, Brown.

Elvestad, Eiri and Angela Phillips. 2018. *Misunderstanding New Audiences: Seven Myths of the Social Media Era*. London: Routledge.

Fallis, Don and Kay Mathiesen. 2019. “Fake news is counterfeit news”. *Inquiry*, online first, doi: 10.1080/0020174X.2019.1688179.

Fletcher, Richard and Rasmus Kleis Nielsen. 2017. “Are news audiences increasingly fragmented? A cross-national comparative analysis of cross-platform news audience fragmentation and duplication”. *Journal of Communication* 67(4): 476–98.

Frank, Thomas. 2016. “Swat Team: the media’s extermination of Bernie Sanders, and real reform”. *Harper’s Magazine*, November 2016, <https://harpers.org/archive/2016/11/swat-team-2/>.

Frankfurt, Harry. 2005. *On Bullshit*. Princeton, NJ: Princeton University Press.

Fricker, Elizabeth. 2006. “Testimony and Epistemic Autonomy”. In Jennifer Lackey and Ernest Sosa (eds), *The Epistemology of Testimony*. Oxford: Oxford University Press, 225–50.

Gelfert, Axel. 2018. “Fake news: a definition”. *Informal Logic* 38(1): 84–117.

Goodin, Robert and Kai Spiekermann. 2018. *An Epistemic Theory of Democracy*. Oxford: Oxford University Press.

Grasswick, Heidi. 2019. “Epistemic Autonomy in a Social World of Knowing”. In Heather Battaly (ed.), *The Routledge Handbook of Virtue Epistemology*. London: Routledge, 196–208.

Grinberg, Nir, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. 2019. “Fake news on Twitter during the 2016 U.S. presidential election”. *Science* 363(6425): 374–8.

Guess, Andrew, Benjamin Lyons, Jacob Montgomery, Brendan Nyhan, and Jason Reifler. 2019. “Fake news, Facebook ads, and misperceptions: assessing information quality in the 2018 U.S. midterm election campaign”. *Democracy Fund report*, <https://www.dartmouth.edu/~nyhan/fake-news-2018.pdf>.

Guess, Andrew, Brendan Nyhan, Benjamin Lyons, and Jason Reifler. 2018. “Avoiding the echo chamber about echo chambers”. *Knight Foundation report*, [https://kf-site-production.s3.amazonaws.com/media\\_elements/files/000/000/133/original/Topos\\_KF\\_White-Paper\\_Nyhan\\_V1.pdf](https://kf-site-production.s3.amazonaws.com/media_elements/files/000/000/133/original/Topos_KF_White-Paper_Nyhan_V1.pdf).

Guess, Andrew, Jonathan Nagler, and Joshua Tucker. 2019. “Less than you think: prevalence and predictors of fake news dissemination on Facebook”. *Science Advances* 5(1): eaau4586, doi: 10.1126/sciadv.aau4586.

Habgood-Coote, Joshua. 2019. “Stop talking about fake news!” *Inquiry* 62(9–10): 1033–65.

Habgood-Coote, Joshua. 2020. “Fake news, conceptual engineering, and linguistic resistance: reply to Pepp, Michaelson, and Sterken, and Brown”. *Inquiry*, online first, doi: 10.1080/0020174X.2020.1758770.

Hatfield, Elaine, John T. Cacioppo, and Richard L. Rapson. 1993. “Emotional contagion”. *Current Directions in Psychological Science* 2(3): 96–100.

Hindman, Matthew. 2009. *The Myth of Digital Democracy*. Princeton, NJ: Princeton University Press.

Hindman, Matthew. 2018. *The Internet Trap*. Princeton, NJ: Princeton University Press.

Illing, Sean. 2020. “‘Flood the zone with shit’: how misinformation overwhelmed our democracy”. Vox.com, 6 Feb. 2020, <https://www.vox.com/policy-and-politics/2020/1/16/20991816/impeachment-trial-trump-bannon-misinformation>.

Kahan, Dan. 2017. *Misconceptions, Misinformation, and the Logic of Identity-Protective Cognition*. New Haven, CT: Yale Law School.

Kakutani, Michiko. 2019. *The Death of Truth*. New York: Tim Duggan Books.

Kramer, Adam D.I., Jamie E. Guillory, and Jeffrey T. Hancock. 2014. “Experimental evidence of massive-scale emotional contagion through social networks”. *Proceedings of the National Academy of Sciences* 111(24): 8788–90.

Levine, Timothy. 2019. *Duped: Truth-Default Theory and the Social Science of Lying and Deception*. Tuscaloosa: University of Alabama Press.

Levy, Neil. 2017. “The bad news about fake news”. *Social Epistemology Review and Reply Collective* 6(8): 20–36.

Lewis, Paul. 2018. “‘Fiction is outperforming reality’: how YouTube’s algorithm distorts truth”. *The Guardian*, 2 Feb. 2018, <https://www.theguardian.com/technology/2018/feb/02/how-youtubes-algorithm-distorts-truth>.

Lodge, Milton and Charles Taber. 2005. “The automaticity of affect for political leaders, groups, and issues: an experimental test of the hot cognition hypothesis”. *Political Psychology* 26(3): 455–82.

McIntyre, Lee. 2018. *Post-Truth*. Cambridge, MA: MIT Press.

McLuhan, Marshall. 1964. *Understanding Media: The Extensions of Man*. London: Routledge & Kegan Paul.

Mayer, Jane. 2017. “The reclusive hedge-fund tycoon behind the Trump presidency”. *The New Yorker*, 17 Mar. 2017, <https://www.newyorker.com/magazine/2017/03/27/the-reclusive-hedge-fund-tycoon-behind-the-trump-presidency>.

Medina, José. 2012. *The Epistemology of Resistance*. New York: Oxford University Press.

Mercier, Hugo. 2020. *Not Born Yesterday: The Science of Who We Trust and What We Believe*. Princeton, NJ: Princeton University Press.

Michaelian, Kourken. 2009. “In defence of gullibility: the epistemology of testimony and the psychology of deception detection”. *Synthese* 176(3): 399–427.

Michaels, David. 2008. *Doubt Is Their Product*. New York: Oxford University Press.

Mill, John Stuart. 1999. *On Liberty*. Peterborough, ON: Broadview Press.

Moss, Jessica. 2007. “The doctor and the pastry chef: pleasure and persuasion in Plato’s *Gorgias*”. *Ancient Philosophy* 27(2): 229–49.

Moss, Robert. 1977. *The Collapse of Democracy*. London: Abacus.

Mott, Frank. 1950. *American Journalism*. New York: Macmillan.

Mukerji, Nikil. 2018. "What is fake news?" *Ergo* 5: 923–46.

Nelson, Jacob and Harsh Taneja. 2018. "The small, disloyal fake news audience: the role of audience availability in fake news consumption". *New Media & Society* 20(10): 3720–37.

Nelson, Jacob L. and James G. Webster. 2017. "The myth of partisan selective exposure: a portrait of the online political news audience". *Social Media + Society* 3(3): 1–13, doi: 10.1177/2056305117729314.

Nguyen, C. Thi. 2018. "Echo chambers and epistemic bubbles". *Episteme*, online first, doi: 10.1017/epi.2018.32.

Nickerson, Raymond. 1998. "Confirmation bias: a ubiquitous phenomenon in many guises". *Review of General Psychology* 2(2): 175–220.

O'Connor, Cailin and James Owen Weatherall. 2019. *The Misinformation Age*. New Haven, CT: Yale University Press.

Oreskes, Naomi and Erik Conway. 2010. *Merchants of Doubt*. New York: Bloomsbury.

Page, Lawrence, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. *The PageRank Citation Ranking: Bringing Order to the Web*. Technical Report, Stanford InfoLab.

Pepp, Jessica, Eliot Michaelson, and Rachel Sterken. 2019. "What's new about fake news?" *Journal of Ethics and Social Philosophy* 16: 67–94.

Pomerantsev, Peter. 2014. *Nothing Is True and Everything Is Possible*. New York: Public Affairs.

Pomerantsev, Peter. 2019. *This Is Not Propaganda*. London: Faber & Faber.

Rini, Regina. 2017. "Fake news and partisan epistemology". *Kennedy Institute of Ethics Journal* 27(2S): E43–E64.

Robb, Amanda. 2017. "Anatomy of a fake news scandal". *Rolling Stone*, 16 Nov. 2017, <https://www.rollingstone.com/politics/politics-news/anatomy-of-a-fake-news-scandal-125877/>.

Robinson, Piers, David Miller, Eric Herring, and Vian Bakir. 2018. "Lying and deception in politics". In Jörg Meibauer (ed.), *The Oxford Handbook of Lying*. Oxford: Oxford University Press, 529–40.

Roose, Kevin. 2019. "The making of a YouTube radical". *New York Times*, 8 June 2019, <https://www.nytimes.com/interactive/2019/06/08/technology/youtube-radical.html>.

Rosenberg, Matthew, Nicholas Confessore, and Carola Cadwalladr. 2018. "How Trump consultants exploited the Facebook data of millions". *New York Times*, 17 Mar. 2018, <https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html>.

Samuel, Alexandra. 2016. "To fix fake news, look to yellow journalism". *JSTOR Daily*, 29 Nov. 2016, <https://daily.jstor.org/to-fix-fake-news-look-to-yellow-journalism/>.

Scarabeli, Ana Clara. 2019. "How did fake news run voters' opinions in the Brazilian elections?". *Diggit Magazine*, 2 Nov. 2019, <https://www.diggitmagazine.com/articles/fake-news-brazilian-elections>.

Shieber, Joseph. 2012. "Against credibility". *Australasian Journal of Philosophy* 90(1): 1–18.

Silverman, Craig. 2016. "This analysis shows how viral fake election news stories outperformed real news on Facebook". *BuzzFeed News*, 16 Nov. 2016, <https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook>.

Silverman, Craig. 2017. "I helped popularize the term "fake news" and now I cringe every time I hear it". *BuzzFeed News*, 31 Dec. 2017, <https://www.buzzfeednews.com/article/craigsilverman/i-helped-popularize-the-term-fake-news-and-now-i-cringe>.

Silverman, Craig and Lawrence Alexander. 2016. "How teens in the Balkans are duping Trump supporters with fake news". *BuzzFeed News*, 3 Nov. 2016, <https://www.buzzfeednews.com/article/craigsilverman/how-macedonia-became-a-global-hub-for-pro-trump-misinfo>.

Snyder, Timothy. 2018. *The Road to Unfreedom*. London: Vintage.

Søe, Sille Obelitz. 2019. "A unified account of information, misinformation, and disinformation". *Synthese*, online first, doi: 10.1007/s11229-019-02444-x.

Sperber, Dan, Fabrice Clément, Christophe Heintz, Olivier Mascaro, Hugo Mercier, Gloria Origgi, and Deirdre Wilson. 2010. "Epistemic vigilance". *Mind & Language* 25(4): 359–93.

Stanley, Jason. 2015. *How Propaganda Works*. Princeton, NJ: Princeton University Press.

Subramanian, Samanth. 2017. "Inside the Macedonian fake news complex". *Wired*, 15 Feb. 2017, <https://www.wired.com/2017/02/veles-macedonia-fake-news/>.

Sullivan, Emily. 2019. "Beyond testimony: when online information sharing is not testifying". *Social Epistemology Review and Reply Collective* 8(10): 20–4.

Vosoughi, Soroush, Deb Roy, and Sinan Aral. 2018. "The spread of true and false news online". *Science* 359(6380): 1146–51.

Welch, David. 1993. *The Third Reich: Politics and Propaganda*. London: Routledge.

Wong, Jennifer. 2019. "The Cambridge Analytica scandal changed the world – but it didn't change Facebook". *The Guardian*, 17 Mar. 2019, <https://www.theguardian.com/technology/2019/mar/17/the-cambridge-analytica-scandal-changed-the-world-but-it-didnt-change-facebook>.

Wylie, Christopher. 2019. *Mind\*ck: Cambridge Analytica and the Plot to Break America*. New York: Random House.

Young, Francis. 2016. "Fake news: a very (early) modern problem", <https://manyheadedmonster.wordpress.com/2016/12/05/fake-news-a-very-early-modern-problem/>.

Zagzebski, Linda. 2012. *Epistemic Authority*. New York: Oxford University Press.

# How Vice Can Motivate Distrust in Elites and Trust in Fake News

*Maura Priest*

## 1. Introduction

### 1.1 Conservatives, Liberals, Elites, and Experts

Distrust of the so-called “elites” has long been a talking point of US conservative commentaries and news outlets. The point was simple: *do not trust the news. What you see, is \*not\* what you get.*<sup>1</sup> This distrust often went beyond what might be considered “news” in the narrow sense, i.e., the newspaper, television, radio, and today, websites and blogs. Distrust had (and continues) to spread to all members of a certain class that conservative identified persons tend to identify as “liberal elites,” “the left,” “progressives,” or, more simply, “liberals.”

Within the United States (and at times, outside these borders) the term, “liberal,” had and has been used in a vague colloquial fashion, seeming to mean, roughly, “persons with political commitments that align closer to the US Democratic Party than the US Republican Party.” These “liberals” are often perceived to overlap (and are thereby associated) with a class you might call “experts,” i.e., persons known to have special and superior intellectual status regarding a limited domain or subject matter, e.g., medicine, economics, health-care policy, international trade, evolutionary theory, climate science, and biology. Members sometimes referenced as part of the class of “liberals,” “the left,” “the elites,” “experts,” or “the liberal elites,” include journalists, academics, or scholars, intellectual experts of any kind, scientists, celebrities, and/or persons holding left-leaning (or perceived to be left-leaning) careers, e.g., artists, musicians, social workers, anthropologists, archeologists, among others.<sup>2</sup> For simplicity,

<sup>1</sup> The footnotes that follow contain many references to works on distrust or the untrustworthiness of elites. Others that specifically mention “liberal elites” include Kearns, 2018; Liao, 2016; and Rankin, 2019.

<sup>2</sup> An apt quote comes from a National Review article, “The Calvin Ball World of Elite White Liberals”: “These liberal-dominated institutions, tertiary education, journalism, entertainment...” (Spiliakos, 2018: para. 4). Along similar lines, both scholarly work and popular articles discussing the tendency of conservatives and conservative-leaning groups to distrust elites and “liberals” and to draw

throughout the remainder of this chapter the term “experts,” and to a lesser extent “elites,” will be used to reference the vague class just mentioned. *Experts* was chosen as the central “catch-phrase” in virtue of its neutrality. Labels like “elites,” and “liberal elites” might be read as unjustifiably elitist on the one hand, or unjustifiably derogatory on the other hand. While “liberal” might not face the same problems, there are other concerns with this turn of phrase. Persons might get caught up on the colloquial vs. the scholarly meaning of the word; a debate is tangential to this chapter.

We should emphasize that the class of experts or liberal elites is defined *not* in virtue of some objective quality shared by all members, but rather, the defining commonality is that members are lumped together and categorized as one class by persons who themselves have a common class categorizing feature, i.e., identifying as “conservative.” While perhaps there are objective qualities that *many or most*, e.g., “liberal elites,” in fact share, these qualities are not essential to class membership. What is essential, rather, is the subjective label bestowed on the class by those who subjectively identify as conservative.<sup>3</sup> They are persons conservatives see as political and ideological opponents, and then having this opposition as fundamental, a cluster of other characteristics are consistently thrown into the mix. The liberals do not exhaust all political opponents of the self-identified conservative. But typically, they are the ones seen (and portrayed) as the most prominent political opponent, and it is this class (or some rough approximation of it) with which there is an extended history of distrust, criticism, and sometimes even contempt.

## 1.2 Distrust, Politics, and Fake News

Distrust of experts has recently intensified, especially during, in the wake of, and in relation to, the 2016 US presidential election.<sup>4</sup> Of course, what qualifies as distrust is a philosophical debate all on its own. But for our purposes specificities of distrust can be set aside. We can think of distrust intuitively, as a vague sentiment (that plays out in words and actions) that a particular entity is unreliable, i.e., not something that one can count on to convey the truth, nor to do what is in the distrusting agent’s interest. The former is the epistemic side of distrust:

an association between the elites and liberals include Berkowitz, 2018; Masciotta, 2016; Mudde, 2007; Phillips-Fein, 2019; Rydgren, 2005; Stark, 1996; Tanehaus, 2017; and Tomasky, 2017.

<sup>3</sup> Of course, not all persons who identify as conservatives bestow the described labels on the amorphous class of “experts.” However, enough of them do that a meaningful discussion of the class is possible.

<sup>4</sup> Articles (news articles and other not necessarily scholarly articles relevant to this “new and improved” mistrust of elites) include Bruinius, 2018; Davies, 2018; and Taub, 2016, and scholarship on this phenomenon includes Eatwell & Goodwin, 2018; Greven, 2016; and Neiwert, 2016.

The type of distrust that labels the distrusted entity as the type of entity that either purposely or accidentally misrepresents reality. The latter type is more closely aligned with either ethics or perhaps just something tied to social interactions/relationships. This type of distrust can be present even without epistemic distrust. It is distrust that is not necessarily skeptical of an entity's ability to represent reality, but rather, it is skepticism of an agent's willingness, or ability, to act in a way that the distrusting agent would find acceptable, appropriate, or helpful. Conservatives' distrust of liberals consists in both kinds of distrust, and in this chapter the context itself can reveal the type at issue (which sometimes includes both types).

Distrust becomes especially notable in recent debates (scholarly and non-scholarly) about *fake news*.<sup>5</sup> “Fake news” taken somewhat literally, seems to reference a contemporary information source that presents as reliable, accurate, and honest; yet which is perceived (maybe accurately) as, false, misleading, dishonest, etc. Political divisions often come alongside divisions over “which news is fake news.” Persons on the political left might see fake news in right-wing media biases that result in misleading or inaccurate reporting. However, many conservative identified persons distrust news sources that they perceive as having “liberal bias.” Moreover, sources considered “mainstream” are often assumed to be entrenched in this unfair leftist perspective. So, it is the mainstream media (that is actually the left-leaning media), and even more so “openly liberal” news sources, that are the propagators of fake news (i.e., these are the propagators according to many who identify as conservative).

### 1.3 Liberal Bias

Based on accusations from the conservative identified, “liberal bias” points to unreliable media/news sources, where the unreliability desire of the source (reporter, news agency, etc.) to promote a particular type of “left-leaning” ideological viewpoint.<sup>6</sup> This fits with the colloquial use of “bias” more generally: rough, “being unfair and unreasonable insofar as persons put their own values and preferences before what is reasonable or right.” Consider the common (often joking) suggestion that persons are “biased” if their statements can be interpreted as self-serving or self-promoting. Here is an example from a prominent philosophy blog:

<sup>5</sup> Associations between distrust and fake news can be found in Freeze et al., 2020; Mcgonagle, 2017; Moravec et al., 2019; and Tubbs, 2019.

<sup>6</sup> For examples of how conservative identified describe the left media bias, see Crews Jr., 2020; Groseclose, 2011; and Haskins, 2017.

I think more students need to seriously consider writing a dissertation in the history of philosophy. There is not only a demand for history going forward, but the skills acquired have many uses outside of academia, i.e. becoming a teacher at a private high school or an archivist. But maybe I'm biased.

*Signed – Professor and Historian of Philosophy*

The “But maybe I’m biased” phrase seems to acknowledge to philosophers an awareness that the poster’s own AOS might have less than innocent epistemic influence, i.e., communicating, “yes, I have a self-interested reason to say this.” In addition, because of the self-interest, there will clearly be “bias”—that is, an epistemically unjustified tendency to believe a certain proposition (or set of propositions.) Lastly, the use of “maybe” seems almost sarcastic, as if to jokingly suggest that *we all know* that the odds of bias are much greater than a mere “maybe” implies.

Given what was just said, *liberal bias* would seem to reference a tendency to assert or believe certain propositions that (1) fall short of the truth and (2) fall short because of an unjustified tendency to assert or believe propositions that support a left-leaning ideology. More specifically, the conservative identified might see liberally biased media sources as falling epistemically short in all of the following ways:

- (1) Promote leftist political ends, i.e., the election of a particular candidate, or the success or failure of a particular piece of legislation. (For example, in US election coverage presents Democrat candidates more favorably than they deserve, and Republican candidates less favorably.)
- (2) Supports the ideological world view of “the left.” (For example, lends support to the claim that unregulated capitalism is ethically suspect and harms innocent persons with no comparable benefit. More generally, lends support to claims that fit with a leftist world view.)
- (3) Calls into question the ideological worldview of “the right.” (For example, questions the accuracy of religious values, and/or suggests that religious persons have false and harmful beliefs.)
- (4) Paints only a partial picture of the truth in a way that either supports the political ends of the left, hinders the political ends of the right, supports the ideological world view of the left, or calls into question the ideological world view of the right.

The above list is *not* an attempt to state facts about the ideological world view of the left, nor the right, nor to make any claims about the reliability of any news or media outlet, nor are any claims being made about the bias or lack of bias within any ideological group. The description points to a possible perspective of (at least

some) persons who identify as conservative. I say “possible” because further certainty would be an empirical exercise in sociology or political science, rather than a theoretical (epistemological and ethical) philosophy paper. The possible perspectives matter regardless of truth, because the ethical and epistemological results will follow nonetheless. Whether or not the left actually has certain biases, it remains true that *if they were perceived to have these biases*, that there are epistemically better and worse ways to respond to this perception. Moreover, even if conservatives do not worry about “the left” in the way described, it still remains true that *if someone were to worry about the left (or any other social group) in this particular kind of way*, that there are better and worse epistemological reactions therein.

Lastly, it’s worth noting that we can put aside questions about politics, fake news, and ideological division, and *still* come away from this chapter with epistemic value. This chapter offers a conceptual account of two under-discussed epistemic vices, i.e., *epistemic insensitivity* and *epistemic obstruction*. These vices can befall anyone (regardless of whether this chapter is correct that the vices are especially problematic when manifest via experts).

## 2. Justifying the Discussion

In what follows, this chapter describes two vices that tend to worsen phenomena that magnify group polarization, distrust, and miscommunication, especially between vices of conservative identified persons and those that conservatives identify as *experts* (or sometimes, “the left,” “progressives,” “liberal elites,” and so on). The claim is *not* that experts (if such a class even exists) manifest either the vice of epistemic insensitivity or the vice of epistemic obstruction at a higher rate than any other class, nor that they manifest these vices to an especially extreme degree. After all, whether or not elites manifest a vice to a greater extent, or in greater frequency, are empirical questions, and so is the following: Are epistemic obstruction and epistemic insensitivity causally responsible for communication problems, and/or responsible for worsening political distrust?

Putting the empirical quandaries aside, we *can* say the following: Given the description of liberal elites by some conservative identified persons, the vices of epistemic insensitivity and epistemic obstruction become noteworthy because the conceptual contours of the former and latter suggest that the vices can enable or motivate distrust, political division, and miscommunication. If we care about minimizing the aforementioned, then we should pay attention to epistemic insensitivity and epistemic obstruction. Attention need *not* be justified on the

grounds that insensitivity and obstruction are *more problematic*, nor *more common*, than other vices. Rather, (1) *all epistemic vices* (by definition) are problematic and worth watching out for, at least or until it is determined the vices are particularly low risk, and (2) if these vices do manifest in “experts,” this comes alongside noteworthy epistemic threats.

### 3. Interpersonal Vice

If Sid is an expert researcher, this expertise speaks epistemically well of Sid (all else held equal). However, suppose that under his tutelage, Sid’s grad student researchers acquire many false beliefs, and many true but unjustified beliefs. Even more, Sid’s grad researchers (under his tutelage) acquire several lasting epistemic vices.

Despite Sid’s admirable epistemic *personal* qualities, his negative epistemic influence over his grad students suggests that Sid also has many *interpersonal* vicious epistemic traits. The vicious traits (all else held equal) make Sid epistemically *worse*. After all, traits that tend toward negative epistemic consequences rarely make an agent epistemically better. So why circumscribe the scope of epistemic vice (or virtue) to *personal* traits alone (to only those traits that influence the epistemic states of the trait holder)? Personal *epistemic traits* (as the term will be used here) are traits that primarily influence *an agent’s own acquisition of knowledge, justified belief, wisdom, understanding, etc.*, but not the epistemic lives of others. An agent’s *interpersonal* epistemic traits, contrastingly, *primarily impact others*. Yes, some traits might be both personal and interpersonal, although most fall closer to one side of the divide than the other. This chapter delineates two *interpersonal* epistemic vices, and the overall epistemic focus is on interpersonal (rather than *personal*) epistemology.

### 4. Epistemic Insensitivity

#### 4.1 Epistemic Insensitivity: An Overview

Epistemic insensitivity, roughly speaking, amounts to the following: *a culpable lack of awareness of contingent, non-epistemic (or not directly epistemic) factors that (rightly or wrongly) interfere with the acquisition of epistemic goods (knowledge, justified true belief, wisdom, etc.).* Directly epistemic can be distinguished from indirectly epistemic via epistemic investigation (see my digression in the boxed text).

## Digression on Direct and Indirect Epistemic Factors

Epistemic investigation is an attempt by an agent, or multiple agents, to acquire epistemic goods (true belief, wisdom, knowledge, etc.) If  $x$  is a *directly* epistemic factor in relation to epistemic endeavor  $y$ , then  $x$  (all by itself) provides reason to become more or less confident in  $y$ -relevant propositions. Said differently, directly epistemic factors, all on their own, speak to the truth or falsity of propositions. Indirectly epistemic factors, on the other hand, are relevant because they speak to *the likelihoods of a successful investigation*. Learning indirect epistemic information can either increase or decrease an agent's confidence that a given investigation will prove successful; these factors do not speak to truth of  $p$  or  $p$ -relevant propositions, but rather to the likelihood that agents will acquire true beliefs about  $p$ .

Both indirect and direct epistemic factors facilitate epistemic investigations. Directly epistemic factors can help an agent get to the bottom of an investigation, but indirectly epistemic factors can let an agent know whether an investigation is worth undertaking at all.

In a world of fully rational agents, i.e., agents immune to irrelevant epistemic forces, epistemic insensitivity would be of little, if any, worry. In this epistemic utopia, non-epistemic factors never threaten interference. Yet in the actual world, agential beliefs are frequently influenced by epistemically irrelevant forces that bear no relevance to the truth or falsity of an investigative belief. Moreover, even agents aware of t epistemic “bad habits” can fall victim. Irrelevant yet powerful epistemic influence includes:

- *Confirmation bias*: We have a tendency to believe propositions that confirm our world view or our expectations, regardless of the evidence.<sup>7</sup>
- *Distractibility*: Epistemic activities that we find dull, epistemic activities that take place while we are tired, hungry, cold, thirsty, emotionally exhausted, heart-broken, etc., are more likely to result in epistemic failure.<sup>8</sup>
- *Personal biases*: Testimony spoken by those we like and respect increases our “willingness” to believe, while testimony from those we dislike decreases this willingness.<sup>9</sup>

<sup>7</sup> See Allahverdyan & Galstyan, 2014; Mercier, 2017; Oswald and Grosjean, 2004; and Silverman, 1992.

<sup>8</sup> See Kpolovici et al., 2014; Renninger et al., 2015; Schiefele, 1999; Silvia, 2008; and Tobias, 1994.

<sup>9</sup> See Pohl, 2017 for articles discussing all kinds of bias, including protectiveness, vindictiveness, and personal biases of various sorts. See also Anderson et al., 2011; Banaji & Greenwald, 2016; Chambers et al., 2013; Weeks, 2015; Wilkowski & Robinson, 2010; and Wu et al., 2013.

- *Vindictiveness*: We want to believe that those we dislike, (1) believe differently than we do, and (2) hold false beliefs.
- *Protectiveness*: Perhaps a sub-category of confirmation bias, we are hesitant to believe propositions that challenge our values.
- *Defensiveness*: Information that is relayed to us with rudeness or disrespect is less likely to be believed, regardless of the evidence.

Wise epistemic discourse requires an awareness of, and sensitivity to, the epistemic shortcomings of interlocutors, i.e., sensitive agents *recognize*, *appreciate*, and *account for* the epistemic flaws of interlocutors. Sensitivity should matter to those who care about positively influencing the epistemic lives of others, because without it, the odds of this influence diminishes. Consider the following example:

**NO EXTENDED DEADLINES:** Professor A has a strict “no extended deadline” policy. Student B (who has thus far been an excellent student) asks Professor A for an extended deadline due to a highly unusual and distressing personal circumstance. Professor A sticks to the original policy, and refuses to offer an extended deadline. Student B becomes distraught and overwhelmed, giving up on the assignment all together. Moreover, Student B’s intellectual enthusiasm takes a turn for the worse, and the student’s performance for the remainder of the semester falls far short of the excellent early semester showing.

We can argue that Professor A’s pedagogical choice is epistemically insensitive. The contingent epistemic features at play are Student B’s circumstances. These are not directly epistemic features, for they do not, by themselves, speak to the truth or falsity of any investigative proposition. However, these factors are relevant to the likelihood that an agent (Student B) will acquire various epistemic goods. Student B’s circumstances make their epistemic success less likely (all things considered equal).

An epistemically sensitive professor would recognize as much, and work to mitigate the negative impact of Student B’s circumstances. In failing to respond this way, Professor A might pass by an opportunity for meaningful epistemic exchange that could improve his pedagogy. If Professor A had shown more sensitivity and extended the deadline, the student might have had epistemic uptake not only regarding the particular assignment, but also the remainder of the semester.

Admittedly, only from a God’s-eye view can we be certain that *epistemic insensitivity* bears the responsibility (partly or fully) for Student B’s epistemic spiral. However, virtuous agents make the best choice with the available information. Student B’s testimony was information that the student might struggle without accommodation. Even if special accommodations *had not* improved Student B’s epistemic state, offering them would still be sensitive, and refusing

to do so would still be insensitive. Sensitivity and its opposite are not virtues that demand achieving certain results, but rather, respond to the circumstances in a way *likely* to achieve certain results.

## 4.2 Epistemic Insensitivity and Defensiveness

Distrust of elites, especially conservative distrust, seems to have long been worsening since the election of Donald Trump, and might be directed at the media (especially the “mainstream media”), expert testimony, scientific publications, physicians, other types of experts, and any politician who fails to support Trump.<sup>10</sup> As bad as this distrust can be all on its own, epistemic insensitivity might exacerbate problems in both degree and scope. This example illustrates how “elite distrust” might develop:

**VALUE CONFLICT:** Samuel (an expert) tries to convince Sunny (a non-expert) about the truth of evolution. Samuel knows Sunny is religious. Nonetheless, Samuel explains evolution while simultaneously empathizing the truth of atheism.

We all have instinctively protected values that shape our world view.<sup>11</sup> If Samuel values his Mercedes, then he will protect it from damage or anything threatening its immaculate condition. Not all values, however, are material. Suppose academics (e.g., experts, liberal elites) value equality and fairness. We should then not be surprised when elites often act protectively when fairness and equality are threatened, and when they are hesitant to believe theories, propositions, or testimonies that conflict with equality and fairness. From the other perspective, when expert testimony conflicts with conservative values (as in VALUE CONFLICT), protective responses from conservatives are unsurprising. VALUE CONFLICT was a hypothetical case, but not far from reality: Elites *do* have a reputation (amongst certain crowds) for being “anti-religious.”<sup>12</sup> Moreover, statistics confirm that academics

<sup>10</sup> Articles that speak to this general sense of distrust include Cage, 2017; Carney, 2016; Freidman, 2017; Kaina, 2008; and Turnage, 2017.

<sup>11</sup> Values can and do change, but typically that change is often slow and moderate, for those who even change at all. Baron and Spranca, 1997, show that people will protect their values by trading off other things of great human importance, like economic security, for instance. Persons do this both consciously and implicitly. See also Ritov and Baron, 1999, and Tetlock, 2003.

<sup>12</sup> The most recent survey on the religious values of professors demonstrated that 38 percent were atheist or agnostic (Dunn, 2016). In the general population, this figure is 7 percent (Lipka and Gecewicz, 2017), which, of course, includes liberals and democrats. Amongst elite universities, almost 50 percent were atheist or agnostic. Moreover, since the survey of the professors is older, and as the population has gotten less religious over time, the gap might be even wider today. There is no research on the patriotism of professors or other specific groups within the elites; however, see Bump, 2019 for information on the wide gap between the patriotic beliefs of those who lean conservative vs. those who lean liberal. Many of the value differences, of course, are connected to political values. Conservatives are not only justified in believing, but know, that academics have much more liberal values than they do,

and the highly educated more generally also eschew religion at higher rates. The difference does not end there. Academics' values and "everyone else's" values are often miles apart on many other issues (many of them emotionally charged), including patriotism, freedom of choice, and communal sacrifice.<sup>13</sup>

Because "expert" values diverge from "non-expert" values, the following seems plausible: When experts assert their values, non-experts can interpret these assertions as attacks on their own values, and hence, respond defensively.

Because cognitive powers are limited, focus on one thing detracts from focus on other things, e.g., focus on value defense will shift focus away from the original discussion. This "value defense mode" impaired focus also impairs epistemic success generally. If Sunny is defending her commitment to theism, her attention has shifted away from Sam's evolutionary argument, and she is less likely to acquire true beliefs about evolutionary theory. Likewise, when Trump supporters are focused on defending his presidency, their focus on legislative policy wanes. Policy conversations between identified conservatives and an identified progressive might epistemically crumble when the former shifts to defending Trump from perceived attacks. If this happens, there are decent odds the progressive acted insensitively, which brought on the epistemic shift. Epistemic insensitivity also is manifest when an agent *fails to notice* these focus shifts. An epistemically sensitive agent, contrastingly, will notice that their interlocutor is defending values instead of discussing the issue at hand. In recognizing this, the sensitive agent redirects the conversation back to what epistemically matters.

## 5. Are Elites Really Insensitive?

Suppose that experts/elites, for example, *sometimes* manifest insensitivity, but comparatively less so than other groups. If so, why worry much about elite insensitivity? Because what matters is not whether the elites are *more* epistemically insensitive, but the potential for negative impact.

Suppose a group G manifests vice V at a *lower* rate than most other groups. However, given G's social position and influence, V as a feature of G has greater epistemic import than V as a feature of other groups. Suppose we learn that the population at large manifests the epistemic vice of gullibility at a rate of 0.75, but the gullibility rate amongst journalists is 0.55. Notwithstanding, it might still make sense to be more worried about gullibility amongst journalists than among the general populace due to journalists' social privileges and powers. Like gullible

and that the more elite the college, the further is this political value divide; see Abrams, 2016; Langbert, 2018; and Langbert et al., 2016. While the extent of the liberal to conservative professor ratio varies according to discipline and university location, even the most conservative heavy disciplines and locations still had conservatives out-numbered, and many times over.

<sup>13</sup> See Harris, 2018; Pew Research Center, 2016; and Schoon et al., 2010.

journalists, epistemically insensitive experts might be cause for special epistemic concern, even if experts manifest insensitivity at a comparably lower rate than the rest of the populace. Epistemic insensitivity, when manifest in elites, might create special epistemic problems or have special potential to worsen existing ones; it might worsen political polarization between conservative identified persons (or persons with conservative sympathies) and those with clashing views.

Increasing political divide typically means decreasing open-mindedness, i.e., decreasing ability to consider alternative perspectives, open-mindedness that allows agents to move from a false belief to a true one. For instance, many conservative identified persons also identify as Christian. And Christians, if they are being consistent, have lots of reasons to feel uneasy about supporting Donald Trump.<sup>14</sup> In the midst of their doubts, suppose a conservative Christian finds themselves in conversation with an expert. What might start out as a productive exchange can quickly go awry because of insensitivity. Agents listening to alternative perspectives one moment might be on the defensive the next, fearing that their values are being challenged or disrespected. Previously open-minded agents can then hold tighter to their previous beliefs, seeking out agents who share their values instead of challenging them, i.e., they might seek out partisan social media groups and only consume partisan news sources.

Even if partisanship and polarization are not made worse, epistemic insensitivity can still interfere with opportunities for valuable epistemic exchange. While experts don't know more about everything in life, on average, those with more education, and/or those who have intellectually oriented careers, are arguably well-suited to educate their interlocutors. Intellectual experts, especially university instructors, are in a privileged position when it comes to teaching "all-purpose intellectual skills," e.g., critical thinking and reasoning, skills in identifying reliable sources, the ability to recognize argumentative fallacies like strawmen, false dilemmas, and so on.

Admittedly, non-experts are better equipped to teach than experts in many "specific skill set" areas of life. A mechanic is in a better position (compared to non-mechanics) to teach the skill set associated with diagnosing a smoking engine; a pre-school teacher is in a superior pedagogical position (compared to a non-pre-school teacher) regarding, "strategies for motivating tired and hungry toddlers."

It is not as though experts/the highly educated/elites are better at everything; it is only that their particular skills place them in a privileged position in respect to certain kinds of epistemic teaching. In this respect, experts who are epistemically insensitive can come with higher epistemic costs compared to the same vice in non-experts. First, this vice tends to manifest in a particular kind of epistemic

<sup>14</sup> See Wehner, 2020.

exchange that is common amongst experts. The mere fact that experts are more likely to engage in these conversations makes the vice especially problematic for experts. Suppose a coward rarely, if ever, faces threatening, fearful, or dangerous situations; in this case, cowardliness might cause little trouble. Cowardliness, however, possessed by a deployed military infantryman, might cause all the trouble in the world. Epistemic insensitivity in experts is more like the latter, and less like the former. Imagine an epistemic exchange between two persons, neither experts. If epistemic insensitivity interferes, there might be little lost at all (because neither party was in an epistemic position to teach the other.) However, if one party is an expert, or if both are experts in different fields, agents might have been well-suited positioned for fruitful epistemic exchange. Hence, epistemic insensitivity in experts might have a particularly high cost.

## 6. Epistemic Obstruction

### 6.1 A Spectrum: Superficiality, Obstruction, Clarity

Epistemic obstruction falls between the virtue of epistemic clarity and the vice of epistemic superficiality. Superficial agents are so concerned with avoiding confusion that they keep discussion superficial, asserting only what has long been understood. Because of this, rarely are epistemic ends advanced, but instead, remain stagnant. Epistemic obstructionists, on the other hand, might advance certain parts of an argument quickly through the use of technical language that latches onto important nuances and distinctions. However, advancements might be made so quickly that it creates more confusion than clarity. While the technical terms might have originally allowed experts to point out scholarly specificities important for advancing a debate, the technicalities are soon misrepresented on a grand scale, putting a debate further back than where it started. Epistemic obstruction stymies advancement by plowing ahead rather than anticipating how, and in what way, scholarship might be interpreted, distributed, and (mis)understood.

The virtue of epistemic clarity hits the virtuous mean between superficiality and obstruction. It does not demand that elites give up technical language; if technicalities have greater potential for epistemic good than bad, then using technicalities is not only compatible with virtue but required. In these cases, using technical language is virtuous and manifests epistemic clarity.

Epistemic clarity means taking into account what we might call, “the greater epistemic good” or, “the best epistemic ends, all things considered.” Epistemically clear agents, (1) are aware of technical language even when it might become second nature to write and speak technically; (2) are aware of, and take into account, the reality that many are unfamiliar with technicalities; (3) make an

attempt to use less technical language while maintaining an expert level of research/scholarship; and (4) when technical language is unavoidable, are prone to explain and clarify technicalities (as a means of preventing misunderstandings). Consider this hypothetical example:

#### RACISM

*Scene: A professor writing an op-ed in a major newspaper*

Quote from the op-ed: “Over 75% of recent hires in large technology corporations were white males. Yet 50% of the US population are females, and almost 40% are non-whites. Hence, it could not be clearer US technology companies are racist.

For the sake of argument, suppose that the above claims are true, i.e., that the hiring demographics of US technology companies are as reported, and moreover, that these demographics can be explained by racism. Hence, any epistemic problems in the quoted text are not “truth problems.” Nonetheless, epistemic trouble can arise *in the way the truths are expressed*. Formally, we can define epistemic obstruction as follows:

**EPISTEMIC OBSTRUCTIONIST:** An agent who (1) is disposed toward technical expression styles that lack clarity and cause confusion without comparable epistemic benefit, and (2) has a corresponding disposition to forget about, or refrain from, explaining and illuminating technicalities.

It is important to note that epistemic obstructionists need not *intend* to cause confusion. Indeed, intention to do this might be better called epistemic malevolence. The fault of epistemic obstruction lay not in intention, but rather in *the failure of attention*; the epistemic obstructionist manifests a special form of negligence; the vice tends to be far less deliberate than epistemic insensitivity. The authors of RACISM might have no intention to mislead their audience; they might not have any idea their quote could so easily be misinterpreted.

I (the author of this chapter) have a non-academic passion, and most others involved in this passion are not academics. We can just call them my “non-academic friends.” I can imagine these non-academic friends reacting to RACISM with either (1) confusion, (2) eye rolling, and/or (3) exasperation at its “ridiculousness.” While the conservative identified might be more likely to have these reactions, I would be unsurprised if others reacted similarly. An explanatory theory will be suggested. But first, let us envision how the liberal elites themselves might react to the just described reactions.

While I do spend significant time with non-academics, I also spend a lot of time with academics. I would not be surprised if, after learning of the critical reaction, many experts scoffed contemptuously. They might assume criticism of RACISM

was itself grounded in racism. This might come alongside the suspicion that the critics are from southern, rural areas of the country. Consider a recent tweet from a philosophy graduate student:

I unironically embrace the bashing of rural Americans... They, as a group, are bad people who have made bad life decisions... and we should shame people who aren't pro-city.<sup>15</sup>

In the above, the tweeter makes unfavorable moral assumptions based purely on residence. Admittedly, this kind of expert criticism of non-experts might latch on to legitimate epistemic shortcomings. Regardless, these faults are compatible with experts manifesting vice through their own communication mistakes. For example, the tweet above manifests *epistemic insensitivity*, and RACISM manifests epistemic obstruction. The quote nicely distinguishes the difference between the two vices. Epistemic insensitivity is not about the use of a specific kind of language, but it is putting certain individuals, or the values of certain individuals, in an unfavorable light. The type of critical scene setting affects the mood, and/or disposition of an interlocutor or other agents in epistemically negative ways. Epistemic obstruction, however, rarely has much to do with personal judgement or values. Instead, it concerns the way language can mislead agents even if they are in ideal epistemic states.

## 6.2 Epistemic Obstruction and Talking Past Each Other

One particularly insidious feature of epistemic obstruction is how often it results in the phenomenon known as “talking past each other.” For example, suppose that several expert researchers are informed about strong scholarly disagreement coming from non-experts. From the experts’ view, the disagreement is focused on claims that seem especially reasonable and modest. Can this be explained by

<sup>15</sup> Parke, 2019. Unsurprisingly, the source I site is Fox News, although it has been verified elsewhere and was widely discussed at the time. The fact, however, that this source is a conservative one speaks to the phenomenon I describe. Many conservatives, justifiably or not, believe they are unfairly judged by members of the liberal elite. The tweeter (whose name I won’t mention) fits their mold: a West Coast academic residing in a city with a reputation of being “extremely liberal.” Even if the tweet is not representative, conservatives can point to it as an example and say, “See the elites make vast sweeping judgements about our character. They clearly have something against us. Why should we trust such a biased source?” This is a case where the tweeter might, just a few seconds ago, have been making epistemically meaningful points that perhaps were under consideration by some persons who identify or lean conservative. But once that tweet is said, the conservatives lose epistemic focus and their focus shifts to one of defense of values. Perhaps it isn’t even fair to attribute the shift entirely to the reaction, because the comment itself was already a shift. The comment was a moral and value judgement that was tangential to the epistemic center of the conversation. To be clear, the issue is not whether elites manifest this vice frequently, or more often than others. The issue is whether experts manifesting the vice have special epistemic detriment.

the fact that non-experts lack expertise? Maybe. However, another possibility is that there is no disagreement at all, but rather differing scholarly groups talking past each other.

“Talking past each other” stories can sometimes be explained by the vice of epistemic obstruction. A disagreeing party might have heard the experts claim  $p$  (an unreasonable claim), when they were actually claiming  $p^*$ . Epistemic obstruction might be responsible for this misinterpretation. The expert might have phrased  $p^*$  in such a way that many reasonable agents interpret it as  $p$ . Feigned disagreement then arises out of thin air. There is no actual disagreement at all. The obstructionist and expert said  $p$ , and meant  $p$ . The non-expert heard the scholar say  $p^*$ .

Let us return to RACISM. Most academics are familiar with “institutional” and “structural” racism, i.e., racism that demands neither vicious nor conscious intention, but instead, is displayed via the organizational structure of private and public institutions, social norms, and other non-intentional systems, schemes, or habits. This might be seen not only in organizations and institutions themselves, but also in behavior patterns and in unreflective attitudes that negatively and disproportionately affect minorities.<sup>16</sup> Structural racism differs from the definition of racism commonly used by non-experts engaging in ordinary discussion (especially ordinary persons who are *not* minorities.) Factors that might contribute to unfamiliarity with structural racism include: minimal formal education, age, growing up in politically “conservative” communities, etc.<sup>17</sup>

Many assume that racism requires explicitly “racist” attitudes toward minority racial groups and/or the intent to harm a member of a minority racial group.<sup>18</sup> This type of “racist” attitude is understood as an attitude of superiority or contempt directed toward minority races for no reason other than race. An

<sup>16</sup> See Altman, 2015; Arthur, 2007; Fullinwider, 2018; James, 2016; and Zack, 2017.

<sup>17</sup> This is not a claim for empirical knowledge. It is just my opinion, based on experiences interacting with, dating, and becoming friends with many persons who lack high school or college degrees, live in economically disadvantaged neighborhoods, and hold blue-collar careers. I could be wrong. But two points: (1) the chapter is not contingent on this particular example. My example might be a bad one, and yet the phenomenon I describe is still real and important; (2) Given the lives that many persons lead, overwhelmed by work, paying the bills, and trying to find enough free time to spend with family, it seems a safe bet that many persons in these types of situations might have missed special ways in which the educated use words. Moreover, those who have taught intro classes at non-elite universities are well aware just how easy it is for students to misunderstand an argument, point, or concept, and how many times a concept (or something else) must be repeated for learning to take place; it seems wise that elites err on the side of caution, i.e., err on the side of defining their terms, and making sure that everyone starts at a similar understanding. After all, philosophers do this in their scholarly publications quite frequently. The introduction of many scholarly philosophy papers is spent overviewing important terms, concepts, and ideas, even though the aforementioned should be fairly obvious to those familiar with the literature. It seems worth it to avoid confusion. Even more so, then, when discussing important ethical and political issues with the public, the same standard should (epistemically) apply.

<sup>18</sup> Campbell, 2018 describes the talking past each other phenomenon and the contested use of “racism” nicely.

alternative “everyday” perspective is that racism is treating persons differently from reason other than skin color. When race-theory experts use the term “racism” differently than the masses, non-elites can quickly fall into epistemic confusion, and soon after, strident disagreement. This seems both understandable and predictable. After all, when (some) non-elites hear statements about (structural) racism, these are interpreted as statements about racist intention, i.e., RACISM might be read as “technology companies are purposely turning down qualified minorities in favor of white persons with less qualifications.” Angry, negative, and intense emotional reactions to “elite” statements about racism might really be angry, negative, and intense emotional reactions to something else entirely, i.e., reactions to a proposition that was never asserted, but which the non-elite could not help but hear.

### 6.3 Feigned Disagreement and Epistemic Excommunication

*Medium* recently published “My Semester with Snowflakes,” by undergraduate Yale freshman, James Hatch. Hatch is a special forces veteran, and 52 years of age. In the excerpt below, this atypical undergraduate describes an exchange that illustrates the way epistemic obstruction can lead to extreme misunderstandings.

Before delving into the quote, consider this background information. Hatch had become friendly with many of his classmates including an outspoken, politically progressive, black female; let us call her, “Elisa.” The two students had vastly different world views and life experiences, yet Hatch admired Elisa’s intelligence, courage, and work ethic. Both students, along with two others, took part in a university, humanities-focused interview. Here is Hatch (Elisa is referred to as “the young woman”):

SAFE SPACES: As the younger students started to express their thoughts, the young woman...used the word “safe space” and it hit me forcefully. I come from a place where when I hear that term, I roll my eyes into the back of my vacant skull and laugh from the bottom of my potbelly. This time, I was literally in shock. It hit me that what I thought a “safe space” meant, was not accurate. This young woman, the one who used the phrase, “Safe Space” isn’t scared of anything. She is a life-force of goodness and strength. She doesn’t need anyone to provide a comfortable environment for her. What she meant by “safe space” was that she was happy to be in an environment where difficult subjects can be discussed openly, without the risk of disrespect or harsh judgement.

(Hatch, 2019)

The potential epistemic obstruction of relevance comes via the term “safe-space.” Let us again emphasize that epistemic obstruction often involves no ill intention.

Notwithstanding, upon hearing “safe spaces” non-elites might lack even an elementary understanding of the concept—after all, they might lack both formal education that explains it, and also a history including the right type of linguistic experience. Within ivory towers, “safe spaces” is used so often that it can be easy to forget that it is *technical*. Most persons lack any reference to guide their interpretation. Moreover, “safe space,” unlike “racism,” is rarely expressed in non-academic contexts. The words are, however, expressed separately; non-expert interpretations of “safe spaces” might derive from this “common sense” or “ordinary language” interpretation of its individual components. As with most technical terms, interpreting “safe spaces” through the lens of ordinary language might result in more mis-understanding than understanding.

Hatch’s surprised response to Elisa’s use of “safe spaces” implies that he had understood the term as referencing environments intellectually comfortable, familiar, ones that scoff at intellectual challenges. Hatch is unlikely the only one inclined to roll their eyes at this type of “safe space.” Many professors might also roll their eyes at the following “university prohibition”:

We, Public University, hereby prohibit, (1) intellectually challenging students, (2) uncomfortable course assignments, and/or (3) classroom disagreement.

This (fictional) university legislation sounds grossly unrealistic to those of working in contemporary scholarship. However, the general public has no easy way of gauging its plausibility. When the public hears (via radio, newspaper, blogs, etc.) that universities are promoting “safe spaces,” they might imagine policies in line with the above. Hatch, after all, sounds as though he had been extremely confident in his previous misguided interpretation of the term. Nonetheless, upon learning his mistake, he quickly shows *sympathy toward the proper definition*. This demonstrates how the distinction between *the perception of a technical concept* on the one hand, and *the concept itself* on the other hand, can be especially stark.

Epistemic obstruction impacts not only belief, but the vice can also shape non-doxastic epistemic attitudes, i.e., trust and distrust, disdain and admiration. Hatch said he would roll his eyes and laugh upon hearing the need for “safe-spaces.” Eye-rolling and laughing can signal an epistemic disdain and despair combination, i.e., a point at which a member of one social group believes epistemic engagement with another social group is beyond hope. Rolling one’s eyes at the use of “safe spaces” illustrates this type of epistemic hopelessness coming from a conservative identified person and directed toward those labeled part of the class of liberal elites. Below, contrastingly, is an example of epistemic hopelessness (a friend’s actual Facebook post) coming from the other side:

ELITE DESPAIR: Sometimes a Trump supporter and a Trump critic will actually talk to each other and hear reasons for why the other side holds the position that

they do. They claim that a lot is learned in these exchanges. As for me, I just defriend Trump supporters.

The post received many “like,” “love,” and “laugh” reactions. Comments on the post ranged from the humorous dismissal of Trump supporters, to self-righteous and indignant dismissals. This post and the reactions express epistemic hopelessness, disdain, despair, and epistemic dismissiveness. Interestingly, I believe Hatch describes very similar attitudes when he admits to rolling his eyes and laughing; it suggests he had no intention of engaging in epistemic exchange. Respectful epistemic exchange rarely starts out with eye-rolling.

The troublesome point made by both Hatch and the Facebook poster runs as follows: “Engagement with *those people* is not worth the epistemic effort. It is better to dismiss them, i.e., that they be excommunicated from the epistemic community.” Surely, at times, epistemic excommunication is epistemically warranted. However, epistemic obstructionism encourages it needlessly. Excommunicated agents might be much like Hatch: Persons who do not actually disagree with the excommunicating agents, even though the fundamental basis of excommunication is this supposed (but unreal) disagreement. Even more, excommunicated members might be sympathetic to the very proposition(s) with which they supposedly disagree (the disagreement responsible for their pariah status).

## 7. Objections

Some might argue that experts are under neither moral nor epistemic pressure to explain and clarify their scholarship to non-experts, nor must they guard against misunderstanding, nor must they be “sensitive” to the irrational reactions of non-experts for altruistic epistemic ends. Grounds for these claims include:

- (1) Technical language is epistemically valuable, while “dumbing-down” scholarship is not. Dumbing down scholarship causes imprecision, and hence, epistemic harm rather than epistemic good.
- (2) If ordinary persons are going to take part in political, ethical, or scholarly discussions, then it is their moral and epistemic responsibility to learn the language.
- (3) Experts use technical terms in a technical context, i.e., in contexts in which technical language is, indeed, the norm. Hence, experts are not using language out of context, and therefore have no duty to use a different language, or to explain technicalities.

## 7. 1 Accessibility Efforts are Epistemically Valuable

Objection (1) involves both empirical and normative assumptions. That said, evidence supporting the relevant empirical assumptions might be unobtainable. For instance, *if elites modified technical terms, would this really improve the epistemic status of non-elites?* (If so, this would suggest that dumbing down language does offer at least some epistemic value.) *Second, would modifying technical terms create scholarly confusion and imprecision?* Alas, we lack data to confirm or deny both hypothetical queries; any such data would be extremely difficult to collect systematically. Not only would we need agreement on what constitutes “epistemically valuable,” but we would also need a means of (i) weighing epistemic harm against good, and (ii) predicating just how much epistemic harm and good result from the use (or disuse) of technical language. Moreover, we would need some means of assessing what qualifies as “creating confusion,” and even more, the ability to recognize when “dumbing things down” caused it. Systematic data collection of this type seems implausible. But even if plausible, *at the moment*, we still lack the data.

Because it is so difficult to obtain scientifically sound assessments concerning (a) technical language and scholarship advancement, and (b) technical language and public misunderstanding, we must rely on other epistemic support. With that in mind, consider the following argument which, admittedly, is both theoretical and anecdotal; this argument far from settles the issue, but it offers *some support* for the following claim: When experts make a conscious effort to discuss scholarship in accessible terms, this effort has epistemic returns, especially for non-experts. Hence, the efforts put toward modifying language have better epistemic results (better for the non-expert) than not making this effort at all. With just this, we can get the further conclusion that ideal epistemic agents at least *consider* accessibility efforts.

Having taught at several “non-elite” universities, I’ve noticed that the way I phrase lectures makes a meaningful difference. While not all students (at either elite or non-elite universities) will grasp the complexity of Kant’s ethics, not even in response to carefully designed lectures and assignments (i.e., designed with an aim toward non-expert comprehension.) Notwithstanding, in using some words rather than others, and in presenting in one fashion instead of another, students seem to get closer to “epistemic enlightenment.” While complete enlightenment might remain in the distance, many students still get closer than if I had made no effort toward accessibility. If this experience is representative, then it seems plausible that something similar would play out in comparable forms of communication. For instance, if Kevin (a Kantian expert) wrote an op-ed on *The Metaphysics of Morals*, his careful consideration of non-expert accessibility offers something of epistemic value (more value compared to making no such efforts).

## 7.2 Why Experts Carry the Burden of Clarification

Let us turn to Objection (2): Is it the responsibility of non-experts to learn expert language? Answering, “yes,” paints an especially elitist picture, one that puts an extreme epistemic burden on non-experts. Non-experts, after all, “do not know what they do not know.” In typical cases, non-experts will assume academics are using words ordinarily. And why wouldn’t they? Without signals of divergence, it is reasonable to begin with an ordinary language assumption. The alternative is to always verify, i.e., to constantly and consistently confirm that words are not being used in typical, rather than technical, styles.

Is scholarship a different epistemic category, and therefore, a category demanding that we begin with an assumption of technical language? This, however, is unreasonable; even highly technical papers use most words in ordinary fashion. It would be bizarre to assume that there is always a divergence between words used in scholarship and words used in ordinary discourse. With this assumption, reading academic articles would become so time consuming as to be impossible; translating every word of a scholarly essay (under the assumption that any word might be technical) sounds similar to learning a new language. Few would bother reading scholarship if each paper demanded acquiring fluency in a foreign tongue. Asking this of non-experts is simply unreasonable.

## 7.3 Why Experts Cannot Circumscribe the Epistemic Scope

And finally, Objection (3), which might run as follows: Even if the technical language is brought away from the scholarly sphere and into the public one, this is not the fault or responsibility of experts. At the least, it is not their fault or responsibility if exposure happened without the recommendation or consent of the scholar.

It seems fair to concede that scholars who speak directly to the public carry a *greater* responsibility to clarify their work. Nevertheless, whenever scholarship reaches the outside world, epistemic consequences are prone to occur, whether intended or not. Virtuous agents have a reasonable awareness of how their scholarship might be consumed and distributed. Consent to this distribution is a different issue. Knowing that public distribution might happen is enough reason to make effort toward clarification. At the least, epistemically virtuous experts who use technicalities should only do so if the risk of epistemic harm from technicalities is outweighed by the potential for epistemic good. When technical language threatens misunderstanding and offers only a minor scholarly benefit, the cost/benefit calculation will often speak in favor of jettisoning the technicalities. Scholarship does not happen in a bubble, even if we wish it did.

## 8. Conclusion

This chapter has discussed epistemic insensitivity and epistemic obstruction. Both vices threaten to worsen the political and social divide between persons self-identified as conservatives on the one hand, and those that the self-identified conservatives identify as “elites,” “liberal elites,” “experts,” “progressives,” or “the left” on the other hand. In worsening this political divide, persons become more likely to distrust news sources that might, in one way or another, be associated with the “other” political perspective. In an extreme form, partisans become more and more suspect of any news source which isn’t clearly in their political bubble; they become likely to see all but the most partisan sources as “fake news,” or some other form of untrustworthy information. At the same time, agents can become *more willing* to trust extreme partisan sources which are the ones that actually convey fake news (i.e., that falsely or misleadingly report world affairs).

Epistemic insensitivity and epistemic obstruction exacerbate troubles that already occur when the politically and socially alienated can turn to highly partisan news outlets. Non-experts can not only acquire deep distrust for news outlets coming from the wrong (untrustworthy) political perspective, but might also develop a more general distrust of expert testimony (because expertise itself becomes untrustworthy, as expertise becomes closely associated with the “wrong” political leaning). Distrust can extend to any source that relies on academic scholarship (as academia itself is seen as under the guiding control of the untrustworthy).

Epistemic insensitivity and epistemic obstruction are vices that tend to have high opportunity costs, resulting in missed chances to share valuable information (that can eventually result in knowledge and other epistemic goods). There are also missed opportunities to share epistemic skills; i.e., in certain cases, if only it weren’t for epistemic insensitivity, an expert might have tutored a non-expert, helping them acquire new epistemic skills. However, victims of epistemic insensitivity are often far too focused on defending themselves to pay the requisite attention demanded of skill acquisition.

This chapter ends with a few clarifications. This chapter does *not* claim that all experts, *or even most* experts, possess the mentioned vices. Nor does it claim that experts possess or display these vices to a greater degree than other groups. The claim is just this: Epistemic obstruction and epistemic insensitivity, when manifest in experts, has particularly worrisome consequences, especially consequences related to political division, miscommunication, and distrust. When distrust of experts is already increasing, and when fake news is becoming a prominent source of concern, we are epistemically wise to pay special attention to any and all vices which exacerbate expertise skepticism and partisan divide.

## References

Abrams, S. (2016). There are conservative professors. just not in these states. *New York Times*. Retrieved from <https://www.nytimes.com/2016/07/03/opinion/sunday/there-are-conservative-professors-just-not-in-these-states.html>.

Allahverdyan, A. E. & Galstyan, A. (2014). Opinion Dynamics with Confirmation Bias. *PLoS ONE*, 9(7). doi:10.1371/journal.pone.0099557.

Altman, A. (2015, August 30). Discrimination. Retrieved January 5, 2020, from <https://plato.stanford.edu/entries/discrimination/>.

Anderson, B. A., Laurent, P., & Yantis, S. (2011). Value-Driven Attentional Capture Is Modulated by Spatial Context. *PNAS*, 108(25), 10367–71. doi:10.1073/pnas.1052894108

Arthur, M. M. (2007). Racism, Structural and Institutional. *The Blackwell Encyclopedia of Sociology*. doi:10.1002/9781405165518.wbeosr013.

Banaji, M. R. & Greenwald, A. G. (2016). *Blindspot: Hidden Biases of Good People*. New York, NY: Bantam.

Baron, J. & Spranca, M. (1997). Protected Values. *Organizational Behavior and Human Decision Processes*, 70(1), 1–16. doi:10.1006/obhd.1997.2690.

Berkowitz, P. (2018, November 6). Opinion | conservatism and populism go back centuries. Retrieved January 4, 2020, from <https://www.wsj.com/articles/conservatism-and-populism-go-back-centuries-1541548438>.

Bruinius, H. (2018, August 27). Who made you an expert? Is America's distrust of "elites" becoming more toxic? Retrieved January 4, 2020, from <https://www.csmonitor.com/USA/Politics/2018/0827/Who-made-you-an-expert-Is-America-s-distrust-of-elites-becoming-more-toxic>.

Bump, P. (2019). How views of patriotism vary by party. *The Washington Post*. Retrieved July 3, 2019, from <https://www.washingtonpost.com/politics/2019/07/03/how-views-patriotism-vary-by-party/>.

Campbell, B. (2018, August 29). The two confusing definitions of racism. Retrieved May 31, 2020, from <https://medium.com/handwaving-freakoutery/the-two-confusing-definitions-of-racism-2d685d3af845>.

Carey, T. (2016). Blame the elites for the trump phenomenon. Retrieved from <https://www.intelligencesquaredus.org/debates/blame-elites-trump-phenomenon>.

Chambers, J. R., Schlenker, B. R., & Collisson, B. (2013). Ideology and Prejudice. *Psychological Science*, 24(2), 140–9. doi:10.1177/095679

Crews Jr., C. (2020, February 18). We're not biased, we're liberals: How cultural leftism will slant social media regulation. Retrieved May 31, 2020, from <https://www.forbes.com/sites/waynecrews/2020/02/17/were-not-biased-were-liberals-how-cultural-leftism-will-slant-social-media-regulation/#31f1c33e503c7612447820>.

Davies, W. (2018, November 29). Why we stopped trusting elites. Retrieved January 4, 2020, from <https://www.theguardian.com/news/2018/nov/29/why-we-stopped-trusting-elites-the-new-populism>.

Dunn, K. (2016, December 22). Faculty faith. Retrieved January 4, 2020, from <https://harvardmagazine.com/2007/07/faculty-faith.html>.

Eatwell, R. & Goodwin, M. (2018). *Who Made You An Expert? Is America's Distrust of "Elites" Becoming More Toxic?* London: Pelican Books.

Freeze, M., Baumgartner, M., Bruno, P., Gunderson, J. R., Olin, J., Ross, M. Q., & Szafran, J. (2020). Fake Claims of Fake News: Political Misinformation, Warnings, and the Tainted Truth Effect. *Political Behavior*. doi:10.1007/s11109-020-09597-3.

Friedman, U. (2017, January 20). Why Trump is thriving in an age of distrust. Retrieved January 4, 2020, from <https://www.theatlantic.com/international/archive/2017/01/trump-edelman-trust-crisis/513350/>.

Fullinwider, R. (2018, April 9). Affirmative action. Retrieved January 5, 2020, from <https://plato.stanford.edu/entries/affirmative-action/>.

Greven, T. (2016). *The Rise of Right-Wing Populism in Europe and the United States. A Comparative Perspective*. Washington, DC: Friedrich Ebert Foundation.

Groseclose, T. (2011). *Left Turn: How Liberal Media Bias Distorts the American Mind*. New York, NY: St. Martin's Press.

Harris, A. (2018, November 12). America is divided by education. Retrieved May 31, 2020, from <https://www.theatlantic.com/education/archive/2018/11/education-gap-explains-american-politics/575113/>.

Haskins, J. (2017, August 26). Yes, there's bias in the media: Liberal bias reaches disturbing new heights. Retrieved May 31, 2020, from <https://www.newsobserver.com/opinion/op-ed/article169576577.html>.

Hatch, J. (2019, December 27). My semester with the snowflakes. Retrieved January 5, 2020, from <https://gen.medium.com/my-semester-with-the-snowflakes-888285f0e662>.

James, M. (2016, February 17). Race. Retrieved January 5, 2020, from <https://plato.stanford.edu/entries/race/>.

Kaina, V. (2008). Declining Trust in Elites and Why We Should Worry About It – With Empirical Evidence from Germany. *Government and Opposition*, 43(3), 405–23. doi:10.1111/j.1477-7053.2008.00260.x.

Kearns, M. (2018, December 8). Liberal elites shouldn't ignore the religious impulse. Retrieved January 4, 2020, from <https://www.nationalreview.com/2018/11/atheists-liberal-elite-ignore-traditional-religious-belief/>.

Kpolovie, P., Joe, A., & Okoto, T. (2014). Academic Achievement Prediction: Role of Interest in Learning and Attitude Towards School. *International Journal of Humanities Social Sciences and Education*, 1(11), 73–100.

Langbert, M. (2018). Homogenous: The Political Affiliations of Elite Liberal Arts College Faculty. *Academic Questions*, 31(2), 186–97. doi:10.1007/s12129-018-9700-x.

Langbert, M., Quain, A., & Klien, D. (2016). Faculty Voter Registration in Economics, History, Journalism, Law, and Psychology. *Econ. Journal Watch*, 13(3), 422–51.

Liao, J. (2016, November 16). The hypocrisy of the liberal elite. Retrieved January 4, 2020, from [https://www.huffpost.com/entry/the-hypocrisy-of-the-liberal-elite\\_b\\_582a3be6e4b057e23e3148e3](https://www.huffpost.com/entry/the-hypocrisy-of-the-liberal-elite_b_582a3be6e4b057e23e3148e3).

Lipka, M. & Gecewicz, C. (2017, June). More Americans now say they're spiritual but not religious. Retrieved January 4, 2020, from <https://www.pewresearch.org/fact-tank/2017/09/06/more-americans-now-say-theyre-spiritual-but-not-religious/>.

Masciotta, D. (2016, November 22). "Real Americans" vs. "coastal elites": What right-wing sneers at city dwellers really mean. Retrieved December 7, 2020, from <https://www.salon.com/2016/11/20/real-americans-vs-coastal-elites-what-right-wing-sneers-at-city-dwellers-really-mean/>.

Mcgonagle, T. (2017). Fake News: False Fears or Real Concerns? *Netherlands Quarterly of Human Rights*, 35(4), 203–9. doi:10.1177/0924051917738685.

Mercier, H. (2017). Confirmation Bias—Myside Bias. In R. Pohl (Ed.), *Cognitive Illusions: Intriguing Phenomena in Thinking, Judgment and Memory* (pp. 109–24). London: Routledge, Taylor & Francis Group.

Merkley, E. (2020, March 20). Analysis | Many Americans deeply distrust experts. So will they ignore the warnings about coronavirus? Retrieved December 29, 2020, from <https://www.washingtonpost.com/politics/2020/03/19/even-with-coronavirus-some-americans-deeply-distrust-experts-will-they-take-precautions/>.

Moravec, P., Minas, R., & Dennis, A. R. (2019). Fake News On Social Media: People Believe What They Want To Believe When It Makes No Sense At All. *Management Information Systems Quarterly*, 43(4), 1343–60. doi:10.2139/ssrn.3269541.

Mudde, C. (2007). *Populist Radical Right Parties in Europe*. Cambridge: Cambridge University Press.

Neiwerk, D. (2016). Trump and Right-Wing Populism: A Long Time Coming. *The Public Eye*, 86, 3–19.

Oswald, M. & Grosjean, S. (2004). Confirmation Bias. In R. Pohl (Ed.), *Cognitive Illusions: A Handbook on Fallacies and Biases in Thinking, Judgement and Memory* (pp. 79–96). Hoven, NY: Psychology Press.

Parke, C. (2019). UC Berkeley Instructor Calls Rural Americans 'bad People' Who Deserve 'Uncomfortable' Lives. *Fox News*, 12 Nov. 2019, from <https://www.foxnews.com/us/uc-berkeley-instructor-rural-americans-bad-people>.

Pew Research Center (2016, April 26). Ideological gap widens between more, less educated adults. Retrieved May 31, 2020, from <https://www.people-press.org/2016/04/26/a-wider-ideological-gap-between-more-and-less-educated-adults/>.

Phillips-Fein, K. (2019, January 31). How the right learned to loathe higher education. Retrieved January 4, 2020, from <https://www.chronicle.com/article/How-the-Right-Learned-to/245580>.

Pohl, R. (2017). *Cognitive Illusions: Intriguing Phenomena in Thinking, Judgment and Memory*. London: Routledge, Taylor & Francis Group.

Rankin, C. (2019, September 19). How does President Trump get re-elected in 2020? With lots of help from California. Retrieved January 4, 2020, from <https://www.sacbee.com/opinion/op-ed/article235221867.html>.

Renninger, A., Hidi, S., & Krapp, A. (2015). *The Role of Interest in Learning and Development*. London: Psychology Press, Taylor & Francis Group.

Ritov, I., & Baron, J. (1999). Protected Values and Omission Bias. *Organizational Behavior and Human Decision Processes*, 79(2), 79–94.

Rydgren, J. (2005). Is Extreme Right-Wing Populism Contagious? Explaining the Emergence of a New Party Family. *European Journal of Political Research*, 44(3), 413–37. doi:10.1111/j.1475-6765.2005.00233.x.

Schiefele, U. (1999). Interest and Learning from Text. *Scientific Studies of Reading*, 3(3), 257–79. doi:10.1207/s1532799xssr0303\_4.

Schoon, I., Cheng, H., Gale, C. R., Batty, G. D., & Deary, I. J. (2010). Social Status, Cognitive Ability, and Educational Attainment as Predictors of Liberal Social Attitudes and Political Trust. *Intelligence*, 38(1), 144–50.

Silverman, B. (1992). Modeling and Critiquing the Confirmation Bias in Human Reasoning. *IEEE Transactions on Systems, Man, and Cybernetics*, 22(5), 972–82. doi:10.1109/21.179837.

Silvia, J. (2008). Interest—The Curious Emotion. *Current Directions in Psychological Science*, 17(1), 57–60. doi:10.1111/j.1467-8721.2008.00548.x.

Spiliakos, P. (2018, October 18). The Calvinball world of elite white liberals. Retrieved January 4, 2020, from <https://www.nationalreview.com/2018/10/elite-white-liberals-double-standards/>.

Stark, S. (1996, February). Right-Wing Populist: Pat Buchanan's presidential campaign is testing a political potentiality that could have a future in downsizing America. *The Atlantic*. Retrieved November 7, 2020, from <https://www.theatlantic.com/magazine/archive/1996/02/right-wing-populist/306027/>.

Tanehaus, S. (2017, June 8). Why conservatives have always distrusted science from the Scopes trial to the “arrogant” atomic elite to climate change. Retrieved January 4, 2020, from <https://www.bloomberg.com/opinion/articles/2017-06-08/why-conservatives-have-always-distrusted-science>.

Taub, A. (2016, July 8). A central conflict of 21st-century politics: Who belongs? Retrieved January 4, 2020, from <https://www.nytimes.com/2016/07/09/world/europe/a-central-conflict-of-21st-century-politics-who-belongs.html>.

Tetlock, P. E. (2003). Thinking the Unthinkable: Sacred Values and Taboo Cognitions. *Trends in Cognitive Sciences*, 7(7), 320–4. doi:10.1016/s1364-6613(03)00135-9.

Tobias, S. (1994). Interest, Prior Knowledge, and Learning. *Review of Educational Research*, 64(1), 37–54.

Tomasky, D. (2017, May 30). Elitism is liberalism's biggest problem. Retrieved January 4, 2020, from <https://newrepublic.com/article/142372/elitism-liberalisms-biggest-problem>.

Tubbs, S. (2019, August 7). "Fake news," diminishing media trust and the role of social media. Retrieved May 31, 2020, from <https://uh.edu/news-events/stories/2019/august-2019/08072019-uh-researcher-explores-fake-news-phenomenon.php>.

Turnage, C. (2017, July 10). Most Republicans think colleges are bad for the country. Why? Retrieved January 4, 2020, from <https://www.chronicle.com/article/Most-Republicans-Think/240587>.

Weeks, B. E. (2015). Emotions, Partisanship, and Misperceptions: How Anger and Anxiety Moderate the Effect of Partisan Bias on Susceptibility to Political Misinformation. *Journal of Communication*, 65(4), 699–719. doi:10.1111/jcom.12164.

Wehner, P. (2020, January 30). There is no Christian case for Trump. Retrieved May 31, 2020, from <https://www.theatlantic.com/ideas/archive/2020/01/there-no-christian-case-trump/605785/>.

Wilkowski, B. M. & Robinson, M. D. (2010). The Anatomy of Anger: An Integrative Cognitive Model of Trait Anger and Reactive Aggression. *Journal of Personality*, 78(1), 9–38. doi:10.1111/j.1467-6494.2009.00607.x.

Wu, L., Zhang, H., Chiu, R. K., Kwan, H. K., & He, X. (2013). Hostile Attribution Bias and Negative Reciprocity Beliefs Exacerbate Incivility's Effects on Interpersonal Deviance. *Journal of Business Ethics*, 120(2), 189–99. doi:10.1007/s10551-013-1658-6.

Zack, N. (2017). *The Oxford Handbook of Philosophy and Race*. New York, NY: Oxford University Press.

# Echo Chambers, Fake News, and Social Epistemology

*Jennifer Lackey*

It's no secret that Donald Trump is a fan of Fox News. Indeed, reports<sup>1</sup> claim that he recently raged at staff members when his wife, Melania, was caught watching "fake news" CNN aboard Air Force One. Apparently, the matter was resolved when Trump's staff confirmed that, moving forward, it would be standard operating procedure to have all TVs tuned to Fox.

Word of this elicited what is now a familiar criticism of Trump—that he is carefully curating his own consumption of news so that it includes information only from likeminded supporters. In so doing, he is said to be illegitimately reinforcing his own beliefs and sheltering himself from criticism. He is, in other words, creating a dangerous *echo chamber* for himself. The problem with echo chambers is vividly captured by a well-known example from Wittgenstein in which he describes someone assuring himself of the truth of a report by reading several different copies of the same newspaper. Whether you read one or one thousand copies of today's *New York Times*, the evidence you take in about today's news is exactly the same. You can't increase the epistemic status of your views by simply consulting multiple versions of the same source.

Echo chambers of this sort are said to be responsible for a host of today's problems, including the degradation of democracy. By insulating ourselves from opposing views and disagreement, we are stunting our ability to engage in effective deliberation about the most pressing issues of our time. As Cass Sunstein says in his book, *#Republic: Divided Democracy in the Age of Social Media*, "In a well-functioning democracy, people do not live in echo chambers or information cocoons" (Sunstein 2017, p. ix.).

In this chapter, I argue that this diagnosis of what is wrong with the sort of consumption of news exemplified by Trump is fundamentally incorrect. In particular, I examine the two dominant explanations of the distinctively epistemic problem with echo chambers and show that each is wanting. Echo chambers, by themselves, are not epistemically problematic. I then highlight how echo

<sup>1</sup> See <https://www.nytimes.com/2018/07/24/us/politics/trump-putin-cnn.html> (accessed November 4, 2018).

chambers are characterized in purely structural terms, but what is needed to capture what is wrong with Trump's exposure to only Fox News is content-sensitive. In other words, it is not that Trump is relying on a single source for news, but that he is relying on one that is unreliable. This makes the problem of discriminating between fake and real news more pressing than ever. I then show that the prevalence of social media bots poses an additional set of epistemological problems for online activity, ones that non-ideal social epistemology should be well-positioned to address.

## 1. Echo Chambers

There are three central components to an echo chamber. First, there is an opinion that is repeated and reinforced, thereby amplifying it, often through re-sharing; second, this occurs in an enclosed system or "chamber," such as a social network, allowing the opinion to "echo"; and, third, dissenting voices are either absent or drowned out. According to Kathleen Hall Jamieson and Joseph N. Cappella, for instance, an echo chamber is a "bounded, enclosed . . . space that has the potential to both magnify the messages delivered within it and insulate them from rebuttal" (2008, p. 76). In a similar spirit, Jon Robson argues that the following three claims are true of echo chambers:

- (i) The accepted view of a group (and particularly its opinion leaders) is frequently repeated and reinforced while dissenting views, if they are present at all, are drowned or ignored.
- (ii) They are in some respect—be it geographical, cultural, or otherwise—enclosed spaces.
- (iii) The judgments of opinion leaders are not merely transmitted but also amplified. (Robson 2014, p. 2520)<sup>2</sup>

Returning to the case involving Trump mentioned at the outset, Fox News is well-known for promoting only conservative viewpoints. Indeed, Media Bias/Fact Check "rate Fox News strongly Right-Biased due to wording and story selection that favors the right" and as "Mixed" regarding its factual reporting "based on poor sourcing and the spreading of conspiracy theories that later must be

<sup>2</sup> C. Thi Nguyen proposes a slightly less orthodox view of echo chambers: "I use 'echo chamber' to mean *an epistemic community which creates a significant disparity in trust between members and non-members*. This disparity is created by *excluding non-members through epistemic discrediting*, while simultaneously *amplifying insider members' epistemic credentials*. Finally, echo chambers are such that *general agreement with some core set of beliefs is a pre-requisite for membership*, where those core beliefs *include beliefs that support that disparity in trust*" (2020, p. 146, original emphasis).

retracted after being widely shared.”<sup>3</sup> It is, thus, no surprise that a Pew Research Center survey found that Fox News was the main source for 40 percent of Trump voters during the 2016 election, and that 47 percent of consistent conservatives name Fox as their central source of political news.<sup>4</sup> Given all of this, the Trump Administration and Fox News do indeed seem to satisfy the conditions of being in an echo chamber. Conservative viewpoints are amplified in the network and are then picked up and repeated by Trump and his staff members. Moreover, by virtue of all of the TVs surrounding Trump and his staff being tuned to Fox, they have created an enclosed system of agreement, where dissenting voices of other sources of news, such as CNN, are entirely absent from the discussion.

Saying that someone is in an echo chamber is not a value-neutral claim. Instead, it is regarded as a criticism or a call for change, at least from an epistemic point of view. That is to say, being in an echo chamber is taken to be epistemically problematic. Let’s now turn to the two central concerns raised with being in an echo chamber and evaluate the extent to which each succeeds in (i) identifying an epistemic problem, and (ii) capturing what is wrong with Trump’s consumption of news.

The first epistemic problem is that the opinions expressed in an echo chamber *lack independence*, yet this is often not at all transparent to those in the closed system. Because of this, a false sense of the amount of epistemic support on behalf of a given opinion can easily be conveyed. This problem is often illustrated by Wittgenstein’s example, mentioned earlier, in which someone assures himself of the truth of a report by reading several different copies of the same newspaper. Each additional newspaper is simply a copy of the original report, so multiple instances of the testimony in question are all dependent on one source, bringing no additional support whatsoever. We can imagine, however, a naïve consumer of news who has no idea how newspapers are produced, believing that each report is from an independent, or at least a different, source. This would present a massively misleading epistemic picture to the consumer.

Our consumption of news is said to bear a striking similarity to Wittgenstein’s newspaper reader. For instance, 61 percent of Millennials report getting their political news on Facebook.<sup>5</sup> We often choose our Facebook friends and Twitter feed on the basis of our personal and professional relations, which increases the likelihood of being surrounded by people who share many of our core beliefs. We then log on and see a number of “shares” of the same articles arguing for political views that we already hold. This creates the illusion of widespread support when, in fact, we are simply consulting multiple copies of the same virtual newspaper.

<sup>3</sup> See <https://mediabiasfactcheck.com/fox-news/> (accessed September 10, 2019).

<sup>4</sup> See <https://www.journalism.org/2017/01/18/trump-clinton-voters-divided-in-their-main-source-for-election-news/> (accessed March 23, 2020).

<sup>5</sup> See <https://www.journalism.org/2015/06/01/facebook-top-source-for-political-news-among-millennials/> (accessed September 10, 2019).

There are, then, two dimensions to this epistemic problem with echo chambers—first, an absence of sufficient independence among the views being expressed and, second, a lack of awareness regarding this absence, or at least the extent of it. The first dimension is a familiar one from the epistemology of disagreement. It is widely accepted in that debate that for an instance of disagreement to require rational revision, it needs to be independent of other instances of disagreement that have already been taken into account. For instance, Thomas Kelly argues that “even in cases in which opinion is sharply divided among a large number of generally reliable individuals, it would be a mistake to be impressed by the sheer number of such individuals on both sides of the issue. For numbers mean little in the absence of independence” (Kelly 2010, p. 148). Similarly, Adam Elga claims that “an additional outside opinion should move one only to the extent that one counts it as independent from opinions one has already taken into account” (Elga 2010, p. 177). Such a thesis, Elga claims, is “completely uncontroversial” and “every sensible view on disagreement should accommodate it” (Elga 2010, p. 178). In other words:

Independence: The opinions of others have epistemic force only to the extent that they are independent of one another.

On this reading, then, those in echo chambers are moved by the number of, say, posts or shares on social media, but each one has no additional value because of a lack of independence from the others.

There are, however, at least two concerns with this diagnosis of the distinctively epistemic problem regarding echo chambers. First, it turns out that dependence is not nearly as epistemically devastating as is often suggested. To see this, let’s start with a plausible online version of Wittgenstein’s newspaper scenario. Suppose every member of a 200-person group on social media posts a comment explaining why Brexit is a mistake. Suppose, further, that each of them is simply repeating what they learned from the same exact online news source, say, *The Guardian*. In such a case, the standard view of echo chambers says that the number of voices commenting on Brexit being a mistake belies the fact that all of them reduce to the same, single source. And it is only this single source that has legitimate epistemic force.

But note that even if we all rely on *The Guardian* in our beliefs about Brexit, we may do so with varying degrees of autonomy. *Autonomous dependence* involves a subject exercising agency in her reliance on a source of information, which in turn involves minimally (1) possessing beliefs about the reliability and trustworthiness of the testimonial source, either in particular or in general, (2) monitoring the incoming testimony for defeaters, and (3) bearing responsibility for expressing the view in question.

Regarding (1), most subjects have countless beliefs that guide their consumption of information, and such beliefs often concern the *contexts* in which the testimony is offered, the *content* of the assertions, and the *speakers* in question.<sup>6</sup> A statement made in the context of a political campaign may be regarded with greater suspicion than one offered over coffee. An assertion about an alien abduction is more readily rejected than one about seeing a coyote in Illinois. And Donald Trump's testimony may be called into question more often than Barack Obama's. In these sorts of cases, subjects rely on their background beliefs to monitor and filter—often automatically—what information is allowed into their doxastic frameworks.

With respect to (2), defeaters can be either *doxastic* or *normative*, and either *rebutting* or *undercutting*.<sup>7</sup> A doxastic defeater is a doubt or belief that you have that indicates that one of your beliefs is either false (i.e., rebutting) or unreliably formed or sustained (i.e., undercutting). A normative defeater is similar, except it concerns doubts or beliefs that you *should have*, given the evidence available to you. So, for example, if I believe that the animal in my backyard is a bobcat by seeing one there, I might get powerful evidence that such a belief is false by you telling me that bobcats have never lived in my state, or that my basis is a poor one by my optometrist reporting to me how much my vision has deteriorated. If I accept both instances of testimony, then I have doxastic defeaters; rebutting in the first case, undercutting in the second. But even if I reject the testimony in question, I am still on the hook for this counterevidence if I do so for no good reason at all. Why? Because it is evidence that I *should have*.<sup>8</sup> The justification that my bobcat-belief might have initially enjoyed, then, has been normatively defeated.

Applying this to the case at hand, if 200 users post about Brexit because of what they autonomously learned in *The Guardian*, the epistemic support for the claim in question goes far beyond that of the author of the original article itself. This is because such a view was filtered through 200 different doxastic frameworks, which brings along potential differences in reliability assessments, belief acceptance, and defeating conditions. Once 200 people post what they learned in *The Guardian*, then, this has the support of having survived all of these additional sets of beliefs, both in having positive assessments leading to acceptance and in avoiding defeaters.

This brings us to the third feature of autonomous dependence: It is widely accepted that there is an epistemic norm governing assertion, which means that

<sup>6</sup> See Lackey (2008).

<sup>7</sup> For various views of defeaters, approached in a number of different ways, see BonJour (1980 and 1985), Nozick (1981), Goldman (1986), Pollock (1986), Fricker (1987 and 1994), Chisholm (1989), Burge (1993 and 1997), Plantinga (1993), McDowell (1994), Audi (1997 and 1998), Bergmann (1997), Williams (1999), BonJour & Sosa (2003), Hawthorne (2004), Reed (2006), and Lackey (2008).

<sup>8</sup> For a very nice development of the notion of “should have known,” see Goldberg (2015).

speakers need to have an appropriate epistemic standing with respect to a proposition in order to properly assert it.<sup>9</sup> Because of this, there is an important sense in which we bear responsibility for what we say. We can, for instance, be criticized for asserting something when it is merely a suspicion or a hunch. And so when there are 200 people offering an assertion about Brexit, rather than simply the author of the original piece, there are 200 additional speakers shouldering the responsibility for this claim. This, by itself, brings additional epistemic support, since a person autonomously vouching for the truth of a claim is certainly evidence. Thus, even though the views about Brexit are all entirely dependent on *The Guardian*, epistemic force is added by each member of the echo chamber through the autonomy of this dependence.

Now, it may be thought that this is all well and good, but people in echo chambers are simply not examples of autonomous dependence. Instead, they are far more like what Alvin Goldman has characterized as non-independence: “If two or more opinion-holders are totally *non-independent* of one another, and if the subject knows or is justified in believing this, then the subject’s opinion should not be swayed—even a little—by more than one of these opinion-holders” (Goldman 2001, p. 99). Goldman characterizes the notion of *non-independence* that is operative here in terms of conditional probability. In particular, where H is a hypothesis, X(H) is X’s believing H, and Y(H) is Y’s believing H, Y’s belief being totally non-independent of X’s belief can be expressed in the following way:

$$[\text{NI}]: P(Y(H)/X(H)\&H) = P(Y(H)/X(H)\&\sim H)$$

According to NI, Y’s probability for H conditional on X’s believing H and H’s being true is equal to Y’s probability for H conditional on X’s believing H and H’s being false. In other words, Y is just as likely to follow X’s opinion whether H is true or false. In such a case, Y is a *non-discriminating reflector* of X with respect to H.<sup>10</sup> “When Y is a non-discriminating reflector of X, Y’s opinion has no extra evidential worth for the agent above and beyond X’s opinion” (Goldman 2001, p. 101). For instance, in the case of a guru and his blind followers, Goldman writes: “a follower’s opinion does not provide any additional grounds for accepting the guru’s view (and a second follower does not provide additional grounds for accepting a first follower’s view) even if all followers are precisely as reliable as

<sup>9</sup> See, for instance, Unger (1975), Brandom (1983), Williamson (1996 and 2000), Adler (2002), DeRose (2002), Reynolds (2002), Cohen (2004), Hawthorne (2004), Stanley (2005), Weiner (2005), Douven (2006), Lackey (2007 and 2008), Whiting (2013), and Goldberg (2015).

<sup>10</sup> I will follow Goldman and talk about Y being *just as likely* to follow X’s opinion whether H is true or false. However, if Goldman wants this likelihood to be specifically tied to the subject’s ability to discriminate the true from the false, it may be more accurate to talk about Y being *just as inclined* to follow X’s opinion whether H is true or false. Otherwise, the likelihood of Y’s following X’s opinion could be affected by factors totally disconnected from Y’s discriminatory abilities, such as features in the environment, and so on.

the guru himself (or as one another)—which followers must be, of course, if they believe exactly the same things as the guru (and one another) on the topics in question” (Goldman 2001, p. 99). The blind follower is, then, a non-discriminating reflector of the guru with respect to the question at hand and thus Goldman claims that disagreement with the follower does not call for doxastic revision beyond that required by the guru’s belief.

In order for Y’s opinion to have additional worth for the agent above and beyond X’s opinion, Goldman argues that Y’s belief needs to be at least partially *conditionally independent* of X’s belief, which can be expressed as follows:

$$[\text{CI}]: P(Y(H)/X(H)\&H) > P(Y(H)/X(H)\& \sim H)$$

According to CI, Y’s probability for H conditional on X’s believing H and H’s being true is greater than Y’s probability for H conditional on X’s believing H and H’s being false. In other words, Y is more likely to follow X’s opinion when H is true than when H is false. Goldman claims that Y’s agreement with X regarding H provides evidence in favor of H for a third party, N, only if N has reason to think that Y used a “. . . more-or-less autonomous causal [route] to belief, rather than a causal route that guarantees agreement with X” (Goldman 2001, p. 102). Such an autonomous causal route is exemplified in cases where (1) “X and Y are causally independent eyewitnesses of the occurrence or non-occurrence of H,” or (2) “X and Y base their respective beliefs on independent experiments that bear on H,” or (3) Y’s belief in H goes partly through X but does not involve uncritical reflection of X’s belief (Goldman 2001, p. 102).

On this reading, then, most of the views of those in echo chambers are non-independent, rather than conditionally independent, in Goldman’s sense. In this way, when, say, 200 people assert what they learned from an article in *The Guardian*, we should not be swayed—even a bit—by the non-independent voices. Yet, all too often, we are unaware of the non-independence, and so we regard the views expressed in the echo chamber as far more epistemically supported than they in fact are.

This is where the second response to the “dependence” diagnosis of the epistemic problem regarding echo chambers arises: Very few of us as consumers of information are literally like the blind followers of a guru. Indeed, consider what this would amount to: We would be just as likely to accept what we are told when it is reported that cockatiels are mammals rather than birds; that onion rings are healthier than broccoli; that the Earth is under, rather than over, 100 years old; that I am shorter, rather than taller, than Michael Jordan; and so on. Even if, say, Democrats are very likely to accept what CNN reports, and Republicans are disposed to believe Fox, surely there are doxastic limits for most of us. Yet, Goldman’s notion of non-independence requires that the trust in our sources be

entirely blind in order for there to be a complete absence of additional epistemic support.

Moreover, notice that Goldman's account of non-independence is characterized in terms of *Y* being a non-discriminating reflector of *X* *with respect to H*, where this is understood in terms of *Y* being as likely to follow *X*'s opinion whether *H* is true or false. But even if *Y* is not a discriminating reflector of *H*, *Y* may nonetheless be discriminating when it comes to *X* or to sources like *X*.<sup>11</sup> To see this, consider the following: I may be a non-discriminating reflector of *The Guardian*'s testimony with respect to the question of whether Brexit is a mistake—that is, I am such that I would accept *The Guardian*'s opinion on this matter whether it is true or false—but I may be supremely discriminating when it comes to the kind of testimony that I generally accept. I may, for instance, know a great deal about *The Guardian*'s practices or I may know that the author of the article in question is competent and sincere in her reporting. Or I may be highly discriminating with respect to news sources in general or my news about Brexit in particular. The fact that I don't have independent evidence about the specific proposition in question does not thereby entail that I do not possess a great deal of other relevant evidence that enables me to function as an epistemic filter, thereby making my testimony have force beyond that provided by *The Guardian*'s report alone.

To put this point another way, suppose that there are two non-discriminating reflectors of *The Guardian*'s testimony with respect to the view about Brexit: Abby and Annie. Both would share *The Guardian*'s opinion on this question whether it is true or false, but only Abby is discriminating when it comes to the source of her information. In particular, Abby would be in such a non-discriminating relationship with a testifier only if she had good evidence of that source's general reliability and trustworthiness. Thus, Abby would be a non-discriminating reflector of *The Guardian*'s belief about Brexit only if she had good reason to think that *The Guardian* is a reliable news source and that the reporter in question is competent and trustworthy. Moreover, Abby is excellent at discriminating among news sources in general. Annie, on the other hand, is non-discriminating "all the way down," that is, she would be in such a non-discriminating relationship with a testifier regardless of the evidence that she possessed about the source's general reliability and trustworthiness. So, Annie would share *The Guardian*'s opinion about Brexit even if she had absolutely no reason to think that *The Guardian* is a reliable news source. Moreover, Annie is very poor at discriminating among news sources in general.

Now, compare a situation where Abby reports on Facebook that Brexit is a mistake solely on the basis of reading *The Guardian* with one where Annie does.

<sup>11</sup> Coady (2006) makes a similar point.

Surely the first situation provides more evidential support on behalf of the claim about Brexit than the one in the second. For even though both Abby and Annie are non-discriminating reflectors of *The Guardian's* opinion, and neither's belief about Brexit is even partially conditionally independent of *The Guardian's* report in Goldman's sense, Abby's trust of this news source is itself well-grounded while Annie's is not. That is, while Abby may be blindly trusting of *The Guardian's* testimony with respect to the question of the status of Brexit, she is neither blindly trusting of *The Guardian's* testimony across the board nor of news sources in general. This contrasts with Annie, whose blind trust extends to *The Guardian's* testimony on all other matters and to other news sources. Another way to express this point is that we need to ask not only whether someone is in an echo chamber, but how that person arrived there. If you are surrounded by like-minded people through careful discrimination, this is obviously epistemically different than ending up in this situation through chance.

What all of this shows is that even if we grant that the views of those in echo chambers are non-independent in Goldman's strict sense, this still would not show that the epistemic support of a group of non-independent believers reduces to that of the original source. When this is combined with the fact that most consumers of news are not literally comparable to a blind follower of a guru, it is reasonable to conclude that the dependence of beliefs within echo chambers does not entirely undermine the epistemic force of the additional views. Indeed, the filtering that occurs with each member brings important support, as does the responsibility that is shouldered for offering an assertion. In this way, voices in an echo chamber are not on an epistemic par with reading multiple copies of the same newspaper, as each voice arises from a different doxastic framework. This has the result that, at least as far as dependence is concerned, echo chambers are significantly better off epistemically than is typically assumed. We should, then, look elsewhere for our diagnosis of the distinctively epistemic problem with echo chambers.

A second candidate for the epistemic problem with echo chambers is a *lack of diverse viewpoints*. The carefully curated social media sites that many people across the globe subscribe to create a self-insulation from difference that Cass Sunstein says is responsible for the degradation of democracy. Recall what he says in his recent book, *#Republic: Divided Democracy in the Age of Social Media*: "In a well-functioning democracy, people do not live in echo chambers or information cocoons" (Sunstein 2017, p. ix.). Sunstein connects his view with the work of John Stuart Mill and John Dewey, both of whom emphasize the importance of being exposed to diverse perspectives.<sup>12</sup> Here is Mill on this topic: "It is hardly possible to overstate the value . . . of placing human beings in contact with other persons

<sup>12</sup> See Mill (1863) and Dewey (1902).

dissimilar to themselves, and with modes of thought and action unlike those with which they are familiar.... Such communication has always been, and is peculiarly in the present age, one of the primary sources of progress" (Mill 1863, p. 252).

What sort of progress? In *On Liberty*, John Stuart Mill famously defends free speech on the grounds that full and open discussion will promote the truth, maintaining that a society should not merely tolerate speech that is objectionable, but embrace it. He defends this conclusion with four arguments: first, because we are fallible, we must be open to the possibility that an opinion that deviates from the mainstream is true: "if any opinion is compelled to silence, that opinion may, for aught we can certainly know, be true. To deny this is to assume our own infallibility" (Mill 1863, p. 101). Second, an opinion that is generally false may nonetheless contain a portion of the truth that is missing from the prevailing view.<sup>13</sup> Third, "even if the received opinion be not only true, but the whole truth; unless it is suffered to be, and actually is, vigorously and earnestly contested, it will, by most of those who receive it, be held in the manner of a prejudice, with little comprehension or feeling of its rational grounds" (Mill 1863, p. 102). And, fourth, in the absence of vigorous debate, "the meaning of the doctrine itself will be in danger of being lost or enfeebled, and deprived of its vital effect on the character and conduct; the dogma becoming a mere formal profession, ineffectual for good, but cumbering the ground and preventing the growth of any real and heartfelt conviction from reason or personal experience" (Mill 1863, p. 102).

Diversity of opinions, then, enables a recognition of our fallibility, provides additional evidence, and encourages debate and deliberation, all of which, according to Mill, promote the truth. And yet, as was noted earlier, 61 percent of Millennials report getting their political news on Facebook, which significantly reduces the exposure to opposing views and disagreement. On this view, then, echo chambers are problematic because they insulate us from different beliefs which, in turn, has negative epistemic consequences.

There are, however, several problems with this diagnosis. First, restricting our information sources is not objectionable by itself, and can even have clear epistemic benefits. If I consult one highly reliable media outlet on a regular basis, rather than a number of conflicting ones, I will block out a lot of noise. With limited time and attention, this streamlines the consumption of news and increases the likelihood of acquiring true, and avoiding false, beliefs. Moreover, if I add other sources simply to avoid worries about insulation and the illegitimate reinforcement of beliefs—without any regard for their reliability—I will end up

<sup>13</sup> "[T]hough the silenced opinion be an error, it may, and very commonly does, contain a portion of the truth; and since the general or prevailing opinion on any subject is rarely or never the whole truth, it is only by the collision of adverse opinions that the remainder of the truth has any chance of being supplied" (Mill 1863, pp. 101–2).

out of an echo chamber and exposed to a diversity of opinions, but far worse off as a knower. Compare: If I learn about climate change from a reputable environmental scientist, there is only the danger of acquiring false beliefs in also consulting a climate change denier.

Second, the “diversity” diagnosis of echo chambers lends itself to a “we’re all in the same boat” attitude about consumption of information. Trump watches only Fox News for his information on climate change and you read only the *Journal of Applied Meteorology and Climatology*: You’re both in echo chambers because of a lack of exposure to opposing or diverse viewpoints. According to Sunstein, then, you are both equally engaging in irresponsible intellectual behavior that is at odds with deliberative democracy. But this is obviously false. If you know that one source is highly politically biased while the other undergoes blind review and fact-checking, there is very little epistemic benefit to adding the former to your news cycle.

Finally, this analysis of echo chambers promotes simple solutions to deep social crises that threaten truth itself. Sunstein, for instance, recommends that we save democracy by having Facebook and other social media sites include buttons where users could click to see opposing viewpoints. If, for instance, you are reading an article linked to via Facebook that explains why Brexit is a mistake, there would be an “opposing viewpoint” option that you could click to see why Brexit is the right course for the British people. However, even if this were successful at exposing people to disagreement—and this is a big if—how does this make us epistemically better off? When I’m reading an article about the impact of climate change on wildlife, what is the benefit of clicking on a button from the perspective of a climate change denier? When I’m reading an article about the Sandy Hook school shooting, what epistemic advantage is there to also learning about the perspective of a Sandy Hook truther? Sure, I will be exposed to different views, but at the expense of something even more fundamental to democracy: truth.

It is helpful to take a step back here and notice a similarity between the dependence and the lack of diversity diagnoses of echo chambers: They both understand the epistemic problem with an echo chamber—such as the one targeting Trump’s consumption of news—as structural in nature. Indeed, echo chambers themselves are *content-neutral*. When we’re criticized for being in one, this raises an objection with *how* we’re taking in information, rather than with *what* we’re taking in. For instance, on the dependence reading, the epistemic problem is a lack of appropriate independence between viewpoints, but it says nothing at all about the content or truth value of the beliefs. On the lack of diversity reading, the objection is simply that there fails to be opposing viewpoints, but, again, there is no mention of what the beliefs amount to. We have seen, however, that both of these readings fail to capture a serious epistemic problem with echo chambers, and thus they also fail to explain what is clearly wrong with Trump’s relationship with Fox News.

## 2. Fake News

If the problem in the opening scenario of this paper is not that Trump is creating an echo chamber, what is it? The real danger with relying only on Fox News is not one of structure, but of content. In other words, it is not that Trump is relying on a single source for news, but that he is relying on one that peddles falsehoods at an alarming rate.<sup>14</sup>

PunditFact provides a scorecard for the truth of statements made on air by Fox News and their pundit guests: they rate only 10 percent as true.<sup>15</sup> One particularly vivid example illustrates this. Recently, Fox News anchor, Charles Payne, spoke to a former ICE agent about the “caravan” of asylum seekers from Central America.<sup>16</sup> During the interview, the agent claimed that the migrants are “coming in with diseases, such as smallpox and leprosy and TB that are going to infect our people in the United States.” *Smallpox*. The agent said that the asylum seekers have smallpox that will spread to us here in the U.S.—a disease that was declared “eradicated in 1980 following a global immunization campaign led by the World Health Organization.”<sup>17</sup> When Payne was later challenged about this claim, he responded that they “have no way of independently confirming this.”<sup>18</sup> So, Fox News responded to an overtly false claim by issuing another patently false statement, as it is not difficult to confirm that the last known natural case of smallpox was in 1977 in Somalia.<sup>19</sup>

Moreover, recall that according to a Pew Research Center survey “Fox News was the main source for 40% of Trump voters” during the 2016 election.<sup>20</sup> And another Pew Survey indicates that “When it comes to choosing a media source for political news, conservatives orient strongly around Fox News. Nearly half of consistent conservatives (47%) name it as their main source for government and political news.”<sup>21</sup> So, the main source of news for Trump supporters and

<sup>14</sup> An analysis of how we should understand “fake news” lies outside the scope of this paper. For various views, see Rini (2017), Gelfert (2018), Mukerji (2018), Habgood-Coote (2019), and Grundmann (unpublished). While there may be important differences between the general category of “false news” and the more specific “fake news,” I will gloss over them in what follows.

<sup>15</sup> See <https://www.politifact.com/punditfact/tv/fox/> (accessed September 10, 2019).

<sup>16</sup> See <https://slate.com/news-and-politics/2018/10/daily-dem-panic-meter-fox-news-anchor-shepard-smith-caravan-hysteria.html?fbclid=IwAR0b3HAbG9vcZUW-VnkX1Zo6vG-t9GENN-dJ01W2WAX2k20N17l4Kr03wNI> (accessed September 10, 2019).

<sup>17</sup> See <https://www.who.int/csr/disease/smallpox/en/> (accessed September 10, 2019).

<sup>18</sup> See [https://talkingpointsmemo.com/news/fox-news-admits-no-way-to-confirm-guests-lie-that-asylum-seekers-will-bring-smallpox?fbclid=IwAR035G102h4PE\\_I5NPGm9gJ6ofNDVzmXhJ95djqQ1jrt6V1FVSxM15Iz5Y](https://talkingpointsmemo.com/news/fox-news-admits-no-way-to-confirm-guests-lie-that-asylum-seekers-will-bring-smallpox?fbclid=IwAR035G102h4PE_I5NPGm9gJ6ofNDVzmXhJ95djqQ1jrt6V1FVSxM15Iz5Y) (accessed September 10, 2019).

<sup>19</sup> See <https://www.who.int/csr/disease/smallpox/en/> (accessed September 10, 2019).

<sup>20</sup> See <https://www.journalism.org/2017/01/18/trump-clinton-voters-divided-in-their-main-source-for-election-news/> (accessed September 10, 2019).

<sup>21</sup> See <https://www.journalism.org/2014/10/21/political-polarization-media-habits/> (accessed September 10, 2019).

conservatives is also said to have pundit guests who speak the truth only 10 percent of the time.

Still further, it has been shown that falsehoods have far greater power and reach than the truth does online. A recent study (Vosoughi et al. 2018), for instance, examined the diffusion of true and false news online by looking at rumor cascades on Twitter, which are unbroken retweet chains with a single common origin. The data set included 126,000 rumor cascades spread by over 3 million people more than 4.6 million times between 2006 and 2017. The authors found that “falsehood diffused significantly farther, faster, deeper, and more broadly than the truth in all categories of information” (Vosoughi et al. 2018, p. 1147). For instance, false news cascades reached between 1,000 and 100,000 people, while the true ones rarely extended beyond 1,000 people. In addition, it took the truth about six times longer to reach 1,500 people than falsehoods did. Still further, these differences between true and false rumor cascades were especially pronounced when it comes to political news. False political news, for instance, reached more than 20,000 people nearly three times faster than all other types of news reached 10,000 people (Vosoughi et al. 2018, p. 1148).

What is especially important for our purposes is that these findings could not be accounted for by structural features of either the networks or the users. As the authors write:

Perhaps those who spread falsity “followed” more people, had more followers, tweeted more often, were more often “verified” users, or had been on Twitter longer. But when we compared users involved in true and false rumor cascades, we found that the opposite was true in every case. Users who spread false news had significantly fewer followers . . . , followed significantly fewer people . . . , were significantly less active on Twitter . . . , were verified significantly less often . . . , and had been on Twitter for significantly less time . . . . Falsehood diffused farther and faster than the truth despite these differences, not because of them.

(Vosoughi et al. 2018, p. 1149)

False rumor cascades are not, therefore, more powerful than the true ones because of the *way* in which the information is dispersed. Instead, the *content* of the cascades seems to be at least one of the central driving forces. For instance, false rumors were significantly more *novel* than true ones across all novelty metrics, such as displaying much higher information uniqueness. Moreover, false news inspired greater responses from users of surprise or disgust, while true information elicited reactions of sadness, anticipation, joy, and trust (Vosoughi et al. 2018, p. 1150).<sup>22</sup>

<sup>22</sup> It is of further interest to note that the conclusions of this study remained the same, with or without the presence of bots. Using a sophisticated bot-detection algorithm, the authors identified all

This raises an urgent problem. At a time when lies, particularly in politics, are propagated in unique ways—and to unprecedented degrees—the human fascination with the novel is leading to an epistemic crisis. Trump, for instance, uses Twitter in ways unseen by previous political leaders, and he reportedly averages about eight lies a day in his public life since taking office.<sup>23</sup> A recent CNN story highlights the thirty-six most outrageous claims Trump made in a single interview.<sup>24</sup> Given the platform that the President of the United States has, combined with the often shocking nature of his tweets,<sup>25</sup> we have the ingredients for a perfect storm against truth: Trump tweets a lie. Because of its original content, social media users are drawn to it. They pick it up and retweet it. News outlets that correct the record may be ignored or drowned out because of the ordinary and familiar nature of their statements. In the meantime, the next interesting falsehood has been tweeted, captivating the attention of the bulk of social media users. The cycle repeats itself.

This, of course, is not to say that echo chambers are never epistemically problematic, and that our attention should be only on fake news. Rather, the point is that to focus on the echo chamber diagnosis of someone like Trump, independently of, or prior to, a discussion of the content of the views in his closed network of followers is to put the epistemic cart before the horse. If you are a member of an echo chamber of proponents of climate change, and the expression of views is autonomously dependent on a single source, this is not only epistemically benign, but can have positive value. An important truth being filtered through hundreds of belief systems and being widely shared on social media is an epistemically powerful way of spreading knowledge. But a pernicious lie being repeated by hundreds of Twitter followers simply because its content is expressed by a Fox News source, or by Trump, or because it is novel, is epistemically destructive. It erodes the reliability of our beliefs, our trust in other members of our epistemic communities, and ultimately the truth itself.

### 3. Bots

I now want to connect issues from the previous two sections to highlight a game-changing epistemological problem currently facing social media users.

bots from the analysis, and found that while the presence of bots accelerated the spread of true and false information, it did so roughly equally. “This suggests that false news spreads farther, faster, deeper, and more broadly than the truth because humans, not robots, are more likely to spread it” (Vosoughi et al. 2018, p. 1150). I will discuss bots in a bit more detail later in this chapter in relation to a 2017 study that showed that bots do spread information at a significantly faster pace than humans.

<sup>23</sup> See <https://www.washingtonpost.com> (accessed 10 September 2019).

<sup>24</sup> See <https://www.cnn.com/2018/11/15/politics/daily-caller-trump-interview/index.html> (accessed 10 September 2019).

<sup>25</sup> See <https://www.usatoday.com/story/news/politics/onpolitics/2018/09/13/trump-denial-hurricane-maria-death-toll-puerto-rico-fuels-fury/1288530002/> (accessed 10 September 2019).

It may be recalled that one of the key arguments I offered against echo chambers being epistemically problematic, by themselves, is that much of what we learn and then report is done with at least a minimal amount of autonomy. As I said, rarely do we literally behave as a blind follower of a guru, and so there is at least some rudimentary discrimination that goes into our belief-forming, and social-media-sharing, practices. At the very least, in order for us to accept a view, it has to pass through our own belief frameworks, and so the more of these that a view passes through, the more epistemic support it has, even where there is a significant amount of dependence among the beliefs.

But all of this epistemic support vanishes when the members of the echo chambers in question are not exercising the relevant kind of autonomy—indeed, when they are simply such that they cannot take on views with autonomy.

A 2017 Pew Research Center study found that 66 percent of all tweeted links to popular news and current event websites come from social media bots—“automated accounts capable of posting content or interacting with other users with no direct human involvement.”<sup>26</sup> Moreover, the percentage of tweeted links from bots is even higher among certain kinds of news sites. For example, 89 percent of tweeted links to popular aggregation sites that compile stories from around the web are posted by bots.

The study also found that a relatively small number of highly active bots are responsible for a significant percentage of the links to prominent news and media sites. For instance, the 500 most active bot accounts are responsible for 22 percent of the tweeted links to popular news and current events sites. In contrast, the 500 most active human users are responsible for an estimated 6 percent of tweeted links to these outlets. So, bots tend to also be more active, and efficient, at spreading news than humans are.

This raises a significant concern with our online consumption of news. One of the reasons I argued that echo chambers are not as epistemically harmful as they are typically said to be is because we can often count on others to exercise at least minimal autonomy in their belief-forming and social-media-sharing practices. We count on users to have belief sets through which the posted content has been filtered. But bots crucially lack these features. They do not exercise autonomy; they do not have sets of beliefs that are filtering content; they do not assess reliability of sources; they cannot be regarded as irrational for having beliefs in the face of defeaters; and they do not bear responsibility for their assertions. Yet from the outside, they appear indistinguishable from human users.

So, while standard echo chambers are epistemically problematic only when the content and truth value of the views in the closed system are also considered, the prevalence of social media bots raises a further, pressing concern. Echo chambers

<sup>26</sup> See <https://www.pewinternet.org/2018/04/09/bots-in-the-twittersphere/> (accessed September 10, 2019).

filled with bots amplify voices that are not really voices at all. The expressed views do not reflect beliefs because there are no believers. In this way, echo chambers become seriously problematic when there are “news-approvers” who appear to be posting with autonomy but in fact are not even human. It is, thus, not only fake news that raises epistemic pitfalls for us online, but also fake news-approvers.

Notice, however, that the two standard diagnoses of the epistemic problem with echo chambers—dependence and a lack of diversity—do not capture what is wrong here. It is not that the “beliefs” of bots or fake news-approvers are inappropriately dependent but appear independent. There are simply no relevant beliefs to even be assessed for their dependence on the beliefs of others. So, this diagnosis seems to miss the mark. Moreover, an absence of diversity surely isn’t relevant here. Bots have been shown to be more than happy to spread a wide range of opposing viewpoints. Indeed, the noise of being inundated with conflicting opinions can often be effective in achieving the sort of epistemic confusion and chaos desired by those who set up bot farms in the first place.

A further point to note is that the prevalence of bots cannot be understood as simply increasing the *quantity* or changing the *location* of epistemic obstacles that we’ve been facing all along. Bernard Williams, for instance, writes:

the Internet shows signs of creating for the first time what Marshall McLuhan prophesied as a consequence of television, a global village, something that has the disadvantages both of globalization and of a village . . . It constructs proliferating meeting places for the free and unstructured exchange of messages . . . The chances that many of these messages will be true are low, and the probability that the system itself will help anyone to pick out the true ones is even lower. In this respect, post-modern technology may have returned us dialectically to a transmuted version of the pre-modern world, and the chances of acquiring true belief by these means, except for those who already have the knowledge to guide them, will be much like those in the Middle Ages. (2002, p. 216)

Williams predicted that the problems we would see in our online interactions would be simply versions of age-old ones—indeed, he compares such a situation to the one facing consumers of information in the Middle Ages. We have always been exposed to gossip, conspiracy theories, lies, bullshit, suspicions, and so on. All that is different, it may be argued, is that the Internet has increased the quantity and the accessibility of such epistemic pitfalls.

What we have seen, however, is that the prevalence of social media bots is an epistemological game-changer. Never before have our epistemic communities been infiltrated by non-human testifiers disguised as being just like you and me. The psychological theories we rely upon to understand speakers and, thus, the tools that we invoke to assess the credibility of testifiers are tailored to humans, not bots. Moreover, the prevalence of online bots upends the expectations we have

as members of an epistemic community. Minimally, we count on most people thinking twice before posting information, and we expect that they will take responsibility for the content of their assertions if challenged. But when two-thirds of the members of some of our online communities are not human testifiers, the very foundation of our systems for sharing information is threatened.

#### 4. Social Epistemology

In this last section, I would like to explore an important conclusion about social epistemology that can be drawn from our earlier considerations about echo chambers, fake news, and bots.

Ideal theory in ethics and political theory, often paradigmatically exemplified by the work of John Rawls in *A Theory of Justice*, is frequently critiqued for all that it leaves out of the theoretical picture. Indeed, it is not the appeal to ideals themselves that is regarded as distinctively problematic, since non-ideal theorists will also invoke moral ideals, but the absence of attention paid to the way the world actually is.

This is a point that is developed extensively by Charles Mills, who characterizes ideal theory in a recent book primarily in terms of what is absent or ignored rather than by what is present. He writes:

What distinguishes ideal theory is the reliance on idealization to the exclusion, or at least marginalization, of the actual . . . [I]deal theory either tacitly represents the actual as a simple deviation from the ideal, not worth theorizing in its own right, or claims that starting from the ideal is at least the best way of realizing it.

(Mills 2017, p. 75)

According to Mills, then, the core feature of ideal theory is not the idealization itself but, rather, the disregard of varying degrees of the actual world and the people and institutions in it. He goes on to argue further that ideal theory will use some or all of a list of concepts and assumptions, including idealized capacities and ideal social institutions.

I want to extend Mills' considerations here and argue that we can distinguish between ideal and non-ideal epistemology, and that the challenges facing consumers of online information make vivid and pressing why we need to do non-ideal epistemology. To this end, let's take a look at some of the representative views in the epistemology of testimony. Since the news we take in online involves accepting the testimony of others, these approaches will be especially relevant for our purposes.

Richard Moran describes testimonial exchanges in the following way: "The speaker, in presenting his utterance as an *assertion*, one with the force of *telling*

the audience something, presents himself as *accountable* for the truth of what he says, and in doing so he offers a kind of guarantee for this truth" (Moran 2006, p. 283, original emphasis). In a similar spirit, Tyler Burge maintains that "A person is entitled to accept as true something that is presented as true and that is intelligible to him, unless there are stronger reasons not to do so" (Burge 1993, p. 467). Matthew Weiner holds that "We are justified in accepting anything that we are told unless there is positive evidence against doing so" (Weiner 2003, p. 257). Robert Audi argues that "Gaining testimonially grounded knowledge normally requires only having no reason for doubt about the credibility of the attester" (Audi 1998, p. 142). And Paul Faulkner claims that:

The idea that conversation be seen as a cooperative endeavor . . . yields a pair of social norms. The prescription that speakers follow the Cooperative Principle and its maxims describes a *social norm of trustworthiness*. And the paired prescription that as audiences we presume this of speakers and act as if we believe that they are following the Cooperative Principle and its maxims describes a *social norm of trust* . . . I . . . take . . . this pair of norms [to describe] our conversational practices. (Faulkner 2010, p. 132)

There are two points to note about the views expressed in these quotes. First, they rely on the idealization of testimonial exchanges, particularly idealized conversational practices. For instance, Faulkner takes cooperative conversations as his theoretical starting point and then argues on behalf of norms of trustworthiness and trust. Indeed, Faulkner suggests that our actual practices are generally cooperative in the ways described by these norms. But even when the authors aren't this explicit, it is clear that they regard trustworthiness and trust as the default position of testimonial exchanges, and so assume that speakers and hearers are generally following the relevant norms. Moran talks about speakers offering a kind of guarantee of the truth regarding what they assert, and Burge, Weiner, and Audi all maintain that hearers are justified in accepting what they're told even when they have no reasons for doing so. The second point to note is that just as Mills argues that ideal theory represents the actual as a simple deviation from the ideal, these theorists relegate to "defeating conditions" the multitude of ways in which things can go awry in testimonial exchanges to the exclusion or marginalization of features of the actual world. In particular, the bulk of the philosophical work in these views focuses on exchanges that go exactly as they should, with virtually no attention paid to understanding or navigating when hearers have "positive evidence against" trusting speakers or have "reason for doubt."

As we saw above, however, online testimonial exchanges are far from cooperative conversations; they are riddled with lies, fake news, and highly efficient bots. Thus, the "defeating conditions" that are relegated to the margins of work in the epistemology of testimony arguably ought to take center stage when considering

speakers and hearers who are online. This is where non-ideal social epistemology needs to step in. Rather than theorizing only about the giving and receiving of testimony in ideal situations, more work needs to be done on how recipients of testimony should navigate the world as it actually is. For instance, given what we know about the amount and efficacy of falsehoods online, it seems unwise to follow Moran, Burge, Weiner, and Audi and accept everything we read on the Internet unless we have specific evidence against it. Similarly, the prevalence of fake news renders many online conversations overtly uncooperative, and it isn't even clear that interaction with bots can be characterized as conversations in the first place. Social epistemologists should thus be looking at ways in which we can promote epistemic success and health online in spite of the myriad ways in which we are set up to fail.

It would be a separate chapter to take this project on myself, but let me close by gesturing toward some of the questions for future research in this branch of non-ideal social epistemology. As noted earlier in the discussion of autonomous dependence, there are at least three broad categories of assessments that are relevant to our acceptance of testimony. First, we evaluate the *contexts* in which testimony is offered, both at the type and token level. Reports offered by a politician in the middle of campaigning or by a random online source may be met with caution, but statements offered at a conference about climate change by the leading scientists in the field might be generally accepted. Second, we assess the *content* of the reports we receive, again at both the type and token level. We may, for instance, be more skeptical of the reports parents offer about the accomplishments of their children or about the latest conspiracy theory than we are about what someone had for breakfast. Finally, we evaluate *sources* or *speakers*, and we do so with respect to their reliability, sincerity, and overall credibility. So, for instance, we might have good reasons for trusting CNN over Fox News and we may have evidence for accepting what Anderson Cooper reports but not what Rush Limbaugh does.

With these three categories in mind, some questions to explore include: How do our assessments of sources, contents, and contexts differ online and offline? How do we evaluate the credibility of unknown or anonymous sources? Are there ways of reliably discriminating between fake and real news? Are there strategies for identifying the reports or activities of bots? How do we individuate testimonial contexts online?<sup>27</sup>

Of course, these represent merely a handful of the relevant questions, but what I hope to have shown is that the changing landscape of our consumption of information raises urgent epistemological questions, and social epistemology is in a unique position to step in and address them.

<sup>27</sup> This is not to suggest that there hasn't already been some important work on these questions.

## 5. Conclusion

In this chapter, I argued that echo chambers are not, by themselves, epistemically problematic, and hence that diagnosing the problem with Trump's consumption of news by saying he is in an echo chamber is fundamentally incorrect. I then showed that we need to look for a content-driven, rather than a structural, explanation of what is wrong with this sort of epistemic behavior, and that this shifts the focus to discriminating between fake and real news. Finally, I claimed that the prevalence of social media bots poses a unique, game-changing problem for the epistemology of testimony, one that should be addressed through a non-ideal lens, especially as our epistemic lives are increasingly being played out online.<sup>28</sup>

## References

Adler, Jonathan E. 2002. *Belief's Own Ethics*. Cambridge, MA: The MIT Press.

Audi, Robert. 1997. "The Place of Testimony in the Fabric of Knowledge and Justification." *American Philosophical Quarterly* 34: 405–22.

Audi, Robert. 1998. *Epistemology: A Contemporary Introduction to the Theory of Knowledge*. London and New York: Routledge.

Bergmann, Michael. 1997. "Internalism, Externalism and the No-Defeater Condition." *Synthese* 110: 399–417.

BonJour, Laurence. 1980. "Externalist Theories of Epistemic Justification." *Midwest Studies in Philosophy* 5: 53–73.

BonJour, L. 1985. *The Structure of Empirical Knowledge*. Cambridge, MA: Harvard UP.

BonJour, L. & E. Sosa. 2003. *Epistemic Justification: Internalism vs. Externalism, Foundationalism vs. Virtues*. Oxford: Blackwell.

Brandom, Robert. 1983. "Asserting." *Noûs* 17: 637–50.

Burge, Tyler. 1993. "Content Preservation." *The Philosophical Review* 102: 457–88.

Burge, Tyler. 1997. "Interlocution, Perception, and Memory." *Philosophical Studies* 86: 21–47.

Chisholm, Roderick M. 1989. *Theory of Knowledge*, 3rd edn. Englewood Cliffs, NJ: Prentice-Hall.

Coady, David. 2006. "When Experts Disagree." *Episteme* 3: 68–79.

<sup>28</sup> For very helpful comments on earlier drafts of this chapter, I'm grateful to Baron Reed and audience members at the 11th Principia International Symposium in Florianópolis, Brazil; the Epistemic Norms for the New Public Sphere Workshop at the University of Warwick; the Ignorance in the Age of Information Conference at Scripps College; and Colorado State University.

Cohen, Stewart. 2004. "Knowledge, Assertion, and Practical Reasoning," in Ernest Sosa and Enrique Villanueva (eds.), *Philosophical Issues* 14: 482–91.

DeRose, Keith. 2002. "Assertion, Knowledge, and Context." *The Philosophical Review* 111: 167–203.

Dewey, John. 1902. "Academic Freedom." *Educational Review* 23: 1–14.

Douven, Igor. 2006. "Assertion, Knowledge, and Rational Credibility." *The Philosophical Review* 115: 449–85.

Elga, Adam. 2010. "How to Disagree About How to Disagree," in Richard Feldman and Ted Warfield (eds.), *Disagreement*. Oxford: Oxford University Press, pp. 175–86.

Faulkner, Paul. 2010. "Norms of Trust," in Adrian Haddock, Alan Millar, and Duncan Pritchard (eds.), *Social Epistemology*. Oxford: Oxford University Press, pp. 129–47.

Fricker, Elizabeth. 1987. "The Epistemology of Testimony." *Proceedings of the Aristotelian Society*, supp. vol. 61: 57–83.

Fricker, Elizabeth. 1994. "Against Gullibility," in Bimal Krishna Matilal and Arindam Chakrabarti (eds.), *Knowing from Words*. Dordrecht: Kluwer Academic Publishers, pp. 125–61.

Gelfert, Axel. 2018. "Fake News: A Definition." *Informal Logic* 38: 84–117.

Goldberg, Sanford. 2015. *Assertion: On the Philosophical Significance of Assertoric Speech*. Oxford: Oxford University Press.

Goldman, Alvin I. 1986. *Epistemology and Cognition*. Cambridge, MA: Harvard University Press.

Goldman, Alvin I. 2001. "Experts: Which Ones Should You Trust?" *Philosophy and Phenomenological Research* 63: 85–110.

Grundmann, Thomas. Unpublished. "Fake News: The Case for a Consumer-Oriented Explication."

Habgood-Coote, Joshua. 2019. "Stop Talking About Fake News!" *Inquiry* 62: 1033–65.

Hawthorne, John. 2004. *Knowledge and Lotteries*. Oxford: Oxford University Press.

Jamieson, Kathleen Hall & Joseph N. Cappella. 2008. *Echo Chamber: Rush Limbaugh and the Conservative Media Establishment*. Oxford: Oxford University Press.

Kelly, Thomas. 2010. "Peer Disagreement and Higher Order Evidence," in Richard Feldman and Ted Warfield (eds.), *Disagreement*. Oxford: Oxford University Press, pp. 111–74.

Lackey, Jennifer. 2007. "Norms of Assertion." *Noûs* 41: 594–626.

Lackey, Jennifer. 2008. *Learning from Words: Testimony as a Source of Knowledge*. Oxford: Oxford University Press.

McDowell, John. 1994. "Knowledge by Hearsay," in Bimal Krishna Matilal and Arindam Chakrabarti (eds.), *Knowing from Words*. Dordrecht: Kluwer Academic Publishers, pp. 195–224.

Mill, John Stuart. 1863 [1859]. *On Liberty*. Boston, MA: Ticknor Fields.

Mills, Charles. 2017. *Black Rights/White Wrongs: The Critique of Racial Liberalism*. Oxford: Oxford University Press.

Moran, Richard. 2006. "Getting Told and Being Believed," in Jennifer Lackey and Ernest Sosa (eds.), *The Epistemology of Testimony*. Oxford: Oxford University Press, pp. 272–306.

Mukerji, Nikil. 2018. "What is Fake News?" *Ergo* 5: 923–46.

Nguyen, C. Thi. 2020. "Echo Chambers and Epistemic Bubbles." *Episteme* 17: 141–61.

Nozick, Robert. 1981. *Philosophical Explanations*. Cambridge, MA: The Belknap Press.

Plantinga, Alvin. 1993. *Warrant and Proper Function*. Oxford: Oxford University Press.

Pollock, John. 1986. *Contemporary Theories of Knowledge*. Totowa, NJ: Rowman and Littlefield.

Reed, Baron. 2006. "Epistemic Circularity Squared? Skepticism about Common Sense." *Philosophy and Phenomenological Research* 73: 186–97.

Reynolds, Steven L. 2002. "Testimony, Knowledge, and Epistemic Goals." *Philosophical Studies* 110: 139–61.

Rini, Regina. 2017. "Fake News and Partisan Epistemology." *Kennedy Institute of Ethics Journal* 27: 43–64.

Robson, Jon. 2014. "A Social Epistemology of Aesthetics: Belief Polarization, Echo Chambers, and Aesthetic Judgment." *Synthese* 191: 2513–28.

Stanley, Jason. 2005. *Knowledge and Practical Interests*. Oxford: Oxford University Press.

Sunstein, Cass. 2017. *#Democracy: Divided Democracy in the Age of Social Media*. Princeton, NJ: Princeton University Press.

Unger, Peter. 1975. *Ignorance: A Case for Skepticism*. Oxford: Oxford University Press.

Vosoughi, Soroush, Deb Roy, & Sinan Aral. 2018. "The Spread of True and False News Online." *Science* 359: 1146–51.

Weiner, Matthew. 2003. "Accepting Testimony." *The Philosophical Quarterly* 53: 256–64.

Weiner, Matthew. 2005. "Must We Know What We Say?" *The Philosophical Review* 114: 227–51.

Whiting, Daniel. 2013. "Stick to the Facts: On the Norms of Assertion." *Erkenntnis* 78: 847–67.

Williams, Bernard. 2002. *Truth and Truthfulness: An Essay in Genealogy*. Princeton, NJ: Princeton University Press.

Williams, Michael. 1999. *Groundless Belief: An Essay on the Possibility of Epistemology*, 2nd edn. Princeton, NJ: Princeton University Press.

Williamson, Timothy. 1996. "Knowing and Asserting." *The Philosophical Review* 105: 489–523.

Williamson, Timothy. 2000. *Knowledge and its Limits*. Oxford: Oxford University Press.

# The Dissemination of Scientific Fake News

## On the Ranking of Retracted Articles in Google

*Emmanuel J. Genot and Erik J. Olsson*

### 1. Introduction

Fake news can originate from an ordinary person carelessly posting what turns out to be false information (Maheshwari, 2016) or from the intentional actions of fake news factory workers (Sydell, 2016), but it can also originate from scientific fraud. Broadly speaking, a published scientific article based on fraud qualifies as “scientific fake news”, although we will prefer the term “fake science”—what looks like science but in fact is not—in an effort to avoid contentious issues regarding the exact definition of “fake news” (cf. McIntyre, 2018). In 2016, a Swedish research group published an article in the journal *Science* claiming that fish eat plastic and that this has dire consequences for the eco-system (Lönnstedt & Eklöv, 2016). This was picked up by many news organizations, including the BBC, which covered the study in an article (McGrath, 2016), including the fact that it was published in the prestigious journal *Science*. A normal reader of the BBC article would be likely to conclude that the Swedish study is true. However, about a year later, *Science* retracted the article on grounds of scientific fraud. In spite of this, one of us found that weeks after the retraction, Google continued to rank the BBC’s coverage of the article higher than content, including *Science*’s own notice, informing about the retraction when searching on “Fish eat plastic”.

Indeed, more than one year later, in January of 2018, search in Swedish Google returned only articles from environmentalist websites covering the original study on the first search result page, often with endorsement (“the study shows that . . .”) and without information about the retraction. This was after, in addition to *Science*, two independent committees and, finally, the vice chancellor of the university in question had declared the study fraudulent (Koffmar, 2017)—just to exclude the potential objection that the Swedish study is solid after all, and Google knows this and the scientific community doesn’t.

One of us re-did the same search a few months later (mid-2018) with similar results. Search on “fish eat plastic” returned the original BBC article on the first Google result page with the most highly ranked information about the retraction,

in an article in *Fortune* (Morris, 2017), being found on the second result page. The same was true for Google Scholar. According to Google, then, the original and now seriously misleading BBC article was still more “important” or “relevant” than the true and non-misleading article about the retraction. This is simply the wrong result. For a user interested in the topic of the article it would be more important or relevant to learn that it has been retracted than to be presented with coverage of the original, fraudulent article.

How common is this phenomenon and how deeply does it affect Google? In this chapter, we conduct a pilot study testing one aspect of the, as it turns out, complex issue of retraction due to scientific fraud in Google, namely, the extent to which Google ranks original content higher than the notice of the retraction, drawing on a sample of articles retracted due to fabrication of data in the Retraction Watch public database (<https://retractionwatch.com/>). As a preliminary, we introduce, in Section 2, the problem of retraction in search engines and conjecture that it might be a “deep” problem given (a) underlying algorithms that prioritize popularity and (b) a hypothesized tendency in humans to think that information about retraction is less interesting, making it therefore less popular, than the content of the original article. In Section 3, we formulate three testable hypotheses regarding the relative ranking of links to original articles and links indicating their retraction. Our results are provided in Section 4 and discussed in Section 5, where we also propose a number of ways to extend our work.

## 2. Background: Why Retracted Articles May Be a Deep Problem for Google

Why might the tech giant Google, with its huge resources, be struggling with the rather fundamental problem of correctly ranking retractions? One aspect is, of course, that Google Search does not have any deeper insight into the actual content of webpages. The algorithm behind it does not understand, at a deeper level, that the BBC’s article is about a scientific article that was later declared fraudulent, failing to semantically connect the one to the other.

A second aspect is that Google, at its core, bases its ranking on popularity with users and web administrators. This is captured not only in the number of clicks but also, characteristically, by Google’s foundational PageRank algorithm, according to which, roughly, more popular content is ranked higher than less popular content.<sup>1</sup> PageRank implements this idea at the level of links (sometimes called “hyperlinks”), so that webpages that receive many incoming links from other webpages are more highly ranked. Moreover, a link from a webpage that itself has

<sup>1</sup> For this aspect of Google, see Surowiecki (2004, p. 16), and Thelwall (2013, p. 77).

many incoming links has more weight than a link coming from any old webpage. By these criteria, the BBC's website will receive a very high ranking in Google: many link to it and many of the links come from other often-linked-to websites, such as other media outlets. Based on these considerations, one would expect an article in the BBC, such as the fish report alluded to in the introduction, to be practically "glued" to the first result page in Google, even after the material it is about was retracted.

However, matters are, in fact, much more complex. As the Google founders noticed in their seminal technical report, the ability of PageRank to track "importance" or "relevance" online depends on web administrators in a sense knowing what they are doing; they have to be capable of distinguishing relevant from irrelevant content and decide to include links to other webpages on the basis of that capacity. Recent studies have confirmed that PageRank tracks desirable properties of webpages, such as importance or relevance, under some rather austere conditions (Masterton et al., 2016, Masterton & Olsson, 2018a). The less austere conditions identified in Masterton & Olsson (2018b) have so far been shown to be realistic only in a minimal sense.

It is not surprising, then, that since PageRank was introduced Google has added some reportedly 300+ "quality signals" to fine-tune its search results. These other signals are subject to change and their precise nature is a trade secret not revealed by Google. However, advice on the Google homepage for how to write highly ranked content<sup>2</sup> suggests that many signals are there for detecting misspellings, grammatical failures, and other signs of low quality at the textual level. Still other signals are used for detecting the position of the search term within the text, so that a text in which the search term figures early on in the article is considered more relevant and ranked higher *ceteris paribus*.

Now, in the case of the information in *Forbes* about the retraction of the fish study, we may assume there are no essential differences between that article and the BBC coverage. Both were well-written, the search term "fish eat plastic" figures in prominent places, and so on. Hence, it is plausible to account for the higher ranking of the BBC coverage in relation to the *Forbes* retraction in terms of the former being more often linked to and generally more popular than the latter. People are simply more interested in reading and sharing an article on how fish eat plastic than in reading and sharing a follow-up on the retraction of that same article.

The reality here is that Google and other search engines that promote popular content may have hit the wall in the form of a proposed law of human psychology, which we might call the *Law of Retraction*: retractions are generally less interesting and therefore less popular with users and webpage administrators than the

<sup>2</sup> See <https://support.google.com/webmasters/answer/7451184?hl=en>.

original information retracted. There are certainly exceptions to this rule. In a highly polarized debate, for instance, the retraction may play into the hands of the other camp. If the latter has sufficient representation in the population, this may make the retraction popular. In a less polarized debate, or a debate in which the camp benefitting from a retraction is weak, we would expect the proposed law to hold. If this is true, Google and other search engines based on similar algorithms may have a fundamental problem that hinders them from effectively preventing fake science from being disseminated.

### 3. Methodology

The general principle we believe should be respected by search engines is the following:

**Criterion A:** True and non-misleading information about X should be ranked higher than false or seriously misleading information about X, *ceteris paribus*.

Based on our reasoning so far, we would expect Google to violate Criterion A by violating the following, more specific principle:

**Criterion A\*:** A true and non-misleading report to the effect that a particular scientific article has been retracted should be ranked higher than the retracted article itself or a now false or seriously misleading coverage of the latter, *ceteris paribus*.

We cannot test directly whether Google satisfies Criterion A\* or not, as it stands. For one thing, what should be the criterion for “satisfaction” here? What is the precise criterion for being a report to the effect that a given article has been retracted? For a report being “true” or “false”? What search terms should be used?

In operationalizing some of these issues, we rely on the Retraction Watch public database of retracted articles (<https://retractionwatch.com/>). Retraction Watch is considered to be a respectable source on retractions and is reportedly collaborating with the prestigious *Science* journal on retractions (Brainard & You, 2018). We assume, therefore, that the information in the database is correct regarding what has been retracted, when the retraction took place and for what reason. Furthermore, our study was restricted to retracted articles satisfying the following criteria:

**B1.** The article reported original research in a peer-reviewed scientific journal according to the standard editorial procedure for that journal.

**B2.** The article has identifiable dates for both its initial publication and its retraction notice.

**B3.** The article has been retracted for reasons that amount to fraud (fabrication of data).

Criterion B1 excludes invited publications falling out of the regular peer-reviewing process (invited contributions that summarize earlier work, editorial mistakes) as well as serious, but non-peer reviewed, publications, such as articles published by the original author of a peer-reviewed article in a non-peer reviewed journal (such as *Scientific American*).

Criterion B2 excludes articles for which either the date of its original publication or the date of the retraction is not identifiable. For instance, according to the Retraction Watch website, some publishers merely overwrite the original article's HTML page with the retraction notice. In such cases, the entry will have matching dates for both the original article and for the retraction notice, making the dates of publication and retraction impossible to separate. Criterion B2 also excludes notices that are weaker than retraction, such as corrections and expressions of concern, which do not always result into later retractions. Criterion B3 excludes retractions that occur for reasons other than fraud, i.e. honest mistakes in data handling or issues with reproducibility.

In order to keep the sample manageable for a pilot study, we restricted our search to all articles retracted between October and December 2018. Our reasoning for not including articles that were retracted later was that a shorter time span between retraction and the time of our study (November 2019) might not leave sufficient time for search algorithms to learn that the article had been retracted and adjust the search results accordingly.

When considering articles for inclusion, we first controlled for compliance with Criterion B3 and selected articles retracted for fabrication of data, which is a clear-cut case of academic fraud.<sup>3</sup> The principles for sample selection yielded twenty-four articles, all of them having distinct dates for publication and retraction, thereby complying with Criterion B2. All selected articles had initially appeared in journals published by reputable, non-predatory scientific publishers (*Elsevier*, *Springer*, *Wiley*, among others) thereby satisfying Criterion B1. The result was, as we will see, quite striking and we do not expect that including more articles would substantially change our results and conclusions.

Searches were by (complete) title in both Google Search and Google Scholar at two locations (to avoid effects of IP geolocation): the university to which the

<sup>3</sup> Thus, in effect, invoking B3 means that we operationalize “false” in Criterion A\* as “retracted due to fabrication of data”. It is certainly possible that a study based on fabrication of data still reaches conclusions that happen to be true, but we conjectured that such cases would, in practice, be very few. For the point that retraction and fraud are not the same, see McIntyre (2019).

authors are affiliated, and the private home of the author carrying out the empirical study. All searches were carried out in the period 18–19 November 2019. In the following, we refer to search on complete titles as “title search”. Searches were done on distinct computers (to avoid identification through MAC addresses) and via anonymous browsing in Chrome. The first Google result page for each query was saved. Studies have shown that users rarely go beyond the first result page (Genot & Olsson, 2017).

By a “retraction link” we mean in the following a link of the form “Retracted: [Title]”, where the title is that of the article that has been retracted or a similar link beginning with “retracted” or “retraction notice”. By an (original) “article link” we mean a link of the form “[Title]” where the title is that of the retracted article. We interpreted Criterion A\* as stating that a retraction link should be ranked higher than the corresponding article link.

Based on the consideration and methodological choices above, we devised two main hypotheses to be tested:

**Hypothesis 1 (H1):** Google Search violates Criterion A\* *more* often than not.

**Hypothesis 2 (H2):** Google Scholar violates Criterion A\* *less* often than not.

As for H2, we conjectured that, given that Google Scholar is tailored specifically for ranking scientific works, there should be relatively few violations of Criterion A\*. In other words, we expected there to be comparatively few cases in which an article link would be more highly ranked than a retraction link. In particular, there would be fewer violations against Criterion A in Google Scholar than in Google Search:

**Hypothesis 3 (H3):** Google Search generally violates Criterion A\* more often than Google Scholar does.

In our Google Scholar study, we also investigated what this resource regarded as the “best result” to a title query—an article link or a retraction link. For the same reasons that we conjectured that H2 and H3 would be true, we also expected retraction links to be considered “best results” by Google Scholar.

#### 4. Results

In our study, we found that Google Search almost always ranked article links higher than the corresponding retraction link. This was so regardless of whether the searches were made from the university (22/24) or from the home of the author carrying out the study (21/24). In the cases in which the article link was

ranked higher than the retraction link, the retraction links were more often than not ranked second to the retracted article. However, multiple instances of article links were often listed first (Figure 10.1). In some cases, finding a retraction link required the manual operation of scrolling down the page on a small laptop screen in addition to running the search. Thus, the results from our test were completely consistent with our hypothesis that Google Search would violate Criterion A\* more often than not (H1). Figure 10.2 summarizes our findings.

However, H2 and H3 were *disconfirmed* by our study in both the university and home searches. Google Scholar, too, violated Criterion A\* more often than not. In fact, its performance was no better than that of Google Search in this regard. Our results are summarized in Figure 10.3.

Furthermore, in 21/24 cases at both the university and at home, Google Scholar first proposed a shortlist of “best result” which, in 12/21 cases failed to list a retraction link but instead often showed just an article link (Figure 10.4).

The retraction link appeared in extended search results upon selecting the option “show all results” in no more than half of those cases, and the retraction link was then often further down on the result page (Figure 10.5). Our results are summarized in Figure 10.6.

Overall, Google Scholar, apart from almost always violating Criterion A\*, failed in both the university and home searches to list retraction links among the “best

The screenshot shows a Google search results page. The search query is "Infusion of amino acid enriched solution hastens recovery from neuromuscular block caused by vecuronium". The results page includes the following details:

- Search bar:** Infusion of amino acid enriched solution hastens recovery from neuromuscular block caused by vecuronium - Google Search
- Header:** Google, Sign in, search icon, settings, tools
- Search filters:** All, Images, Shopping, News, Videos, More
- Result count:** About 6 450 results (0.42 seconds)
- First result:**
  - Title:** Infusion of amino acid enriched solution hastens recovery ...
  - URL:** <https://www.ncbi.nlm.nih.gov/pubmed>
  - Text:** by Y Saitoh - 2001 - Cited by 15 - Related articles
  - Text:** Br J Anaesth. 2001 Jun;86(6):814-21. Infusion of amino acid enriched solution hastens recovery from neuromuscular block caused by vecuronium. Saitoh Y(1) ...
- Second result:**
  - Title:** Infusion of amino acid enriched solution hastens recovery ...
  - URL:** <https://www.sciencedirect.com/science/article/pii>
  - Text:** by Y Saitoh - 2001 - Cited by 15 - Related articles
  - Text:** Infusion of amino acid enriched solution hastens recovery from neuromuscular block caused by vecuronium Y. Saitoh<sup>1</sup>, K. Kaneda<sup>2</sup>, Y. Tokunaga<sup>2</sup> and M. ...
- Third result:**
  - Title:** RETRACTED: Infusion of amino acid enriched solution ...
  - URL:** <https://bja.oxfordjournals.org/article/fulltext>
  - Text:** Retraction notice to 'Infusion of amino acid enriched solution hastens recovery from neuromuscular block caused by vecuronium' [Br J Anaesth 2001; 86: ...]
- Fourth result:**
  - Title:** Retraction notice to "Infusion of amino acid enriched solution ..."
  - URL:** <https://bja.oxfordjournals.org/article/fulltext>
  - Text:** by Y Saitoh - 2019
  - Text:** Nov 29, 2018 - Retraction notice to "Infusion of amino acid enriched solution hastens recovery from neuromuscular block caused by vecuronium" [Br J Anaesth ...]

[https://www.google.com/search?q=infusion+of+amino+acid+enriched+solution+hastens+recovery+from+neuromuscular+block+caused+by+vecuronium&q=infusion+of+amino+acid+... 1/3](https://www.google.com/search?q=infusion+of+amino+acid+enriched+solution+hastens+recovery+from+neuromuscular+block+caused+by+vecuronium&q=infusion+of+amino+acid+...)

**Figure 10.1** Example of how multiple article links are often ranked higher than retraction links in Google Search

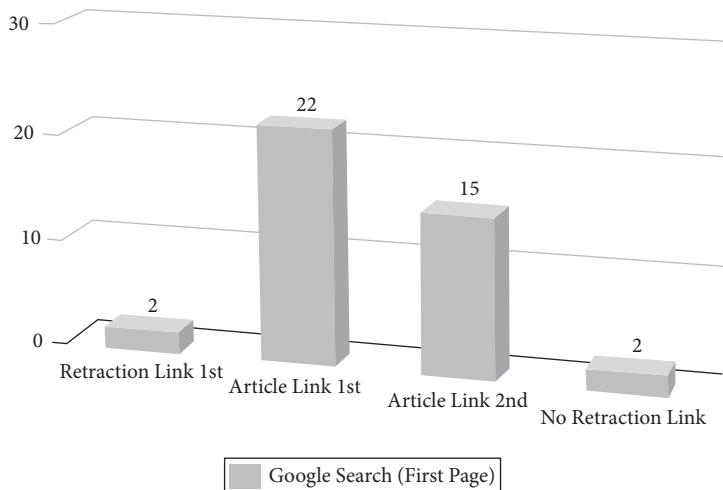


Figure 10.2 Summary of findings for Google Search (first result page)

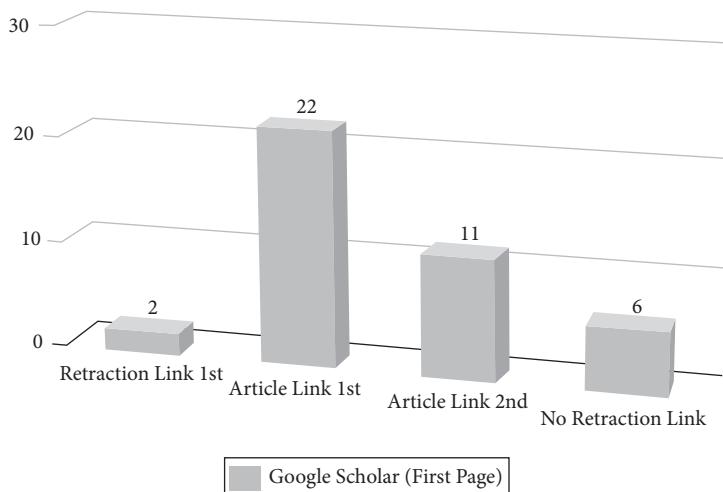


Figure 10.3 Summary of findings for Google Scholar (first result page)

results" half of the time and to list a retraction notice on the first page of search results one quarter of the time. Furthermore, in our sample, Google Search failed in two cases to include a retraction link at all on the first result page, and Google Scholar failed to do so in no less than five cases. In these cases, the Google user cannot tell that the article has been retracted just by looking at the first search result page.

19/11/2019

Infusion of amino acid enriched solution hastens... - Google Scholar

Infusion of amino acid enriched solution hastens recovery from neuromuscular block caused by vecuronium  
 Y Saitoh, K Kaneda, Y Tokunaga... - British journal of..., 2001 - academic.oup.com  
 We investigated the effect of an amino acid infusion on neuromuscular block produced by vecuronium, and on rectal temperature and surface temperature over the adductor pollicis muscle. Sixty adult patients undergoing general anaesthesia were randomly divided into four groups of 15 patients each: amino acid (AA)-post-tetanic count (PTC); AA-train-of-four (TOF); control (C)-PTC; or C-TOF group. In the AA-PTC and AA-TOF groups, after a bolus of vecuronium 0.1 mg kg<sup>-1</sup>, a continuous infusion of an 18 amino acid enriched solution ...  
 ☆ 99 Cited by 15 Related articles All 11 versions

Showing the best result for this search. See all results

**Figure 10.4** Example of an article link being proposed as “best result” for a title search in Google Scholar even though the article has been retracted

19/11/2019

Infusion of amino acid enriched solution hastens... - Google Scholar

Infusion of amino acid enriched solution hastens recovery from neuromuscular block caused by vecuronium  
 Y Saitoh, K Kaneda, Y Tokunaga... - British journal of..., 2001 - academic.oup.com  
 We investigated the effect of an amino acid infusion on neuromuscular block produced by vecuronium, and on rectal temperature and surface temperature over the adductor pollicis muscle. Sixty adult patients undergoing general anaesthesia were randomly divided into ...  
 ☆ 99 Cited by 15 Related articles All 11 versions

Role of amino acid infusion in delayed recovery from neuromuscular blockers  
 S Kalra, R Wadhwa - Indian journal of anaesthesia, 2010 -.ncbi.nlm.nih.gov  
 ... heat production by administration of amino acid infusion[9]. Amino acids infused during general ... prevent postoperative hypothermia[9]. Previous studies have found that IV amino acid infusions exert enhanced ... 7. Saitoh Y, Kaneda K, Tokunaga Y, Murakawa M. Infusion of amino ...  
 ☆ 99 Cited by 6 Related articles All 7 versions

Drugs to facilitate recovery of neuromuscular blockade and muscle strength  
 Y Saitoh - Journal of anaesthesia, 2005 - Springer  
 ... Söderén E, Brundin T, Wahren J (1984) Augmented thermic effect of amino acids under general anaesthesia: a mechanism useful for ... 2. After intravenous infusion ... RA, Leijerent TH, Likoff MJ (1987) Efficacy and safety of sustained (48 hour) intravenous infusions of milrinone in ...  
 ☆ 99 Cited by 9 Related articles All 8 versions

A Prospective Randomized Double Blind Study to Evaluate the Effect of Infusion of Amino Acid Enriched Solution on Recovery from Neuromuscular Blockade  
 N Gupta, R Sehgal, R Kumar, KR Sharma... - Indian journal of..., 2009 -.ncbi.nlm.nih.gov  
 ... Results. The four groups were similar in demographic profile, type and duration of surgery, ambient temperature and time of fluid infused (Table 1) ( $p>0.05$ ). Saitoh Y, Kaneda K, Tokunaga Y, Murakawa M. Infusion of amino acid enriched solution hastens the recovery ...  
 ☆ 99 Cited by 5 Related articles All 10 versions

Retraction notice to “Infusion of amino acid enriched solution hastens recovery from neuromuscular block caused by vecuronium”[Br J Anaesth 2001; 86: 814–821]  
 Y Saitoh, K Kaneda, Y Tokunaga... - British journal of..., 2019 - Elsevier  
 Statistical analysis suggests that the data may be fabricated. Y Saitoh provided a statement in a personal communication to a member of the editorial board of British Journal of Anaesthesia that the study was not approved by the Institutional Review Board and that no ...  
 ☆ 99 All 6 versions

[https://scholar.google.se/scholar?lookup=0&q=infusion+of+amino+acid+enriched+solution+hastens+recovery+from+neuromuscular+block+caused+by+vecuronium&hl=en&as\\_sd=0](https://scholar.google.se/scholar?lookup=0&q=infusion+of+amino+acid+enriched+solution+hastens+recovery+from+neuromuscular+block+caused+by+vecuronium&hl=en&as_sd=0)

1/2

**Figure 10.5** Example of a rather anonymous retraction link further down in the search result list in Google Scholar

## 5. Discussion

The fact that the first link in our study on Google Search and Google Scholar were almost always an article link and not a retraction link, even though the article had been retracted, is particularly disturbing in relation to what has been referred to as “top link heuristics” (Pan et al. 2007, Salmerón et al., 2013). Salmerón (2019, p. 1, references removed) writes:

Search engine results pages (SERPs) are a frequent gateway to Internet content. Prior research has extensively documented strong effects of SERPs (e.g. rank

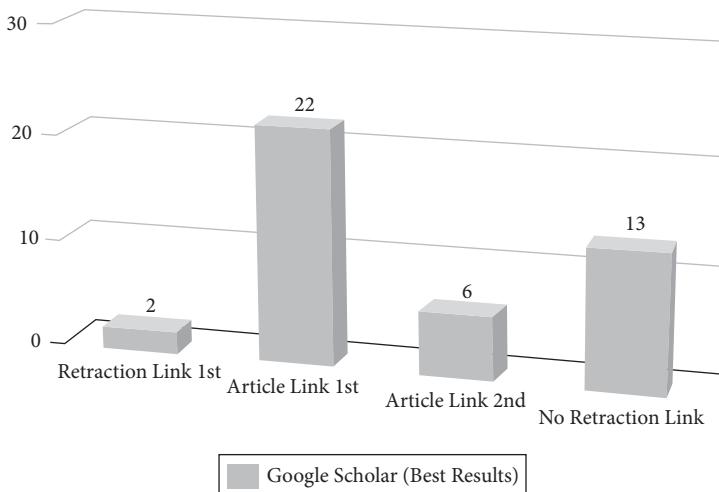


Figure 10.6 Summary of findings for Google Scholar (best results)

order or the spatial distribution of the results) on users' attention to and selection of particular Web pages. In the context of Web search, a common user behavior is the 'top link' or 'Google trust' heuristic, that is, the inspection and selection of only the first few search results presented by the search engine, without evaluating all other search results available. This heuristic behavior allows users to find information in an efficient way, as search engines tend to provide relevant documents on top of the list, especially when it comes to simple facts. But just relying on the top results of the SERP to access information may not be as efficient when users search for learning purposes about controversial topics, such as climate change, for at least two reasons. First, users can be easily misled by, for example, commercially biased Web Pages located on top of the SERP. Second, by looking at just few hits users miss the opportunity to use SERP information to reflect on the relationships between available web pages, an essential step when learning about controversial topics.

Our study indicates that the top link heuristic may not be "efficient" even when it comes to some "simple facts", such as the indubitable fact that a scientific article has been retracted.

Effectively, rather than dealing with retractions itself, Google "outsources" the problem, counting on scientific journals to include a retraction notice on the webpages of retracted articles. In our sample, many journals had indeed included a retraction notice on the webpage of the retracted article, so that when users click on the article link in Google Search, they are directed to a page with a visible retraction notice. Unfortunately, it has been observed that "[n]ot all publishers, for

instance, publicize or clearly label papers they have retracted, or explain why they did so" (Brainard & You, 2018).<sup>4</sup>

This practice of outsourcing supports our earlier conjecture that retracted articles are a deep problem for Google and similar search engines relying on popularity metrics. After all, had there been a quick fix, Google (or Alphabet, as the company is now called) would have found it and solved the problem inside the search engine itself rather than relying on journals to visibly include information about retractions.

Our study raises a number of questions for future work. The overarching question is of course whether actual users of Google are in fact misled by the way Google handles retractions due to fraud. Do they, as a result, come to entertain false scientific beliefs? Our results indicate that Google almost always ranks an article link higher than a retraction link. In order to investigate whether users are actually misled by Google's practice, it would probably be necessary to set up an empirical experiment with real subjects and study their search behaviour and beliefs before and after search. To the best of our knowledge, no such experiment has yet been carried out. A factor that might be of interest here is the relative proportion of links to the original article and retraction links on the first Google page. Conceivably, a majority of retraction links in relation to article links might convince a user that the article has been retracted and vice versa.

A further, pressing issue for future study concerns identifying the best policy for handling retracted articles in order to prevent fake science from being disseminated through search engines such as Google. One approach is to leave search engines more or less as they are, including their ultimate reliance on popularity as a significant determinant of the ranking of search results, and to outsource the task of informing about retractions to the scientific journals. As we have seen, this is essentially how Google at the time of writing solves this problem. A difficulty with this approach is that links to retracted articles receive unwanted attention in the list of search results in a way that could potentially mislead the user. A further issue is that the strategy assumes that journals generally take their responsibility vis-à-vis retractions, which, again, is not always the case. A second kind of solution would be to solve the problem within the search engine itself, i.e. to develop search algorithms that have an "organic" preference for retraction links.

<sup>4</sup> Cf. Teixeira da Silva & Bornemann-Cimenti (2017, p. 365): "COPE guidelines state that the 'retracted status should be indicated as clearly as possible', but this is definitely not true for many retracted publications. Likewise, databases do not consequently link retracted articles with the notice of retraction. Furthermore, many papers are deposited in the 'original', i.e. pre-retraction version on personal or institutional websites or online repositories. Similarly, printed 'stock files' are obviously unaffected by a retraction. Clear identification of a retracted article using a watermark and in databases is a crucial step while incorporation of an electronic 'retraction check' in reference management software and during the online submission is necessary to detect and avoid citing retracted literature. Solving this problem needs the close attention of everybody involved in the publishing process: authors, reviewers, and publishers."

However, search engine providers will be unwilling to make radical changes to a technology that has served them so well. Thus, it is unlikely that they will consider moving away from the underlying “populist” technology. A less radical proposal for algorithmic change would be for Google and others to store retracted information in a special database, much like Retraction Watch does, and tweak their algorithms so that the fact that information has been retracted overrides other considerations, such as popularity. Of course, these two approaches are not mutually exclusive: it is likely that both search engine providers and scientific journals will have to take greater responsibility for how they handle retracted studies.

In our study, the search term constituted the (complete) title of the article. Initial tests suggest that searching on key words in the title rather than on the title itself can give different results vis-à-vis the ranking of article links versus retraction links. In one case, a retraction link was ranked second when searching on the title but ranked somewhere in the middle when searching on keywords. Other possible options would be combinations of keywords attached to the abstract of the article, when available, or titles of popular science or generalist media articles referencing the retracted article, when available. As for the latter, anecdotal evidence suggests that article titles are seldom included verbatim in popular science or generalist media coverage, but that the articles are usually linked to. Thus, a non-specialist reader following the link could have access to the article title as published, and later search for it. Alternatively, titles and taglines of popular science and generalist media coverage are likely the first exposure of non-specialists readers to the content of retracted article, and are thus a natural choice for online searches.

In cases of popular science or non-specialist accounts of articles that were later retracted, we may ask how often the outlet in question includes a retraction notice after the retraction. If this does not happen on a regular basis, would it be useful to have a service that informed the outlet that content on its website relies on a retracted scientific article?

## 6. Conclusion

Our initial case study showed that fake science, in the sense of articles that have been retracted due to fraud, can be visible in Google, and thus potentially disseminated via the search engine, even after the articles has been retracted. We hypothesized that the reason for this lies in the popularity-based logic governing Google, in particular its foundational PageRank algorithm, in conjunction with a psychological law which we refer to as the “law of retraction”: a retraction is typically (though not universally) taken to be less interesting and therefore less popular with internet users than the original content retracted. We performed a pilot study of the relative ranking of retractions due to fraud (fabrication of data) drawing on records of retracted articles in the Retraction

Watch public database. The study tested, specifically, the extent to which article links, i.e. links with the title of the retracted article, are still highly ranked in Google—and more so than retraction links, i.e. links starting with “retraction” or “retraction notice” followed by the title of the retracted paper.

In our sample, Google Search almost always ranked article links higher than retraction links, apparently judging article links to be more important or relevant than retraction links. This was in conformity with one of our main hypotheses (H1). We thought that the problem would be less severe in Google Scholar (H2, H3). However, this turned out to be false: Google Scholar, too, consistently ranked article links higher than retraction links. Moreover, Google Scholar often failed to list retraction links among the “best results” for a title search, and in five cases failed to include a retraction link at all on its first result page!

The study reported here is a small-scale pilot study comprising twenty-four retracted articles from a given period in time, and one should be careful about drawing general conclusions. In particular, the results should not be taken as proof that users will generally be misled by the way Google Search and Google Scholar handle retractions due to fraud. Even so, the results are quite striking and it would be surprising if including more retracted articles would yield a radically different result vis-à-vis our hypotheses, unless, of course, Google chooses to update its current search algorithms in decisive ways to deal with the problem.

To clarify, we do not claim that Google is guilty of producing fake science, of course; those responsible are the scientists in question. We do think, however, that our study indicates that Google risks disseminating fake science through its ranking algorithms. Finally, our study supports our initial conjecture that it is not easy to handle the problem of retracted articles within Google itself. Rather than tweaking its algorithms to make retraction links more highly ranked than the corresponding article link, Google has in effect “outsourced” the handling of retractions to the scientific journals issuing them, counting on these journals to include retraction notices on the articles’ webpages, which, as we saw, is an ideal yet to be universally implemented.<sup>5</sup>

## References

Brainard, J. & You, J. (2018), “What a massive database of retracted papers reveals about science publishing’s ‘death penalty’”, *Science*: <https://www.sciencemag.org/news/2018/10/what-massive-database-retracted-papers-reveals-about-science-publishing-s-death-penalty>.

<sup>5</sup> The authors’ work on this chapter was financed by the Swedish Foundation for Humanities and Social Sciences (Riksbankens Jubileumsfond) within the project Filter Bubbles and Ideological Segregation Online (P18-0656:1).

Genot, E. & Olsson, E. J. (2017), "Do We Trust Blindly on the Web?" *Iride: Journal of Philosophy and Public Debate* 30: 87–105.

Koffmar, L. (2017), "Universitet konstaterar oredlighet bakom plaststudie", Uppsala University, 7 December: <https://www.uu.se/nyheter-press/nyheter/artikel/?id=9818&typ=artikel>.

Lönnstedt, O. M., & Eklöv, P. (2016), "Environmentally Relevant Concentrations of Microplastic Particles Influence Larval Fish Ecology", *Science* 352(6290): 1213–16. Retracted 25 May 2017.

Maheshwari, S. (2016), "How fake news goes viral: a case study", *New York Times*, 20 November: <https://www.nytimes.com/2016/11/20/business/media/how-fake-news-spreads.html>.

Masterton, G., Olsson, E. J., & Angere, S. (2016), "Linking as Voting: How the Condorcet Jury Theorem in Political Science is Relevant to Webometrics", *Scientometrics* 106(3): 945–966.

Masterton, G. & Olsson, E. J. (2018a), "From Impact to Importance: The Current State of the Wisdom of Crowds Justification of Link-based Ranking Algorithms", *Philosophy and Technology* 31: 593–609.

Masterton, G. & Olsson, E. J. (2018b), "PageRank's Ability to Track Webpage Quality: Reconciling Google's Wisdom-of-Crowds Justification with the Scale-free Structure of the Web", *Heliyon* 4: <https://www.sciencedirect.com/science/article/pii/S2405844018340738>.

McGrath, M. (2016), "Fish eat plastic like teens eat fast food, researchers say", *Science and Environment*, BBC 2, June: <https://www.bbc.com/news/science-environment-36435288>.

McIntyre, L. (2018), *Post-Truth*, Cambridge, MA: MIT Press.

McIntyre, L. (2019), *The Scientific Attitude: Defending Science from Denial, Fraud, and Pseudoscience*, Cambridge, MA: MIT Press.

Morris, D. Z. (2017), "Scientific Journal Retracts Controversial Paper Claiming That Fish Prefer Plastic Over Real Food", *Fortune*, 7 May: <https://fortune.com/2017/05/07/science-retraction-plastic-fish/>.

Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., & Granka, L. (2007), "In Google We Trust: Users' Decisions on Rank, Position, and Relevance", *Journal of Computer-Mediated Communication* 12(3): 801–23, <http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00351.x/full>.

Salmerón, L., Kammerer, Y., & García-Carrión, P. (2013), "Searching the Web for Conflicting Topics: Page and User Factors", *Computers in Human Behavior* 29: 2161–71.

Salmerón, L. (2019), "Search Interfaces and Learning about Controversial Topics", abstract, SALMM '19, 21 October, Nice, France: <https://dl.acm.org/citation.cfm?id=3356729>.

Surowiecki, J. (2004), *The Wisdom of Crowds: Why the Many Are Smarter than the Few and How Collective Wisdom Shapes Business, Economies, Societies, and Nations*, London: Little Brown.

Sydell, L. (2016), "We Tracked Down a Fake-News Creator in the Suburbs. Here's What We Learned", National Public Radio, 23 November: <https://www.npr.org/sections/alltechconsidered/2016/11/23/503146770/npr-finds-the-head-of-a-covert-fake-news-operation-in-the-suburbs?t=1604766393716>.

Teixeira da Silva, J. A. & Bornemann-Cimenti, H. (2017), "Why Do Some Retracted Papers Continue to Be Cited?", *Scientometrics* 110(1): 365–70.

Thelwall, M. (2013), "Society on the Web", in W. H. Dutton (ed.), *The Oxford Handbook of Internet Studies* (pp. 69–85), Oxford: Oxford University Press.

# The Virtue of Epistemic Trustworthiness and Re-Posting on Social Media

*Sarah Wright*

## 1. Re-Posting and Its Epistemic Risks

While there are many worrying aspects of the generation and spread of fake news, I will focus on a feature of its role in our contemporary epistemic lives that I believe calls for new applications of epistemic virtues. I assume that readers of this chapter are not producers of fake news, but many of us run the risk of spreading it. This risk has become more pronounced with the development of social media and the changing role that it plays in the information landscape. Almost whenever one is reading an article online these days, one is confronted with an option to share that article on social media; often this option is encoded directly into the original site publishing the article. Re-posting is as easy as pressing a button at the top of the page. Not only is re-posting easy to do, it is also becoming a major source of information for others. Rather than receiving information directly from newspapers, magazines, or television, we now receive much of our news from other individuals through links on social media. A 2017 Pew Report found that 67 percent of U.S. adults get some news through social media; the same report found that this number is increasing year-on-year.<sup>1</sup>

Re-posting on social media exposes its audience to two different risks of epistemic harm. Testimony in general exposes the recipient to the risk of false belief. When we trust the word of others, there is always a risk that they have a false belief they will pass on to us. This is a classical concern about testimony and can easily be captured by veritistic accounts of our epistemic practices; veritism recognizes the value of true belief and the harm of false belief, and looks to social systems that maximize this epistemic value for the participants.<sup>2</sup>

To understand the second kind of epistemic harm re-posting exposes recipients to, we need to consider how re-posting is a kind of credentialing. Ordinarily, when I cite someone else's writing I endorse not only the content but also the source.

<sup>1</sup> Shearer and Gottfried 2017.

<sup>2</sup> The veritistic approach to evaluating social epistemic practices was introduced by Alvin Goldman in his 1999.

When I respond to a query by saying “Well, I read P in the *New York Times*,” I imply both that I believe P and that I trust the *New York Times* as a source. We can think of this second aspect as a kind of credentialing. By citing a source I am often endorsing it as a source. This implication depends on conversational norms of relevance; if I don’t trust the source, why bring it up in conversation? Similar considerations apply not only in conversation but also in re-posting. If you pose a question on social media and I respond with a link to an outside article, my reposting can be taken as both an endorsement of the content (“Here is an answer to your question”) and as credentialing the source (“Here is a source where you could find answers to similar questions”).

Credentialing a source through reposting articles from it exposes audiences to a different kind of risk—the risk of misguided trust in a source. The epistemic harm of misguided trust includes, but also goes beyond, the harm of being led into false belief. Misguided trust in an unreliable source clearly exposes an audience to a risk of developing many false beliefs. It is interesting to note that re-posting, more than just conversational citing, puts the reader in touch with the source. By citing a source and linking to it, I invite your trust in that source, and give you an easy way to get more information from that source. Your contact with and trust in an unreliable source will tend to lead you to form further false beliefs.

But beyond the risk of developing further false beliefs, misguided trust can be seen as kind of epistemic harm in itself. Not only does it contain the false belief that the source is trustworthy, it also represents a misunderstanding about one’s epistemic position in the world. We can think of the misunderstanding as going deeper than the mere formation of false beliefs. Understanding one’s position in the world is central to one’s self-conception and is required in order to skillfully navigate the world in any way. If we see understanding as an epistemic good distinct from the other epistemic goods of true belief or knowledge, we should recognize that a misunderstanding about who to trust keeps one from a full understanding of one’s epistemic location.

## 2. Apparent News Sources as Sources of False Belief and as Objects of Misguided Trust

If we recognize misguided trust as a distinctive kind epistemic harm, we can explore similarities between sources that can be its target. This could be individuals credentialing each other; I can introduce you to someone as my friend, and this can lead you to trust them when you should not. They could be deceptive or simply incompetent. Misguided trust can happen in practical or epistemic cases.

However, the category that I find most relevant to the activity of re-posting (and to this volume) is that of sources which in some way resemble news sources. When articles are presented in a particular format, using specific forms of

language and presentation, they can easily be taken as news sources, and they can be trusted on that basis. This can lead to misguided trust, whether the source intended to deceive its audience or is simply slipshod about the production of articles with insufficient regard for their truth. I want to explore three categories of these apparent news sources here. They do not all meet existing definitions of fake news, but I think that they are helpfully addressed together because of the ways that they expose their audiences to misguided trust in a source, even when their content happens to be true.

First, and perhaps most visible, is politically motivated fake news. By now we are all familiar with the way that the Internet Research Agency employed thousands of Russians to write fake pro-Putin news stories about Ukraine, as well as mostly pro-Trump fake news stories in the run-up to the 2016 U.S. election. Those employed in this project reported being asked to write these articles and re-post them through fake accounts at a breakneck speed. Particularly on the issues involving Russia, those employed were required to write a set number of pro-Russia articles a day, as well as to re-post and comment on those of other employees.<sup>3</sup> These articles reached a large audience and may have been sufficient to change the outcome of the 2016 election.

Jonathan Freedland and his guests on the BBC radio show “The Long View” have argued that, although the format has changed, this kind of fake news has been happening for centuries.<sup>4</sup> This radio show looks for connections between contemporary news stories and historical events, and here they focused on the story of the murder of William of Norwich in 1144. While there was no evidence that this murder was committed by members of the Jewish community, this false “blood libel” story may have been motivated by envy of the political position of those communities. What makes this a strong parallel with contemporary fake news was the publishing by Thomas of Monmouth of a multi-volume work, *The Life and Miracles of St. William of Norwich*, in which he asserted that William was murdered by Jews, and cited evidence of miracles performed at his grave site as evidence of his being martyred. This story became canonical (literally) and it and stories like it led to violence against Jewish communities through the Middle Ages. It may seem odd to characterize Thomas of Monmouth as a purveyor of fake “news” before the regular publication of news. But his is clearly a narrative that an audience is encouraged to believe through its publication from an apparently authoritative source. (Monmouth being a Benedictine monk in the local monastery.) This example shows that trusting apparently authoritative reports could lead to misguided trust even before the invention of news (and hence of fake news).

<sup>3</sup> MacFarquhar (2018) gives details of these operations from individuals who were employed in them.

<sup>4</sup> Freedland 2017.

Second, we might consider economically motivated apparent news sources. People around the world can profit from placing Google ads on their websites and by directing enough traffic through those sites. This was the strategy of a group of teenagers from Veles Macedonia in the run-up to the 2016 U.S. election.<sup>5</sup> They took their profit model from a pair of older men in the community who has made a substantial profit from websites offering health advice. These websites published whatever was profitable, without any regard to whether it was true—e.g., one article suggested sleeping with a bar of soap under one's sheets as a cure for leg cramps. The teenagers experimented with different topics and discovered that posting pro-Trump articles was the most profitable. One teenager reported earning \$16,000 in four months, approximately ten times the average salary in Macedonia. These teenagers were not politically motivated; they were simply responding to a pure profit motive. In fact, some of these teenagers reported being worried that they may have affected the outcome of the U.S. election.

Tim Wu (2016) has argued that the profit motive for generating apparent news stories has been around at least as long as the model of supporting publishing through ads. In 1833, Benjamin Day shook up the newspaper publishing industry by starting *The Sun*, a New York paper that sold for one cent, undercutting the market rate of six cents. Wu points to this as the first instance of what he calls “selling attention”; the product was not the paper itself, but rather the attention of the readers who were exposed to its advertising. This economic model incentivized *The Sun* to publish whatever would sell papers, including stories of the most gruesome and sensational crimes of the day. Famously, this led them to publish a series of stories about discoveries on the moon of a new human-like species they called moon bats (*Vespertilio-homo*). While this advertising model didn’t involve any clicks, this article would clearly be classified as “click-bait” today. The economic motivation to produce salable stories that look like news does not require deception. It does not even require that the stories produced are false. While the moon-bats story is transparently false, many of *The Sun*’s reports on sensational crimes may have contained many truths. As a result, these economically motivated sources do not seem to meet the criteria that have been offered in definitions of fake news. Regina Rini and Axel Gelfert both require fake news to have false content and to be produced with an intent to deceive.<sup>6</sup>

But though they may not be produced with an intent to deceive, when they are trusted by a reader, economically motivated articles can have the same two epistemic harms as fake news. The first is in leading the reader to believe a falsehood, and the second is in generating misguided trust in the reader. Neither of these harms depends on deceptive intent, and the second (though not the first) can occur even when the article’s content is true. Trusting a source that publishes

<sup>5</sup> See Subramanian 2017.

<sup>6</sup> Rini 2017 and Gelfert 2018.

only out of a desire to sell advertising, and not at all from a concern with the truth, is a clear case of misguided trust. We might think of this category of apparent news sources as a close parallel to “bullshit,” in Frankfurt’s sense, with the added caveat that it is presented in a format that causes others to classify it as news.<sup>7</sup> Truth is simply irrelevant to the production of click-bait.

Finally, we should consider a kind of apparent news source that is not clearly motivated by political or economic gain but rather comes from fiction or comedy being mistaken for news. There are many sites, like the *Onion* or *The Babylon Bee*, presenting satire which mimics the form of news stories. When we go to the sites directly, there are often cues in the masthead or context which indicate that the stories are intended as entertainment; but in a social media feed, stories are ripped from that context. Often only the headline and an image are visible in the post.

This sort of mistake can be generated by any fiction that takes the form of news. It has a very close analog to the way that listeners were fooled by the *War of the Worlds* radio broadcast in 1938. Conflicting programming led many listeners to tune into the program after its introduction as a dramatic presentation by The Mercury Theatre on the Air. This radio play sounded like a series of news bulletins interrupting a music performance, and those news bulletins were taken as real by many in the listening audience, apparently causing some to panic. Although recent investigations have undermined the most implausible claims about extreme mass panic caused by the *War of the Worlds* broadcast, it is still plausible that some individuals were fooled by this show, at least in the short term.<sup>8</sup>

While satirical news websites and the *War of the Worlds* radio broadcast are both examples in which the audience might be confused and deceived, the intention of the producers is simply to entertain. Using the format of a news broadcast is a helpful narrative device towards that end. But using that narrative device opens up the possibility of the fiction being misunderstood by its audience. When it is misunderstood, we have an audience who is receiving what they take to be news, which was produced with a goal other than that of giving information. If the audience trusts this as a news source, their trust is misguided.

### 3. Re-Posting and the Further Risks of “Bent Credentialing”

Re-posting on social media is a very easy way of spreading information that we have not produced, and, as noted above, it is a growing source from which people report getting their news. In the historical examples above, it was possible to do

<sup>7</sup> Frankfurt 2005.

<sup>8</sup> Memmott 2013. Interestingly there is also speculation that the over-reporting of panic caused by this radio show might have been motivated by the desire of newspaper editors to undermine the credibility of radio, which would make the reporting on the *War of the Worlds* a kind of economically motivated false reporting.

something like re-posting, and in the process to credential the source being cited. A medieval merchant could pass on a copy of Thomas of Monmouth's work to a friend, encouraging them to read it and, "get the real story." An early nineteenth-century New Yorker could give their copy of *The Sun* to a neighbor, encouraging them to read about moon bats. So the activity of re-posting has historical analogs. However, the ease with which we can re-post and the number of people who report getting their news from re-posting, shows us that it is becoming a more central element of our epistemic environment.

In addition, re-posting is an epistemic act which has not (yet) developed clear communication norms. Regina Rini (2017) has noted this and she calls re-posting a "bent kind of testimony." Re-posting displays the article for others to see without clearly endorsing the content. Ordinary cases of testimony often pass on information from other sources, but the format of ordinary testimony makes clear the attitude of the testifier. I can quote someone else to critique them. But, when you are asking for information and I respond by telling you what someone else said, this testimonial act is a kind of endorsement. And, at least on the assurance view of testimony, one is responsible for their testimonial assertions passing on the views of others. If those views turn out to be false, the hearer can criticize the testifier for passing them on. Re-posting, on the other hand, can convey belief in the content of an article, or it can hold that article or its content up for ridicule. Re-posting without comment does not disambiguate the purpose of sharing. If the content of the re-posted article turns out to be false, the re-poster can deny responsibility. As Rini notes, there is a common reply, "a retweet is not an endorsement." Thus the speech act of re-posting is in some ways like testimony, but in other ways it is ambiguous. This is why Rini calls it bent testimony.

Bent testimony is an intriguing feature of our current practices of re-posting. I would like to take Rini's insight one step further to consider another aspect of the speech act that is bent. If re-posting can cause one's commitment to the content of the post to be bent, it can also bend one's commitment to the credibility of the source. Call this "bent credentialing." A standard case of citing the work of others is, particularly in response to a question, one in which the source and the content are endorsed. Of course, it is also possible, and even common, to cite someone's work as a way to criticize it. Either intention is generally signaled in the text of a work making a citation or the context of a conversation in which a source is sited. But re-posting allows one to cite and link to an article without either signaling agreement or disagreement. The link can simply be presented without any context to disambiguate the re-poster's attitude toward the source cited.

Regardless of the poster's intention, re-posting puts audience members in touch with a source, and it provides the audience with at least some evidence that the re-poster trusts the site. These implications of the speech act of re-posting can be observed through empirical studies about the ways that audiences interact with reposting on social media. When the re-poster is trusted by the audience as

knowledgeable on the subject, this increases the audiences trust in the content of the article and also leads them to seek out more information from the linked source.<sup>9</sup> Re-posting runs the risk not only of leading's one's audience into misguided mistrust, but also of doing so contrary to the intentions of the person re-posting. Whether earnest or ironic, re-posting from an unreliable source can lead others into misguided trust of the original source.

#### 4. Why Do the Risks of Re-Posting Call for an Epistemic Virtue to Address them?

So far we have explored the epistemic risks that re-posting brings with it, noting that this includes not only the risk of leading others to a false belief but also a distinct risk of leading others to a state of misguided trust. This trust is encouraged through the credentialing than can be implied or explicitly stated in the act of re-posting. Since re-posting is a speech act with evolving norms, and since it is possible to repost without providing any disambiguating context, some of the credentialing that leads to misguided trust may unintentional or “bent.” All these epistemic risks should lead us to be careful in our re-posting behavior. I propose that a good way to do this is through looking to epistemic virtues; virtues are intended to help us navigate through complicated situations, and we have just explored how complicated the risks of re-posting are. For those who are doubtful, I suggest three further reasons to adopt a search for relevant epistemic virtues.

First, the epistemic good at risk in cases of misguided trust is understanding, which is an epistemic good that plays a central role in many virtue epistemologies, particularly those that aim to model their epistemic virtues on traditional moral virtues.<sup>10</sup> Just as moral virtues aim not just at correct individual acts, but at good moral character, epistemic virtues aim not just at the production of true beliefs, but at the development of good epistemic character traits. Understanding is a natural target here, as it cannot be reduced to individual true beliefs. Rather, understanding requires interrelations between complex sets of beliefs and ways of seeing the world. It is closer to the virtue ethical goal of a good life than to the punctate goal of correct individual acts.

Second, the epistemic risks associated with re-posting stories highlight the fact that we have taken on a new epistemic role. Rather than simply being consumers of purported news stories, we are also transmitters of those same stories. This is a role not unlike that of being an editor in a traditional newsroom; we choose what to publish on our own social media pages. What are the demands of this new role? Traditional virtue ethics, going back to its ancient Greek roots, has recognized the

<sup>9</sup> Turcotte et al. 2015.

<sup>10</sup> This is a model of responsibilist epistemic virtues established by Linda Zagzebski in her 1996.

importance of social roles in understanding how the virtues are manifested in individual lives.<sup>11</sup> So, virtue epistemology, and particularly a version of virtue epistemology that focuses on our robust epistemic character traits, may have an answer to this question, which now takes a new form. What epistemic virtues as called for by our new role as transmitters of purported news stories? Our epistemic role as re-posters generates new ways that others depend on us. We should recognize that our actions in re-posting can expose others to epistemic risks; yet, others depend on us not to harm them in this way. In addition to not misinforming others, we should also take care not to generate misguided trust in others.

Finally, the harm of misguided trust falls under the more general category of misunderstanding our own epistemic position with respect to others, which can also include misguided *mistrust*. It is interesting to note the testimonial injustice that Miranda Fricker has developed and explored is a kind of misguided *mistrust*.<sup>12</sup> When a hearer assigns too little credibility to a speaker based on her racial, gender, and other social prejudices, that hearer will fail to take a speaker's testimony as seriously as she ought—she will fail to trust someone who is trustworthy.

Fricker's proposed solution to the problem of testimonial injustice is developing the virtue of testimonial justice. She argues that the antidote to testimonial injustice must be a virtue because the credibility deficits that lead to this injustice come from the way that we perceive others. Our credibility judgments about others are rarely explicitly argued for; rather we most often simply perceive others as authoritative or hesitant, trustworthy or sketchy. Since these prejudices will often be implicit, Fricker argues that we need a well-developed virtue of testimonial justice to identify and to correct for them. The epistemic virtue of testimonial justice helps us improve the way that we see the world; this is intended to be parallel to the way that moral virtues change our perception of others and hence change our motivations and actions. Just as with Fricker's testimonial injustice, misguided trust has to do with a biased perception of others. Testimonial injustice has to do with seeing others as lacking in credibility. Misguided trust comes from seeing a source as more credible than it is. Correcting for either requires a change in our ways of perceiving the world, which is a central aim of virtues.

<sup>11</sup> See, for example, Epictetus (1995), *Discourses* II, 10, 10–11: “If furthermore, you are on the council of any city, you should remember that you are a councilor, if a youth, a youth; if an old man an old man. For each of these names, if rightly considered, always point to the acts appropriate to you.”

<sup>12</sup> Fricker 2007.

## 5. Epistemic Virtues Relevant to Re-Posting

One virtue that might seem obviously relevant to the activity of re-posting is that of trustworthiness. If others are trusting us, and putting their trust in sources on that basis, we ought to strive to be worthy of that trust. Karen Jones (2012) explores how trustworthiness can be generated through our recognition that others depend on us. To be trustworthy is not only to be competent in a particular area, but also to take the dependence of the trusting person as a compelling reason to act on that competence. Thus, Jones grounds her account of trustworthiness on what she takes to be basic elements of the human condition. We are social and finite, and so depend on others. She is considering these limitations in the general realm of action, but they apply equally well in the realm of the epistemic. A classic motivation for moving to a social epistemology is the fact that we are very limited in what we can know on our own. In order to satisfy our epistemic ends we must depend on others. But this dependence opens us up to risk. To this, Jones adds that we are reflective creatures. Not only can I depend on you, you can recognize my dependence on you, and I can recognize that you recognize my dependence. You can also see my dependence as a compelling reason to act; when you do so and you are competent to support my dependence, you are acting in a trustworthy manner. Jones' account of trust can be practical or epistemic. But the target here is epistemic, and so we should look for a model of epistemic virtues that are relevant.

It will be useful to make a distinction between the kinds of virtues that are often the target of virtue epistemology and the epistemic virtues relevant here. Most of the literature on epistemic virtues tends to focus on *self-regarding virtues*, or those that help the individual in pursuing her own epistemic ends. For example, arguments that open-mindedness is a virtue generally focus on whether practicing this virtue will lead its possessor to better epistemic outcomes, whether those are truth, knowledge, or understanding. Epistemic virtues aim at epistemic ends, and this is generally understood to be ends for the individual possessing them. Self-regarding virtues relevant to the activity of re-posting might include both open-mindedness and epistemic carefulness in our evaluation of stories.

In contrast, the focus of *other-regarding virtues* concerns the epistemic ends of those other than the possessor. In addition to desiring the appropriate epistemic ends for ourselves, we should also be concerned with the epistemic ends of those that we interact with, particularly those who we testify directly to. Jason Kawall (2002) notes that this is the parallel of many ethical virtues which motivate us through our concern for others. Generosity, for example, does benefit its possessor, but it is primarily judged by its effects on the recipient of generosity. Traditional other-regarding epistemic virtues might include honesty in our testimony to others.

Now that we see that other-regarding epistemic virtues will be relevant to re-posting, we can ask which of them is called for in our new social role as editors of content for others. While there has been comparatively little discussion of other-regarding epistemic virtues, Daukas' virtue of epistemic trustworthiness is plausibly understood as other-regarding:

A is epistemically trustworthy if and only if A is disposed to behave (when contextually appropriate) as though her epistemic status is S if and only if her epistemic status is S. [This] implies that an (ideally) epistemically trustworthy agent (sincerely) confidently asserts that P only if she knows that P, expresses doubt about P only if she has reason to doubt that P, asserts that P is possible only if P is consistent with her standing beliefs, and so on.<sup>13</sup>

This virtue may sound like it only involves self-evaluation. However, once we recognize that the behaviors here involve others, we can see that it involves the other-evaluation as well, and so is relevant to evaluating sources that we re-post:

Epistemic trustworthiness is a *social* epistemic virtue, then, insofar as it depends on appropriate attitudes towards others, as well as toward oneself, as epistemic agents.<sup>14</sup>

We may have traditionally thought of the relevant “behaviors” here as acting on one’s belief, or asserting what one believes. But once the act of re-posting articles on social media is added as a new epistemic behavior, we can see the relevance of the epistemic interests of others to that behavior.

Within the broader category of other-regarding epistemic virtues, we may also focus on a sub-category of *community-regarding epistemic virtues* which are concerned not only with our epistemic effects on particular others, but also on the epistemic community as a whole.<sup>15</sup> This goes beyond her interest in particular others; she should consider all those she may affect and how she might be exposing them to epistemic risk. This will be more difficult to judge than the direct effects of her actions, but it should be part of her considerations.

While we expect a virtuous epistemic agent to have virtues of all these kinds, community-regarding virtues are particularly relevant to the new role of re-posting articles through social media. Posting a link to an article online is in some ways like other social media posts announcing an event or posting personal picture. But it is unclear exactly how links to articles are received by readers. Rini (2017) highlighted this when she called re-posting a kind of bent testimony. Similarly

<sup>13</sup> (Daukas 2006: 111). <sup>14</sup> (Daukas 2006: 113–14).

<sup>15</sup> Since we are members of our own epistemic communities, community-regarding epistemic virtues do have a self-regarding element. We benefit epistemically when we make our epistemic communities better. Still, the action in community-regarding epistemic virtues is other-directed.

with bent credentialing. One will most often post from trusted sources, but one might sometime re-post to shame the source of the article. While the intent of the person posting might be obvious to her close friends, this intent is harder to read at a distance. Posts have a larger and less directed audience as they may appear in the feeds of friends, friends-of-friends, etc. I might post something satirically, to mock it or as an inside joke, and my close friends might understand that. But if the post is read by a larger audience, they may take me to be endorsing the source and agreeing with the content. A person with community-regarding epistemic virtues should take this effect on the larger community into account when she acts. A widely shared post runs the risk of “context collapse,” the dangers of which Karen Frost-Arnold (forthcoming) has explored. She is generally concerned that messages shared with an intended audience can take on a different meaning when moved outside of that context. Even the simple act of re-posting can take on this character, when the context of the inside joke or implied criticism is removed. Since a single message can be interpreted differently by different audiences, and we may often not be able to anticipate how it will be received by a particular audience, we need to consider the wider epistemic ends of the community, over and above our intended audience, when we engage in re-posting.

Social epistemology asks us to recognize our general dependence on the epistemic inputs of others. Virtue epistemology, particularly the virtue of trustworthiness, asks us to recognize the epistemic dependence of others on ourselves. We can now ask how being epistemically trustworthy applies to our new social role of being transmitters of potential information through re-posting. To answer this question we need to consider exactly how others are epistemically depending on us.

## 6. The Psychological Evidence about the Risks of Re-Posting

Recent psychological studies on the revisability of mistaken beliefs give us a reason to be particularly concerned about the ways that others depend on us epistemically. If we can easily use our critical faculties to revise our beliefs, then having a false belief for a short time may not be that serious an epistemic risk. However, if a falsehood, once accepted, is very hard to revise, we have reasons to be very concerned to avoid false beliefs in ourselves and not to expose others in our epistemic community to this risk.

Starting with the simplest, and perhaps most surprising cases, it is difficult to revise false beliefs even on neutral topics and when the correction is presented almost simultaneously with the original misinformation. Johnson and Seifert (1994) conducted a study where participants were exposed to a series of snippets of information, written in the style of a set of ongoing news reports about a fire in a warehouse. One of these reports indicated that there were flammable cans of paint

and pressurized gas cylinders in a closet. This information was then corrected with a report stating that the closet was in fact empty. When the correction came far after the original report, it might seem unsurprising that subjects continued to make reference to the misinformation in their explanations of the fire. However, this behavior was present even when the correction was given *immediately after* the misinformation. In these and other studies, correction of misinformation reduced the use of the misinformation in explanations given between at most by half, but in some studies correction did not reduce the use of misinformation in explanation at all.<sup>16</sup>

Johnson and Seifert's (1994) further studies demonstrated that the resiliency of misinformation against revision is strongest when the participants are presented with the misinformation in ways that encourages its use in constructing a causal explanation. Their explanation of this resiliency cited the difficulty of replacing a causal explanation with only a lack of any explanation. They did see more revision when participants were offered an alternative explanation to replace the one that they had constructed using the misinformation. So we can note not only that the revision of misinformation is difficult, but that it is particularly difficult when that misinformation is used by the recipient in an explanation.

A second type of study concerned a common myth/fact presentation of information. This kind of presentation, often used in health promoting literature, lists a commonly held myth followed by a correction with the facts. While this format might seem like it would be effective, Schwarz et al. (2007) found that, even when the myths are clearly presented as such, those presented with a myth/fact health flyer remember more myths as facts after thirty minutes than those who were not presented with the flyer. What seems like an obvious way to dispel misinformation does more harm than good.

The mechanism proposed to explain this result is our use of a familiarity heuristic. The more familiar a claim seems, the more likely we are to judge it to be true. Unfortunately, this is true even if the claim was originally heard in the context of being explicitly presented as a myth. This points to a further danger of re-posting. If many people post the same story, this may lead the recipients to judge that the content of the story is true because the story, or even the headline, is familiar.

The studies above were intentionally carried out on subject matters that participants would not have standing views about. The first talked about a fictional fire and, while some of the myth/fact presentations concerned vaccines, others concerned a brand new medical treatment for a little-known condition. However, much of the information that is posted online concerns topics about which we

<sup>16</sup> See Lewandowsky et al. (2012) for a review of these studies.

already have set opinions, and on these topics it can be even more difficult to revise our beliefs.

Neil Levy (2017) presents a whole host psychological studies that he thinks provide “Bad News about Fake News.” Among them he considers the backfire effect—those exposed to information contrary to their political views tend to become more committed to those original views. This effect has been demonstrated in a number of studies, including information debunking the original claims of weapons of mass destruction in Iraq (Nyhan and Reifler 2010) as well as in reaction to evidence of the effect of climate change on health (Hart and Nisbet 2012). In both of these studies, conservative participants responded to new information undermining their political positions by retrenching and coming to believe their original views with greater strength.

A proposed explanation for the backfire effect is that we use our current views to evaluate sources of new information. Those presenting new information that is discordant with our current beliefs are perceived as untrustworthy. This effect has been shown to carry over to other topics unrelated to the original discordant information. Marks et al. (2018) showed that people prefer to hear from those with similar political views, even for tasks as unrelated as categorizing shapes.

The backfire effect illustrates the danger of being exposed to misleading news sources; correction of a settled false belief is difficult and attempts at correction may only lead to deeper conviction in the original false belief. But further, commitment to false content can misdirect our trust (or mistrust) in new sources. If we only trust those who share our beliefs, we will mistrust the very sources that could help us correct our errors. And high levels of education don’t seem to protect against the backfire effect. In fact, Hamilton (2011) found a stronger backfire effect in more educated subjects.

Finally, there is a group of studies that are particularly relevant not to correcting misinformation but instead to the specific context of posting on social media. We have seen above that content, once accepted, will be very hard to revise. These studies show why content presented in a social media context might present a special kind of risk, over and above the general risk of encountering misleading sources in another context.

Schula et al. (2008) demonstrate what they identify as the “benefits of distrust.” When trust is elicited in subjects, they perform non-routine tasks less well than those in whom distrust is elicited. These non-routine tasks (matchstick puzzles) are taken to be a stand-in for the kind of careful thinking and questioning that might be required to identify misinformation. Interestingly, the stimulus used to elicit trust was a smiling face, to which participants were asked to match words like “warm,” “happy,” and “trustworthy.” This study gives us reason to believe that encountering an article from an apparent news source in a social media feed is particularly risky for recipients. The context in which a re-posted article is presented is one tailor-made to elicit trust. The article is likely to be surrounded

by smiling faces in vacation photos and party pictures from others' posts. Further, it may follow directly after the profile picture of the re-poster, an image that is very likely to elicit trust. Presenting an article in this context makes it particularly risky, since the recipient's trust has been elicited, and they are unlikely to be thinking particularly critically. While the Schula et al. study used elements that might be present in a social media feed, there have been more recent experiments with trust elicited directly by social media, particularly Facebook contexts. In a study by Antochi et al. (2019), subjects who had recently read a discussion thread in which a topic was discussed in a civil manner showed more willingness to trust others in a trust game. While trust in response to civil discussion seems a positive outcome, it does leave the reader vulnerable to trusting misleading news sources presented in that context.

Some contextual norms of social media may also quiet explicit dissent from an audience, which may eventually lead to more acceptance by that audience. Postings on social media include many of the kinds of things that there is a social norm to accept. It would be rude to ask for confirmation of someone's birthday, a trip they took, or a party they threw. Particularly with close friends or with those to whom we have strong social ties, we may feel some social pressure to act as though we believe them, or at least not to explicitly dissent. It would be rude to publicly question something that your aunt or dear friend re-posted. This behavior might seem like a harmless politeness, but there are potential concerns. There is experimental evidence that acting as though we believe someone can lead us to believe them. Cognitive dissonance experiments show that those who defend a position counter to their own interests on the basis of situational pressure are later more likely to self-ascribe belief in that position (Cooper 2007). Thus, just getting others to politely act as though they believe a post from a misleading news source may put them at risk of eventually coming to believe it.

## 7. Empirical Evidence about the Benefits of Re-Posting

The psychological evidence shows the epistemic risks posed to the community by posting articles from misleading news sources; misinformation that enters our epistemic communities is very hard to weed out. This gives us a reason to be very cautious about the way that we post links on social media.

But in addition to epistemic risks there are also epistemic rewards. We should also consider the positive role that we might play through re-posting and actively taking on the new social role of editor and disseminator of information. Our re-posting might serve to counteract the formation of "filter bubbles" or "echo chambers." These bubbles and chambers are formed, in part, by the algorithms that curate one's news feed. These algorithms are aimed to present one with the sorts of stories that one has clicked on before, to keep one engaged. But these

algorithms may present one with a skewed view of the world. And this is not undermined by our good practices or reading only news sources that we trust ourselves.

But there is some recent research that shows that our posting of news articles on social media might actually benefit others in our epistemic communities by helping to undermine the echo chamber effect. A 2017 study by Fletcher and Nielsen showed that people who encounter news incidentally on social media are exposed to a more diverse set of news stories than those who did not use social networks. This diversity might be explained by the “weak ties” that we have with co-workers, distant relatives, etc., on social media. These people are more likely to have a different political identity than those with whom we share strong ties.<sup>17</sup> As a result, people to whom we have weak ties are poised to provide a wider range of sources to us, and we are poised to do the same for them. This is a kind of dependence the recognition of which is central to Jones’ conception of trustworthiness. Of course, dependence brings risk; if those we have weak ties to post from misleading news sources, we are exposed to epistemic risk. However, social networks with weak ties have potential to become a potent source of information from diverse, but also trusted, sources.

## 8. How Should the Epistemically Trustworthy Respond to this Evidence?

Since epistemic trustworthiness is an epistemic virtue, we should not expect to be able to generate a list of rules that the epistemically trustworthy person will follow in her re-posting actives. Developing this virtue will instead involve a sensitivity to the relevant kind of considerations in deciding how to act in each case.<sup>18</sup> These are considerations that the epistemically trustworthy will recognize as reason-giving, and which might not be seen as reason-giving in others.

We can start by considering the content of articles re-posted. It should go without saying that the epistemically trustworthy person will be careful not to simply re-post articles with false, or even misleading, content. To do so would be to expose others to epistemic risk of false belief. In addition to the content of the article, the epistemically virtuous person will also take the time to ensure that she is not re-posting from an untrustworthy news source, even when she is confident that the content of the article is true. As noted above, economically motivated sources run a particular risk of posting a mix of true and false stories. Rather than having a specific bias, they have a reason to post whatever will get clicks and advertisement dollars. Readers of a re-posting from these sites may follow the link

<sup>17</sup> This explanation is suggested by Bakshy et al. 2012 and 2015.

<sup>18</sup> Thanks to the reviewer who pressed this point.

to the original article and so be exposed to many other false articles. But even if they happen to stumble on only true articles in the original source, linking to an untrustworthy source still exposes the reader to an epistemic risk of misguided trust. This is a risk that the virtuous person will not expose others to unnecessarily.

The psychological literature on belief correction gives particular weight to the epistemic risks here. Learning of the difficulty of correcting false beliefs once formed will give the epistemically trustworthy person reason to be particularly cautious in her re-posting. The production of false beliefs that are resilient to correction in others is a worse epistemic harm than the production of passing false beliefs. The psychological research has primarily focused on the correction of false beliefs about content, but it is likely that some of the same mechanisms will also make misguided trust resilient to correction. Once an audience trusts a source, it may be very difficult to lead them away from that trust, leaving them trapped in a misunderstanding of their own epistemic situation.

Our considerations so far concern the epistemic risks of re-posting. But there are epistemic rewards as well. Re-posting has the potential to inform others by connecting them to content and sources that they would not have consulted on their own. The epistemically trustworthy person will recognize this opportunity both to inform others and to correctly guide their trust. Exposing others to new credible sources while also credentialing those sources provides others with a way out of a potential echo chamber.

In addition to the value of exposing others to new credible sources, the epistemically trustworthy person will recognize that re-posting is an opportunity to rectify some of the harms of testimonial injustice. Fricker's virtue of testimonial justice concerns our own attitudes to sources, asking us to recognize and compensate for our own biases. Trustworthiness can take a further step to help others overcome their biases. Re-posting articles from undervalued sources, particularly those sources that are undervalued as the result of cultural biases, can help bring other's attention to sources that they ought to be trusting, but whose voices are being silenced by prejudice. Through a concern to address testimonial injustice, the epistemically trustworthy can benefit both the audience to whom they re-post and the original author.

Recognizing the epistemic harm of misguided trust, the epistemically trustworthy person will consider both the content and source for the articles that she re-posts; she will also carefully consider how the context of her re-posting can be adjusted to mitigate epistemic risk to others. The epistemically trustworthy person will consider the effect that her re-posting will have on the whole community, not only on her intended audience. Sensitivity to the possibility of context collapse may lead the epistemically trustworthy person to more clearly signal her intentions about how the audience is to read a re-posting, while also considering how wide her actual audience is. Inside jokes might be best posted only to one's inside circle.

Finally, there may be circumstances where the need to re-post a false story or re-post from a fake news source overwhelms the epistemic risks of this activity. This might happen if there is a particular pervasive false story making the rounds, or when one wants to point to the gaps in reasoning in a persuasive article. When re-posting to reveal problems in an article's content or source, it is helpful to consider ways that empirical psychology has identified for making our corrections of false beliefs effective. In their survey of the literature, Lewandowski et al. note that

To date, only three factors have been identified that can increase the effectiveness of retractions: (a) warnings at the time of the initial exposure to misinformation, (b) repetition of the retraction, and (c) corrections that tell an alternative story that fills the coherence gap otherwise left by the retraction.<sup>19</sup>

The epistemically trustworthy person should consider these facts about human psychology in judging how to best present information that is problematic. This may mean re-posting problematic articles only with a clear evaluation of them appearing before the link. It may also require giving not only a negative evaluation but also an explanation of the error; this could be an explicit discussion of an error in reasoning, or giving an alternative explanation of why the author might want to persuade the audience of the point in question. With regard to comedy news sites, this may mean indicating in the comments that this is intended to be humorous or sharing comedy only with a limited audience.

Weighing all these considerations will take more time and mental effort than simply re-posting in an unrestricted manner. But the epistemically trustworthy person should be willing to take that time and effort because she recognizes the epistemic risks (and potential rewards) that come from re-posting.

## 9. Conclusion

In this chapter, I have explored the new epistemic risks that come with re-posting on social media, and have developed the epistemic virtues that can help us to mitigate these risks. I have further considered how an epistemically trustworthy person should regulate her re-posting behavior in light of the psychological evidence that retracting false beliefs is far more difficult than we might have supposed. Behaving in an epistemically trustworthy way requires being responsive to the real risks that our actions expose others to, as well as recognizing the real ways that others depend on us. Balancing epistemic risk and dependence will lead

<sup>19</sup> Lewandowsky et al. 2012: 116.

to a very careful pattern of re-posting articles on social media. This requires being vigilant about what we re-post in order to minimize the risk of exposing others to misleading news sources. It also requires care in the way that we contextualize posts that might be misunderstood by others. But the balance of risk and dependence does not require or even encourage us to stop re-posting altogether. Despite the risks, we can serve as important sources of real news for others, and the diversity of our social groups can help expose others to real news that they might otherwise not hear at all.<sup>20</sup>

## References

Antochi, Angelo, Laura Bonelli, Fabio Paglieri, Tommaso Reggiani, and Fabio Sabatinit (2019), Civility and Trust in Social Media. *Journal of Economic Behavior and Organization* 160: 83–99.

Bakshy, Eytan, Dean Eckles, Rong Yan, and Itamar Rosenn (2012), Social Influence in Social Advertising: Evidence from Field Experiments. *Proceedings of the 13th ACM Conference on Electronic Commerce*, 146–61, <https://arxiv.org/abs/1206.4327>.

Bakshy, Eytan, Solomon Messing, and Lada A. Adamic (2015), Exposure to Ideologically Diverse News and Opinion on Facebook. *Science* 348(6239): 1130–2.

Cooper, Joel (2007), *Cognitive Dissonance: 50 Years of a Classic Theory*, Thousand Oaks, CA: Sage Publications.

Daukas, Nancy (2006), Epistemic Trust and Social Location. *Episteme* 3: 109–24.

Epictetus (1995), *The Discourses, the Handbook, Fragments of Epictetus*, ed. Christopher Gill, London: Everyman.

Fletcher, Richard and Rasmus Kleis Nielsen (2017), Are People Incidentally Exposed to News on Social Media? A Comparative Analysis. *New Media & Society* 20(7): 2450–68.

Frankfurt, H. G. (2005), *On Bullshit*, Princeton/Oxford: Princeton University Press.

Freedland, Jonathan (2017), The Long View of Targeted Fake News, BBC Radio and Podcast, <https://www.bbc.co.uk/programmes/b08jb6rt>.

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

<sup>20</sup> My thanks for the helpful suggestions and comments from audiences at the Fake Knowledge Conference by the Concept Center at the University of Cologne, the Southeastern Epistemology Conference, the University of Alabama at Birmingham, and the Works in Progress Group at UGA. I am particularly grateful for suggestions made by the anonymous reviewers, Karen Frost-Arnold, Sven Bernecker, Aaron Meskin, and Beth Preston.

Frost-Arnold, Karen (forthcoming), The Epistemic Dangers of Context Collapse Online. In Jennifer Lackey (ed.), *Applied Epistemology*. Oxford: Oxford University Press.

Gelfert, Axel (2018), Fake News: A Definition. *Informal Logic* 38(1): 84–117.

Goldman, Alvin (1999), *Knowledge in a Social World*, Oxford: Oxford University Press.

Hamilton, L. C. (2011), Education, Politics and Opinions about Climate Change Evidence for Interaction Effects. *Climatic Change*, 104: 231–42.

Hart, P. S. and E. C. Nisbett (2012), Boomerang Effects in Science Communication: How Motivated Reasoning and Identity Cues Amplify Opinion Polarization about Climate Mitigation Policies. *Communication Research*, 39: 701–23.

Johnson, H. M. and C. M. Seifert (1994), Sources of the Continued Influence Effect: When Misinformation in Memory Affects Later Inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20: 1420–36.

Jones, Karen (2012), Trustworthiness. *Ethics* 123: 61–85.

Kawall, Jason (2002), Other-Regarding Epistemic Virtues. *Ratio* 15(3): 254–75.

Levy, Neil (2017), The Bad News About Fake News. *Social Epistemology Review and Reply Collective* 6(8): 20–36.

Lewandowsky, S. et al. (2012), Misinformation and Its Correction: Continued Influence and Successful Debiasing. *Psychological Science in the Public Interest* 13: 106–31.

MacFarquhar, Neil (2018), Inside the Russian Troll Factory: Zombies and a Breakneck Pace, *New York Times*, Feb. 18, <https://www.nytimes.com/2018/02/18/world/europe/russia-troll-factory.html>.

Marks, Joseph, Eloise Copland, Eleanor Loh, Cass R. Sunstein, and Tali Sharot (2018), Epistemic Spillovers: Learning Others' Political Views Reduces the Ability to Assess and Use Their Expertise in Nonpolitical Domains. *Harvard Public Law Working Paper* No. 18-22, [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3162009](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3162009).

Memmott, Marc (2013), 75 Years Ago, 'War of the Worlds' Started a Panic. Or Did It? *NPR: The Two Way*, Oct. 30, <https://www.npr.org/sections/thetwo-way/2013/10/30/241797346/75-years-ago-war-of-the-worlds-started-a-panic-or-did-it>.

Nyhan, B. and J. Reifler (2010), When Corrections Fail: The Persistence of Political Misperceptions. *Political Behavior* 32: 303–30.

Rini, Regina (2017), Fake News and Partisan Epistemology. *Kennedy Institute of Ethics Journal*, 27(2): E-43–E-64.

Schula, Yaacov, Ruth Mayo, and Eugene Burnstein (2008), The Value of Distrust. *Journal of Experimental Social Psychology*, 44: 1293–1302.

Schwarz, N., L. J. Sanna, I. Skurnik, and C. Yoon (2007), Metacognitive Experiences and the Intricacies of Setting People Straight: Implications for Debiasing and Public Information Campaigns. *Advances in Experimental Social Psychology* 39: 127–61.

Shearer, Elisa and Jeffrey Gottfried (2017), News Use Across Social Media Platforms 2017, PEW Research Center, <https://www.journalism.org/2017/09/07/news-use-across-social-media-platforms-2017/>.

Subramanian, Samanth (2017), Inside the Macedonian Fake News Complex, *Wired*, Feb. 15, <https://www.wired.com/2017/02/veles-macedonia-fake-news/>.

Turcotte, Jason, Chance York, Jacob Irving, Rosanne M. Scholl, and Raymond J. Pingree (2015), News Recommendations from Social Media Opinion Leaders: Effects on Media Trust and Information Seeking. *Journal of Computer-Mediated Communication* 20: 520–35.

Wu, Tim (2016), *The Attention Merchants: The Epic Scramble to Get Inside Our Heads*, New York: Alfred A. Knopf.

Zagzebski, Linda (1996), *Virtues of the Mind*, Cambridge: Cambridge University Press.

# 12

## Fake News and Epistemic Rot; or, Why We Are All in This Together

*Sanford C. Goldberg*

### 1. Fake News and Epistemic Harms

By “fake news,” I have in mind

the dissemination of an item of “news” that can be traced back to a source who regards the item as false or misleading in a significant way, or else who lacks a proper regard for its truth-value,<sup>1</sup> where the original source in question has significant broadcasting abilities, and where the source’s aim is to have the item obtain uptake as a serious item of news in a given community (often though not always for political or financial ends).<sup>2</sup>

The sort of uptake in question need not involve acceptance; it might involve only being taken seriously enough to raise doubts on topics on which previously there was confidence.<sup>3</sup> (These doubts are not just the topic of the report itself, but on any related matters that might be rebutted by the topic of the report.<sup>4</sup>) And the aimed-at audience need not include everyone in the community, or even a majority of members; it might target only a select few people.<sup>5</sup>

In this chapter, I will be adopting an epistemological orientation towards fake news. My aim in doing so is twofold. First, I aim to see how epistemological theorizing can help us to characterize the distinctly epistemic harms that are associated with the prevalence of fake news. Second, I aim to see how the prevalence of fake news might force us to complicate our epistemological theories themselves—in particular our account of the epistemology of testimony. I will be

<sup>1</sup> Here I have in mind the sort of thing Harry Frankfurt called “bullshit.”

<sup>2</sup> It is worth noting that there are other phenomena that aim to spread, not news, but inflammatory images, and in this way to influence our behaviors by provoking outrage. I distinguish these cases from the phenomenon of fake news proper. (With thanks to Allan Hazlett for noting this point to me.)

<sup>3</sup> See the influential Oreskes and Conway (2010), as well as Tufekci (2017: ch. 9).

<sup>4</sup> With thanks to an anonymous referee.

<sup>5</sup> While the label “fake news” is relatively new, the phenomenon is not: within the same broad family I would include disinformation campaigns and propaganda. See Pepp et al. (2019); Habgood-Coote (2019).

arguing for three related points. First, in order to understand the epistemological harms inflicted by the prevalence of fake news, as well as to discern the sources of such harms, we need to appreciate the various roles people play in our information economy. These roles include but go beyond those typically highlighted in the epistemology of testimony. Second, while the most salient epistemological harm of fake news obtains when one is taken in by a “fake” news report, this is far from the only epistemological harm of fake news, and arguably it is not the most worrisome one. Third, and finally, our obligation to be responsible consumers of the news is grounded, at least in part, in what we owe to one another as social information-seeking creatures.

In order to reach as wide an epistemological audience as I can, I will be employing weak epistemological assumptions. In particular, I will be assuming only two kinds of epistemic harms: (1) acquiring a false belief; and (2) failing to take advantage of a manifest opportunity to acquire a true belief.<sup>6</sup> This assumption is compatible with the claim that, in addition to (1) and (2), there are other epistemic harms (involving, e.g., justified belief or knowledge). In addition, my assumption is compatible with the claim that, once a harm of type (1) or (2) obtains, there are further downstream epistemic harms—as when the false belief in (1) constitutes a defeater for another (antecedently justified) belief one has, leading one to give up that antecedently justified belief; or when the truth in (2), had it been believed, would have justified one in abandoning a false belief one has.<sup>7</sup>

Some additional comments about epistemic harms are in order. First, one might think that the epistemic harm of (1) is worse than the harm of (2). What I have to say will be consistent with this ranking, but does not assume it.<sup>8</sup> Second, I will speak of a “mismatch” in the recipient’s reaction to a news report when either (a) the recipient accepts as true a report that is false or (b) the recipient does not accept a report that in fact is true. In the latter case, the recipient might be agnostic, or else she might reject the report outright (believing it to be false). Given my characterization of epistemic harms, cases involving a mismatch are cases in which the recipient suffers an epistemic harm. Third, whether an act which brings about an epistemic harm is *immoral* depends in large part on such things as the intentions behind it, the precautions that the actor took to guard against such epistemic harms, the precautions that the audience took (as well as those it would be reasonable to expect her to take) to avoid being taken in, and so forth. These are not my topic here; I am focusing only on the epistemic dimension of the harm.

<sup>6</sup> While I will not talk of credences, the account could be tweaked to do so.

<sup>7</sup> I thank an anonymous referee for indicating the need to make this point.

<sup>8</sup> Indeed, a generic harm of fake news, when it is prevalent, is to induce doubts in the minds of an audience—doubts that undermine their sense of their own competence to distinguish what is true from what is false in a given domain. This is a point that has been emphasized by Oreskes and Conway (2010) and Tufekci (2017: 228–9).

I want to make one final preliminary point. It will be noticed that as I am construing the epistemic harms of fake news, these harms are no different in kind than the epistemic harms of false news reports (and false testimonies) more generally. Whence then the focus on the problem of fake news (in contrast to that of false news)? Simply put, while the problems posed by fake news differ only in degree from that of false news, this difference is significant nevertheless. For one thing, fake news reports are often constructed with an eye on the prejudices of the target community. As we will see below, this means that less critical members of the community have an enhanced risk of accepting these reports and passing them on in turn. Additionally, when a community of news consumers becomes aware of the prevalence of fake news—that is, of news reports that were produced by an original source that, e.g., had no interest in its truth—trust is degraded. While degraded trust can be expected to bring some epistemic benefits in an environment of fake news—it serves as some protection against the acceptance of fake news reports, and so is a guard against (1)—even so, degraded trust also risks making recipients less trusting of true reports as well—thereby increasing the prospects for (2).<sup>9</sup> And of course the failure to take advantage of an opportunity to acquire a true belief can have further downstream effects, both for oneself and for one's fellows. Accounting for these epistemic effects, and for the way in which they become pervasive, is the aim of this chapter.

## 2. Epistemic Harms of Fake News: The Simple Model of Communication

My topic, then, is the distinctly epistemic harms of fake news. I am exploring the extent to which theories of the epistemology of testimony can shed light on these harms: their source, and the nature of the challenge a community faces as it confronts this phenomenon.

I start my exploration with what I will call the **Simple Model** of communication. The model is simple in two ways. First, it recognizes only two individuals as relevant to a communicative exchange: the one who provides the information—henceforth *the reporter*<sup>10</sup>—and the one who receives the information—henceforth *the recipient*. Second, the Simple Model assumes that the considerations that are relevant to the epistemological challenge faced by the recipient are exhausted by (i) the recipient's background beliefs and (ii) whatever information the recipient picks up in observing the report itself. The information in (i) will include all of the information the recipient already has at his disposal—the information he brings with him to the exchange itself, as it were. And the information in (ii) will include

<sup>9</sup> I thank anonymous referees for indicating the need for this point.

<sup>10</sup> She need not be the original source of the information.

all of the information that the recipient picks up in the course of observing the report (hearing it or reading it): information about the reporter, the content of the report, the manner in which the report was made, the context in which the report was made, etc.

Because of these two commitments, the Simple Model is highly restricted in how it can represent the epistemic harms of fake news. This is both a virtue and a drawback. It is a virtue: since the model focuses on a very restricted set of factors, it enables us to single out the role each plays in the generation of epistemic harms. It is a drawback insofar as these are not the only relevant factors. I will develop these points in order.

Let us start by representing the epistemic task faced by every recipient on any occasion on which she encounters a news report as the task of determining whether the report itself is true or false.<sup>11</sup> According to the Simple Model, the recipient is restricted to the information in (i) and (ii) as she makes this determination. The Simple Model, then, construes the problem fake news poses as deriving from the difficulty of this task.

This construal is not implausible. To see this, consider a recipient who is well-educated, politically aware, social-network-savvy, and who is a news junky as well. If we can show that even such a recipient, utilizing the information in (i) and (ii), is at risk of a mismatch, this will make clear that success in the task identified by the Simple Model is not always easy. To this end, let us assume that our recipient is discriminating in the news sources she relies on—for the news she relies on the *New York Times* and the *Wall Street Journal*, and she avoids Breitbart and Slate<sup>12</sup> (or reads them with a very jaundiced eye when she does). She is open-minded, in that she recognizes that sources have political orientations, and tries to correct for that by reading what she regards as the responsible sources on both sides. She follows the news every day, and is also active on social media, where she often follows her friends' Facebook and Twitter posts as well as the interesting news links they share (when these show up on her news feed).

Even if such a recipient is restricted to (i) and (ii), such a recipient appears to be in a position in which to do well in her attempt to discern true news reports from false ones. And indeed she does do well—at least when she is able to identify the original source itself. Here I am going to assume that her policy is epistemically conservative: she accepts the reports she reads in the news sections of the *Times* or the *Journal*<sup>13</sup> or any other trusted source, except in those cases in which she has

<sup>11</sup> It will serve the aim of clarity to distinguish between news items, news reports, and observations of news reports. There may be other relevant distinctions, too. Even so, the tripartite distinction will do for present purposes.

<sup>12</sup> I confess that, while I am trying not to introduce political distractions in this chapter (e.g., by singling out certain media as disproportionately responsible for “fake news”), I have been impressed with the empirical studies suggesting that irresponsible journalism by, e.g., Breitbart and Fox News is in a class by itself. See Benkler et al. (2018).

<sup>13</sup> The opinion pages of these newspapers are another matter, but she reads these critically.

strong determinate evidence that supports non-acceptance; and she treats with skepticism any report whose source is not among her trusted sources (or whose original source she cannot determine). While this policy is epistemically conservative, given a good selection of reliable trusted sources it will yield a very high percentage of true beliefs. In effect she is depending on those trusted sources to do the sifting for her: reports that show up in the *Times* or the *Journal* or... have a seal of approval she recognizes as legitimate. When a report lacks this seal of approval—when it comes from Breitbart or Slate, or when she can't tell what source it comes from—she regards the report with skepticism, on the generic grounds that, when people speak or write about these topics, they have all sorts of motives for lying, distorting the truth, or highlighting only those things that support their side. She would seem to be a model citizen.

But now consider two implications of her conservative news-consuming policies.

First, since any news that does not come from a source antecedently regarded as trustworthy will be greeted with skepticism, the sources that she does regard as trustworthy are invested with a great deal of authority. With respect to them, it is not only what they report, but also what (taken collectively) they *don't* report, that informs her reaction to reports from unrecognized sources. If she encounters a report of something she hadn't previously encountered from a trusted source, she often does not accept it—especially if it is the sort of news item that (had it been true) she would have expected to have been reported in a source such as the *Times* or the *Journal*. This puts a burden on the scope of the coverage of her trusted news sources. But it also puts an added burden on her: she must be properly attuned to the sorts of things that the *Times* and the *Journal* would have found newsworthy, and she must be sensitive to the investigative limitations of even such large news outfits as the *Times* and the *Journal* (so that her expectations of them are not unduly demanding). The effect of her conservative approach to news from unrecognized sources is that she is likely not to accept a good deal of true reports. (After all, the *Times* and the *Journal* don't cover everything—despite the *Times'* motto of publishing "All The News That's Fit To Print"!)<sup>14</sup>

Second, a good deal of the news reports she encounters come at second hand, or in any case at some remove from the original report itself. Sometimes her friends report things on Facebook or Twitter, without revealing the source from which they got the news item. While sometimes she can follow up and ask for the

<sup>14</sup> It might be objected that the *Times* and the *Journal* might nevertheless choose to publish something after other newspapers have done so, in which case even this conservative approach to news consumption by a reader relying on these two newspapers alone would be less problematic. I agree, though when we bear in mind how irresponsible some news sources have been (for which, see Benkler et al. 2018), and how fake news that is spread by non-trusted sources can nevertheless affect the epistemic environment for someone like our epistemically conservative subject (for which, see below), this is not much consolation. (With thanks to Thomas Grundmann for indicating the need to address this concern.)

original source, she can't always do this; and even when she does, the reporter doesn't always remember. In such cases, the information in (ii)—the information she acquires in the act of observing the report itself—will be somewhat impoverished. To be sure, for many of the people whose reports she encounters, she has a good deal of background knowledge of *them* (how reliable they are, how discerning they are, how credulous they are, what their background politics are, etc.). But this is not always so: she has a great many Facebook friends and follows many Twitter feeds! So she remains at risk of failing to accept many true reports from unrecognized or unconfirmed reporters.<sup>15</sup>

What is more, if we supplement the Simple Model by adding a few (broadly supported and independently plausible) empirical assumptions, we can make clear that the task facing the recipient is even more challenging still. I have in mind the following three empirical assumptions:

#### TIME & ENERGY

We have limited time and energy, and thus are under pressure to be efficient in the time and energy we devote to our consumption of the news. To be sure, when it is important, we can increase the time and energy we put in; but in most everyday occasions, we employ a variety of heuristics as we approach the challenge of discriminating true news reports from false ones.

#### LIKE-MINDEDNESS

To a very large degree, we associate with people who are a lot like us. Like-mindedness is not only a matter of what we believe but also what we value.

#### SOCIAL IDENTITY

Our social identity—our membership in certain groups—is very important to us. We often do what we can to preserve our status in the groups with which we identify.

I will not bother defending these but instead will take them for granted.<sup>16</sup> If they are true, the Simple Model would lead us to predict that the task facing each of us

<sup>15</sup> Can this opportunity loss be offset by the *benefits* accruing to the conservative news consumption policy—benefits that can be seen in any informational environment teeming with fake news? In such an environment, an agent who adopts this conservative news consumption policy will avoid believing a good number of falsehoods. I agree that there can be circumstances in which the cost–benefit analysis favors the conservative policy. My current objection is not that this policy is never justified, but that the current model (which highlights the need for such a policy) fails to offer a complete picture of our predicament in connection with fake news. It fails to do so for being overly individualistic in its analysis (for which, see below). With thanks to an anonymous referee for indicating the need to address this point.

<sup>16</sup> Even so, the empirical literature on these matters is extensive. For recent work pertaining to SOCIAL IDENTITY, see Ledgerwood (2014), Kahan (2017), van Bavel and Pereira (2018), and Williams (forthcoming). See also Sunstein and Hastie (2008, 2015) (which bear on LIKE-MINDEDNESS as well as SOCIAL IDENTITY).

as recipients—the task to discriminate the fake from the real as we respond to the news reports that reach us, using only the information in (i) and (ii)—is harder than we have appreciated so far.

Given TIME & ENERGY, we will often not avail ourselves of all of the information we have in storage, but will bring to bear only those pieces of information that are brought to mind by the report itself. Of course, our reactions to reports presumably exploit unconscious information-processing as well, so the fact that we only have some knowledge “in mind” when processing the report does not rule out our reliance on other information; but having only some information “in mind” does limit what we can *say* (to ourselves or others) as we seek to rationalize our reaction. In addition, the limitations of what is brought to conscious attention affects the heuristics we deploy: any heuristic that makes inferences from information that is “in mind” will be thus restricted in the information it has available to it. Here one thinks of the availability heuristic, the recognition heuristic, the familiarity heuristic, the representativeness heuristic, and perhaps others.<sup>17</sup>

Given LIKE-MINDEDNESS, the second-hand reporters who provide us with news will be disproportionately people who read many of the same newspapers and follow many of the same blogs we do, and whose friends on Facebook overlap to some non-negligible degree with ours. But this just means that each of us will spend a disproportionate amount of time and energy on the same news items, and that our “reach” beyond the standard (print, TV, or electronic) news sources will not be significantly enhanced. Consequently, we will risk over-reliance on what in fact are a select few original news sources, and we will still face the challenge of how to properly respond to news that isn’t covered by the relied-upon sources. This is the sort of worry people tend to bring up under the labels “echo chambers” and “informational bubbles.”<sup>18</sup>

Assuming that we value our membership in various groups (as per SOCIAL IDENTITY), we will have a vested motive to do what we anticipate will preserve our membership and enhance our standing in those groups. The result is a kind of *motivated scrutiny*, whereby a recipient has the tendency to subject news reports to greater or lesser scrutiny according to whether she considers it to be “unhelpful” or “helpful” to the cause. This can lead to pressure to conform to what one anticipates as the group’s reaction to a piece of news, even if one oneself isn’t disposed to doing so on the basis of one’s evidence alone. (I will return to this below.)

<sup>17</sup> For some of the classic work on heuristics and biases (and their critics), see Tversky and Kahneman (1975), Gigerenzer (1991), and Gilovich et al. (2002).

<sup>18</sup> See, e.g., Nguyen (2020a, 2020b). For a dissenting opinion to the effect that LIKE-MINDEDNESS might be a virtue in certain normative domains—the virtue of “epistemic partisanship”—see Rini (2017).

### 3. Epistemic Harms of Fake News: The Revised Model of Communication

For all that the Simple Model can illuminate regarding the sources and types of epistemic harm that arise in connection with fake news, I don't think that this model fully illuminates the phenomenon: it is too simple. My main gripe against it is that it construes the epistemological challenge of fake news in *overly-individualistic terms*. We can see this both in the (restrictive) assumption it makes about the relevant players in a communicative exchange, and in the (restrictive) assumption it makes about the information the recipient brings to bear on assessment. In the starker terms, the Simple Model assumes that each of us consumes reports as an isolated individual, relying only on our own background information and on what we glean from observing the report itself. This neglects the fact that news is often confronted in highly social contexts—contexts in which *your fellows' reactions to the news reports you jointly observe is itself an important part of the evidence you use in grounding your own reaction to the report*.<sup>19</sup> Without recognizing this, we will fail to appreciate some of the characteristic “group-like” features of the fake news phenomenon.

How might we revise the Simple Model in order to accommodate the fact that others' reactions to a mutually observed report provide an important source of (higher-order) evidence as to the report's credibility? To begin, we need to expand the list of the participants in a communicative exchange. Here I imagine a model that postulates not only a reporter and one recipient, but potentially *many* recipients, all of whom are regarded as potentially observing one another as they all encounter the report. In addition, we also need to expand the relevant information with which recipients determine whether to accept the report. For any given recipient, this information will include not only (i) her background beliefs and (ii) whatever information the recipient picks up in observing the report itself, but also (iii) the evidence offered by others' reactions to this report. I will call this the **Revised Model** of communication.

The Revised Model agrees with the Simple Model about the nature of the epistemological challenge deriving from fake news. Both agree that this challenge centers on the task facing individual recipients on each occasion of encountering a news report—namely, the task of discerning true reports from false ones. Where the two models diverge is over the materials available to each recipient as she addresses this task, and over the relevant players in a communicative exchange.

<sup>19</sup> It is worth considering whether one of the more significant effects of social media on our patterns of news consumption is to increase the extent of the phenomenon whereby news is consumed in highly social contexts. After all, social media decreases the significance of temporal and geographic constraints on “togetherness.” The result is that we are often “together,” if only mutually—and among other things this affects how we consume the news. On this basis, I would speculate that in the age of social media the evidence of others' reactions to the news has become even more important.

To see how the addition of (iii), the evidence offered by others' reactions, might affect how a recipient responds to a report, consider a scenario in which there are multiple recipients each of whom observes the others' reactions to a mutually observed report. In such cases, any one of the recipients may rely on the reactions of any or all of the other recipients to the publicly observed news item. In this way, they exploit the *monitoring competence* of other recipients, as they each seek to assess the report for credibility. Here is a highly schematic example. Suppose it is manifest to all parties that you are watching as Sally (the speaker) reports to Rochelle (the recipient) that p. If you doubt Sally's report, you might make this manifest—that skeptical look on your face!—with the result that even if Rochelle had been otherwise inclined to accept the report, observing your doubtful appearance prompts her to think twice before doing so. Alternatively, if you have no reasons for doubt, you might remain quiet, allowing the report to pass in silence. In that case, if Rochelle has reasons to think both that you were paying attention and that if you had reasons for doubt you would have said so (or otherwise would have made this manifest in your doubtful appearance), she might regard your silence as indicating your assent. If Rochelle also takes you to be knowledgeable about the general subject matter on which Sally was speaking, she might take your silence (not only as indicating your assent but also) as further evidence that what Sally said is true—since if there had been reasons for doubt she believes that you likely would have known of them. I will call this role that a third party might play in a communicative exchange the *local monitoring* role.

The local monitoring role distributes the task of scrutinizing a report for credibility, so that it is a role played by several. Here it is worth making explicit that whenever a third party observes a communicative exchange, she is potentially playing the local monitoring role—whether or not she is among those to whom the report was addressed, and so whether or not she is among the intended audience. (This makes clear that that the distinction between the intended audience and what we might call mere overhears is less epistemically significant than is sometimes supposed.<sup>20</sup>) And whenever a third party is *taken* to observe a communicative exchange, she is potentially *regarded as* playing the local monitoring role—whether or not she was paying any attention. (This makes clear that contexts of communication are often *highly social contexts* involving *many* people, and that insofar as each of us might draw inferences from how others react to a piece of mutually observed testimony, we are epistemically relying on them—regardless of whether they were among the intended audience.)

<sup>20</sup> Here I have in mind so-called assurance views of testimony (such as Hinchman 2005, Moran 2006, and McMyler 2011) as well as trust-based views of testimony (Faulkner 2011). Both treat the fact that a testimony is addressed to a specific audience as generating a specific sort of reason for belief. I take overhearer cases to support the idea that whatever grounds there are for accepting a testimony, they can be engaged with reasons that others might offer against the truth of the say-so or the reliability of the report itself.

When all goes well, local monitoring can be a great epistemic boon in a given community. Consider the following illustration:

### IDEAL CONFERENCE

You give a paper at a conference. As you do, you recognize that a sizable number of the leading experts on the topic of your paper are in the audience. You think to yourself that if anyone is in a position to discern flaws in your position, they are. On this basis you reason that if there is relevant evidence you failed to consider, or you've made an error in your reasoning, or you've neglected a relevant alternative, etc., they are likely to spot it and let you know. Consequently, when you see that they are impressed with your argument, you increase your confidence in your line of argument and in the conclusion you've reached: after all, not only does your view square with your evidence (something you thought before the conference), it would appear that you just got some confirmation that it squares with whatever additional evidence is possessed by the experts in the audience (if it didn't, they would have said so); you also now have greater confidence that you didn't make any subtle errors of reasoning (had you done so, one of them would likely have picked up on it and let you know about it); and you have greater confidence that you haven't overlooked any relevant alternative (if you had, they would have called you out on this).<sup>21</sup>

In this idealized scenario, you enjoy a *conference-generated enhancement of your degree of warranted confidence*. This is a boon, not just for you as the speaker, but for *anyone* in the audience who was aware of the audience's level of expertise on the topic at hand. A lovely social epistemology effect!

Of course, real conferences are much messier affairs. A variety of considerations make it hard to determine precisely what epistemic significance to assign to audience reactions. There are differentials in power and social status, often reinforced by membership in social categories such as race, gender, and so forth, and these affect who speaks, for how long, and with what authority ascribed to the say-so.<sup>22</sup> Norms of participation vary greatly by culture and by context: sometimes silence merely manifests prevailing norms of politeness.<sup>23</sup> These complications make clear that the inferences we draw from audience reactions might well be erroneous: we might mistake the level of audience attentiveness, we might mischaracterize the actual knowledgeableness of (certain members of) the audience, we might misconstrue the audience reactions, e.g., by taking them to endorse

<sup>21</sup> I discuss cases of this sort at length in Goldberg (2011).

<sup>22</sup> Some ascriptions of authority are epistemically unwarranted and sometimes epistemically unjust as well. Here, of course, I have in mind the work of Fricker (2007), Dotson (2011), Medina (2013), Mills (2014), and others on epistemic injustice.

<sup>23</sup> For a discussion of the possible conversational contributions made by silence, see Goldberg (2020).

reports when in fact they harbor substantial *but silent* doubts, and so forth. So we should also consider

### THE PROBLEMATIC CONFERENCE

As above, except the audience, which is taken to be attentive and knowledgeable, is neither; and so their silence in the face of the presentation, which is construed as endorsement, is in fact owed to something else entirely.

We can imagine various different versions: audience silence reflects inattentiveness, lack of understanding, ignorance of the topic at hand, a decision to opt out of participating in the session's conversation, or.... Here the various individuals who are under a misapprehension of the audience's silence will likely increase their degree of confidence in the worthiness of the paper's claims, in ways that do *not* reflect the audience's critical monitoring capabilities.

When others are taken to play the local monitoring role, this affects the challenge a recipient confronts when she observes a report. On the one hand, others' reactions to the report serve as additional evidence for her regarding the acceptability of the report. Since the background information a single recipient brings to the scene is limited, this evidence can be highly useful: she can use their observed reactions, together with her knowledge of how discerning they are in discriminating true from false news reports, as further evidence in support of acceptance or rejection. Unfortunately, this further evidence can also be highly misleading: perhaps others are not attentive, or are not as knowledgeable as she took them to be; or perhaps she erred in construing their reaction. Mistakes of these sorts can make the evidence seem (much) stronger in one direction than in fact it is.<sup>24</sup>

What we now see is another dimension of the challenge of fake news—one that is missed by the Simple Model. The task of reacting correctly to the news reports one encounters is made both easier but also more challenging when news reports are observed in highly social contexts. Any recipient who relies on others' reactions to mutually observed reports—each of us at some point—not only must be good at discerning true from false news reports, but also must be good at discerning (the epistemically relevant aspects of) *others' reactions* to news reports.

The three empirical assumptions I introduced earlier only reinforce the challenge. Presumably, we rely on others to play this local monitoring role in part because of our limited background information, but also because of the constraints of TIME & ENERGY. But given LIKE-MINDEDNESS, we may overestimate how much *additional* information—that is, information beyond what one oneself already possesses—our fellow recipients bring to bear. This may lead us to

<sup>24</sup> See, e.g., Goldberg (2017) and Nguyen (2020a, 2020b).

overestimate the additional scrutiny they bring, and so may incline us to regard them as having more (dis)confirming evidence than in fact they have—leading to a degree of confidence on our part that does not reflect that additional evidence. In addition, given SOCIAL IDENTITY, any arbitrary recipient may feel pressure to wait to speak up in response to a report until she has a sense of what others think first; she may misconstrue the silence of others as indicating their assent, when in fact they are, like her, waiting for others to speak up first; and finally she may also feel pressure to conform to what she anticipates as the group opinion, for reasons having nothing to do with the evidence—heightening the prospects for motivated scrutiny. All of this is the recipe for the sorts of things that have been lamented in the social psychology literature: pluralistic ignorance, information cascades, and groupthink.<sup>25</sup> (While the move to consume news on social media may have exacerbated these problems, it didn't create them.<sup>26</sup>)

In short, our reliance on others as local monitors appears to bear a good deal of responsibility for creating and reinforcing the information bubbles that have been much-remarked of late. And when it comes to fake news, the epistemic risks that come in the wake of being in such a bubble are clear. Insofar as the purveyors of fake news do their homework, they will tailor the reports to maximize the attention that the news report will garner among the targeted audience. As a result, recipients are at an increased risk for a mismatch, either that of *accepting a false news report* (as when the news reported is something that flatters one's candidate or political party or one's in-group or . . .), or that of *rejecting a true news report* (as when the news reported is something that casts one's candidate in a bad light, or does the same with respect to one's political party or one's in-group or . . .). In this way, the fact that we distribute the task of local monitoring can help to both create and to reinforce our already worrisome tendencies as individuals.

We have seen that the Revised Model outperforms the Simple Model when it comes to identifying certain distinctly social features of the fake news phenomenon—features deriving from the fact that reports are consumed in highly social environments, and that the observed reactions of our fellows are among the important pieces of evidence we use as we determine whether to accept a news report. Even so, the Revised Model is not adequate. It fails to identify all of the factors that make the task of discriminating false news reports from true ones as challenging as it is. More importantly, it fails to identify an important but highly indirect type of epistemic harm that comes in the wake of the prevalence of fake news.

<sup>25</sup> Seminal work on pluralistic ignorance includes Taylor (1982) and Miller and McFarland (1991). For a general discussion of the pitfalls of group deliberation, see Sunstein and Hastie (2015).

<sup>26</sup> For an attempt to discern the distinctive contributions to these problems made by social media itself, see Sunstein (2017), Benkler et al. (2018), and (in the context of social movements) Tufekci (2017).

#### 4. Epistemic Harms of Fake News: The Modified Revised Model of Communication

It is clear that our two models so far have yet to characterize the full scope of the challenge that we confront with fake news. In this light, consider the phenomenon whereby certain news items are not disseminated widely, or (when they are disseminated) their dissemination is much slower and less extensive in a given community. Neither of the two models above illuminates this sort of phenomenon. This is because they focus on the problems that arise when a recipient *encounters* a news report; but the phenomenon just described is one in which *not all of the news reports out there will reach you in the first place*. To understand this, we need to postulate yet another epistemological role to be played in communication: the *background gardening* role.<sup>27</sup>

Consider all of the news items that have been circulating in your community for the last week. (I will call this your “ambient news environment.”) What percentage of the true news reports on topics of importance to you reach your ambient news environment? What percentage of the news items in your ambient news environment reach you? What percentage of those news items that succeed in reaching you are true? The percentages in question—the percentage of significant news reports out there that make it to one’s ambient news environment; the percentage of news items in one’s ambient news environment that reach one; and of those news items that reach one, the percentage that are true—can vary from community to community and from context to context. This is something that will not be in focus if we endorse either the Simple Model or the Revised Model.

What affects these percentages? In the context of a single individual’s news consumption online, one of the most important is the role of the algorithm: the search algorithm used by the individual subject’s preferred search engine (e.g., Google), or the algorithm that selects what is presented, and in what order, on the subject’s social media news feed (e.g., Facebook, Twitter).<sup>28</sup> But as we expand our focus from an individual (and the news items *she* encounters online) to a community (and the news items encountered by *any one* of its members in *any* venue), other factors become relevant. These include the communication technologies available to the community, how often people communicate with one another, the percentage of people who are informationally well-connected, and the structure of the networks that spread the information.<sup>29</sup> Here, however, I want to highlight a different class of factors: the pattern of news consumption by one’s fellows, and the patterns of what one’s fellows do once they have reacted to a given report.

<sup>27</sup> In previous work I have called this role the background policing role. I thank Lisa Miracchi for the suggested replacement.

<sup>28</sup> For discussion, see Miller and Record (2013).

<sup>29</sup> See, e.g., the work of Zollman (2007, 2010, 2013), Olsson (2011), and Olsson and Vallinder (2013).

News items that are endorsed are likely to be passed on; news items that have been publicly rejected are less likely to be further disseminated person-to-person, or else will be passed on in ways that would give would-be readers pause before accepting the news.<sup>30</sup> In this manner, others play a role akin to that of a gardener of a public garden: they provide background gardening of the shared ambient news environment. As such, their patterns of (public) acceptance and non-acceptance can affect such things as what news items make it to the community's ambient news environment, the speed with which a news item passes through a community's social network, and the extent to which the item gets passed along (and so how many community members it reaches).<sup>31</sup> This can have an effect on each of the percentages above.

We can now appreciate the need to modify the Revised Model, to arrive at what I will call the **Modified Revised Model** of communication: in addition to postulating a reporter and the recipient(s) who observe and react to the report (and to each other's reactions), the Modified Revised Model conceives of communicative exchanges as taking place in information environments in which background gardening has been taking place all along.

What is the epistemic significance of background gardening? It is not obvious that the existence of this phenomenon forces us to complicate our picture of the information which the recipient brings to bear in trying to arrive at a reaction to a news report. At the same time, there can be little doubt but that background gardening is epistemically significant in at least three other ways.

First, recipients now have another source of evidence to be handled properly. I have in mind one's evidence—or what one takes to be one's evidence—regarding the epistemic goodness of the background gardening in one's environment. Suppose one has false beliefs regarding the epistemic goodness of the background gardening that is done in one's environment: one takes it to be either epistemically better or epistemically worse than it is. In that case, one will likely misconstrue the epistemic impact of that gardening: one will misrepresent the percentage of observed reports that are true, and so will likely misconstrue the challenge of the task of discerning the true reports. This can affect the effort one brings to bear in assessing the news; and this in turn can lead to reports that are too-quickly accepted (or rejected), and to overconfidence when the news report is accepted (or rejected).<sup>32</sup>

Second, and relatedly, background gardening can affect the difficulty of the task that recipients face when they try to discern the true from the false reports. In a

<sup>30</sup> For critical discussion, see Rini (2017).

<sup>31</sup> There is reason to think that titillating news travels further, and more quickly, on social media—and that purveyors of false news take advantage of this. See Vosoughi et al. (2018).

<sup>32</sup> In this respect it is worth noting that the fear of the prevalence of fake news might lead one to form overly pessimistic beliefs as to the quality of the gardening being done, which in turn will have downstream epistemic effects on one's consumption of news reports.

community in which the background gardening of one's ambient news environment is epistemically good—few false reports are accepted by one's peers, so that false news items are less disseminated, and are disseminated much more slowly—the result will be that the percentage of false news items that reach one will be diminished accordingly. Of course, in a community in which the background gardening of one's ambient news environment is epistemically bad, the result will be that the percentage of false news items that reach one will be higher.

This is seen clearly in the extreme case, in which a community is perfectly successful in its background gardening, so that no false news items are left in the ambient environment. Here, a recipient who accepts every news report she encounters will end up with a 100 percent truth score in her news-based beliefs. We may not think that this alone qualifies her news-based beliefs as justified or as knowledge: perhaps epistemic responsibility requires that individual recipients monitor for credibility in any case.<sup>33</sup> Even so, the burdens on individual recipients, as they seek to accept only true reports, decrease in proportion to the epistemic goodness and the prevalence of background gardening. This is an epistemic effect of background gardening.

The lesson here is a general one: how effective one is in discerning true from false reports will depend, in part, on how effective the background gardening in one's community has been. Two recipients with the same background information and the same on-board competence at discerning true from false news reports may fare quite differently depending on whether the background gardening has been epistemically good or epistemically bad. The recipient in the good community will likely attain a higher percentage of true news-based beliefs, compared to her peer in the bad community. Since the two recipients do not differ in their background information or their on-board competence in discerning real from fake news, this result is attributable to the work done by the background gardening itself. Simply put, whether the epistemic gardening work is done by technology, individual humans, social practices, norm-governed institutions, or some combination therein, the epistemic goodness of the background gardening in one's ambient news environment can affect the difficulty of the task one faces as a recipient of the news.

So far, I have highlighted two ways in which background gardening is epistemically significant: by providing us with another source information bearing on the acceptability of news reports; and by affecting the degree of difficulty of the task of discrimination itself. But background gardening is epistemically significant for a third, more far-reaching reason: it can affect what news reports reach you in the first place. News items that your fellows publicly reject will disseminate less

<sup>33</sup> This appears to have been the basis of Elizabeth Fricker's well-known (1987, 1994, 2006) charge against anti-reductionist accounts of the epistemology of testimony. For a response, see Goldberg and Henderson (2006).

quickly, and in some cases stop being disseminated, making it less likely that you will encounter them; and news items that your fellows publicly accept will disseminate more quickly and more extensively, making it more likely that you will encounter them.<sup>34</sup> As a result, there will be cases in which *true news items fail to reach one in the first place*, as when the background gardening weeds them out or else slows (and affects the extent of) the dissemination. As above, given both LIKE-MINDEDNESS and SOCIAL IDENTITY, we would predict that these will be cases in which the true news report involves news that casts one's preferred candidate, one's preferred political party, one's in-group, etc., in a negative or unflattering light. Insofar as fake news exacerbates this, e.g., by producing reports of highly inflammatory news items, it will exacerbate the tendency individuals have to reject unhelpful news items and endorse helpful news items.

More radically still, features of the background gardening can affect which news items reach your community, and so which news items constitute part of your ambient news environment (whether you encounter them or not), in the first place. It is useful in this connection to consider how background gardening can lead to what Kristie Dotson has called *testimonial smothering* (Dotson 2011). This involves cases in which a source fails to testify out of the anticipated sense that she will not be believed by her audience. Dotson's own example concerns women and people of color attesting to conditions of oppression, though her point can be generalized. Insofar as we tend to associate with like-minded people and know this about one another, the result will be an increase in a certain kind of testimonial smothering: when factionalization is anticipated, "unfriendly" news reports will likely be stifled or "smothered." This will only reinforce the like-mindedness of members of the community itself, since in effect it is likely to have a dampening effect on the dissemination of "unfriendly" news within the community. In its most extreme form, we might even think of a *news desert*, wherein little or no unflattering news reports are disseminated in the community. Such a community may be dying of epistemic thirst without even realizing it.

There is an additional epistemic effect deriving from fact that background gardening affects the speed and extent to which news items are disseminated in a given community. In particular, this fact affects the reliability of a certain familiar kind of reasoning in the face of a surprising news item. This familiar reasoning involves cases in which we do not accept a surprising news report on the grounds that if it were true we would have heard about it by now. Under certain conditions, this can be a reliable way to react to surprising news (Goldberg 2011). But when we have distortions in the speed and prevalence of dissemination of certain types of reports, as described above, this sort of reaction will be much less

<sup>34</sup> See, again, Vosoughi et al. (2018).

reliable. Stronger, it will be unreliable in those domains in which testimonial smothering is prevalent. In this way we are at risk of rejecting surprising news that happens to be true. And while we think we are justified in doing so—after all, we are using a pattern of reasoning that under certain conditions is epistemically kosher—we are blind to the fact that present conditions are not propitious for using this sort of reasoning. Once again, this dampens the prospects that unfriendly but true news reports will be accepted, increasing the prospects for a mismatch. This makes it clear that the harms of fake news on this score go beyond the risk of failing to take advantage of a manifest opportunity to acquire a true belief—they include the risk of failing so much as to *come across* certain real news items that are otherwise circulating in the greater informational environment. By definition, those who are victims of this kind of epistemic harm will not be aware that they have been so victimized.

## 5. Epistemic Rot

I want to conclude by introducing a metaphor through which I hope to make vivid the epistemic problems that arise from the prevalence of fake news. The metaphor is *epistemic rot*. Roughly, the verb “to rot” means to (cause to) decay through the behavior of bacteria and fungus; the noun form, “rot,” designates the resulting decayed condition itself. Both of these meanings are useful to the metaphor I am trying to develop. We might think of our information environment as exhibiting degrees of health or sickness, where this is a function of (a) the degree of information saturation in the environment (how much of the news that is important to the community is actually reported in that environment) and (b) the percentage of the ambient reports that are true. We can think of fake news items, when they are widely disseminated and widely accepted, as adversely affecting the health of our epistemic community. And we can think of fake news itself as a form of (intentionally cultivated) *epistemic rot*, both in the sense that the *dissemination* of fake news is part of a process of epistemic decay in a community, and in the sense that the persistence of particular items of fake news are themselves evidence of that very decay (the rot in the system, as it were).

As an expression of the phenomenon of fake news, “epistemic rot” is apt. The parallels are many. For one thing, just as a rotting tree is one whose decaying condition is the result of organisms whose thriving brings about that decay, so, too, an epistemic community suffering from epistemic rot is one whose decaying epistemic condition is the result of purveyors of fake news—organisms whose thriving brings about that very decay. For another, in both cases the conditions that enable the process of rotting are present long before their effects (the rot itself) are seen. Finally, there is a point past which the prevalence of the decay—the rot itself—undermines any realistic hope of the recovery of health. When a tree

suffers from rot, it can reach the point at which it can no longer regain its health; when an epistemic community suffers from epistemic rot, it can reach a point at which it will be unable to recover its epistemic health on its own.

The notion of epistemic rot also enables us to make vivid our own role in the process, and the fundamentally social nature of the phenomena that were described above. We play a role as agents who unwittingly aid the process of epistemic decay when we fail to distinguish true news reports from false ones—whether by giving our support to and helping to spread false reports, or not accepting (and so helping to slow or kill off) true reports. As noted above, we are not only recipients of reports in our own right, but also cogs in the system which aid (or harm) others in their attempts to discern the fake from the real. In the same way that the effects of fungi and bacteria can be slowed, if not stopped entirely, by ensuring that conditions are not propitious for decay, so, too, the effects of fake news can be slowed, if not stopped entirely, by ensuring that the conditions are not propitious for its spread.

What is to be done is a matter I leave for another occasion. However, I want to highlight one lesson that might otherwise go unnoticed. When we think about our responsibilities as believers—our epistemic responsibilities—we often think of these responsibilities in terms of our own efforts at attaining truth and avoiding error. This is fine as far as it goes. But it does not go far enough. If I am right that we play important roles for one another in how we manage the flow of information in our environment, then it stands to reason that the demand to be epistemically responsible is one that derives, at least in part, from *what we owe to one another*. In the same way that each of us is responsible for how we behave towards others when we interact as practical subjects, so, too, each of us is responsible for how we conduct our doxastic lives, and for the information that we help to spread (or to kill off), when we engage one another as epistemic subjects. If the epistemic problem of fake news—the prospect for a slow descent into epistemic rot—is in very large part a social one, so, too, the way to address it must be social. This includes thinking of the demands of epistemic responsibility as being based, at least in part, in what we owe to each other as epistemic subjects.<sup>35</sup>

<sup>35</sup> Special thanks to Eric Bayruns, Emmalon Davis, Kate Elgin, Carolina Flores, Karen Frost-Arnold, Allan Hazlett, Melissa Koenig, Lauren Leydon-Hardy, Michael Lynch, Anna Sara Malmgren, Rachel Ann McKinney, Lisa Miracchi, Carry Osborne, Paul Silva, Miriam Solomon, Ernie Sosa, and the other members of the audience at a conference on Fake News at the University of Pennsylvania, where I gave this chapter as a talk. Thanks as well to the hive-mind at the Social Epistemology Network Facebook page, where I solicited and received excellent references for related empirical work. Last, but not least, thanks to Thomas Grundmann, Sven Bernecker, and two anonymous referees, for comments on an earlier version of this chapter.

## References

Benkler, Y., Faris, R., and Roberts, H. 2018: *Network Propoganda* (Oxford: Oxford University Press).

Dotson, K. 2011: "Tracking epistemic violence, tracking practices of silencing." *Hypatia* 26:2, pp. 236–57.

Faulkner, P. 2011: *Knowledge on Trust* (Oxford: Oxford University Press).

Fricker, E. 1987: "The epistemology of testimony." *Proceedings of the Aristotelian Society Supplementary Volume* 61, pp. 57–106.

Fricker, E. 1994: "Against gullibility." In B. Matilal and A. Chakrabarti, eds., *Knowing from Words* (Dordrecht: Springer), pp. 125–61.

Fricker, E. 2006: "Varieties of anti-reductionism about testimony—a reply to Goldberg and Henderson." *Philosophy and Phenomenological Research* 72:3, pp. 618–28.

Fricker, M. 2007: *Epistemic Injustice* (Oxford: Oxford University Press).

Gigerenzer, G. 1991: "How to make cognitive illusions disappear: Beyond 'heuristics and biases.'" *European Review of Social Psychology* 2:1, pp. 83–115.

Gilovich, T., Griffin, D., and Kahneman, D., eds. 2002: *Heuristics and Biases: The Psychology of Intuitive Judgment* (Cambridge: Cambridge University Press).

Goldberg, S. 2011: "If that were true I would have heard about it by now." In A. I. Goldman and D. Whitcomb, eds., *Social Epistemology: Essential Readings* (Oxford: Oxford University Press), pp. 92–108.

Goldberg, S. 2017: "Can asserting that p improve the speaker's epistemic position (and is that a good thing?)" *Australasian Journal of Philosophy* 95, pp. 157–70.

Goldberg, S. 2020: *Conversational Pressure: Normativity in Speech Exchanges* (Oxford: Oxford University Press).

Goldberg, S. and Henderson, D. 2006: "Monitoring and anti-reductionism in the epistemology of testimony." *Philosophy and Phenomenological Research* 72:3, pp. 576–93.

Habgood-Coote, J. 2019: "Stop talking about fake news!" *Inquiry* 62:9–10, pp. 1033–65.

Hinchman, E. S. 2005: "Telling as inviting to trust." *Philosophy and Phenomenological Research* 70:3, pp. 562–87.

Kahan, D. 2017: "Misconceptions, misinformation, and the logic of identity protective cognition." *The Cultural Cognition Project*, Working Paper No. 164, <http://www.culturalcognition.net/browse-papers/misconceptions-misinformation-and-the-logic-of-identity-prot.html>.

Ledgerwood, A. 2014: "Evaluations in their social context: Distance regulates consistency and context dependence." *Social and Personality Psychology Compass* 8:8, pp. 436–47.

McMyler, B. 2011: *Testimony, Trust, and Authority* (New York: Oxford University Press).

Medina, J. 2013: *The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and the Social Imagination* (Oxford: Oxford University Press).

Miller, B. and Record, I. 2013: "Justified belief in a digital age: On the epistemic implications of secret internet technologies." *Episteme* 10:2, pp. 101–18.

Miller, D. T. and McFarland, C. 1991: "When social comparison goes awry: The case of pluralistic ignorance." In J. Suls and T. Wills, eds., *Social Comparison: Contemporary Theory and Research* (Hillsdale, NJ: Erlbaum), pp. 287–313.

Mills, C. W. 2014: *The Racial Contract* (Ithaca, NY: Cornell University Press).

Moran, R. 2006: "Getting told and being believed." In J. Lackey and E. Sosa, eds., *The Epistemology of Testimony* (Oxford: Oxford University Press), pp. 272–306.

Nguyen, C. Thi. 2020a: "Cognitive islands and runaway echo chambers: Problems for epistemic dependence on experts." *Synthese* 197:2, pp. 1–19.

Nguyen, C. Thi. 2020b: Echo chambers and epistemic bubbles." *Episteme* 17:2, pp. 141–61.

Olsson, E. J. 2011: "A simulation approach to veritistic social epistemology." *Episteme* 8:2, pp. 127–43.

Olsson, E. J. and Vallinder, A. 2013: "Norms of assertion and communication in social networks." *Synthese* 190:13, pp. 2557–71.

Oreskes, N. and Conway, E. 2010: *Merchants of Doubt* (New York: Bloomsbury).

Pepp, J., Michaelson, E., and Sterken, R. 2019: "What's new about fake news?" *Journal of Ethics and Social Philosophy* 16:2, pp. 67–94.

Rini, R. 2017: "Fake news and partisan epistemology." *Kennedy Institute of Ethics Journal* 27:2, pp. 43–64.

Sunstein, C. 2017: *Republic.com 2.0* (Princeton, NJ: Princeton University Press).

Sunstein, C. and Hastie, R. 2008: "Four failures of deliberating groups." *University of Chicago Public Law & Legal Theory*, Working Paper No. 215, [https://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?article=1213&context=law\\_and\\_economics](https://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?article=1213&context=law_and_economics).

Sunstein, C. and Hastie, R. 2015: *Wiser: Getting Beyond Groupthink to Make Groups Smarter* (Cambridge, MA: Harvard Business Press).

Taylor, D. G. 1982: "Pluralistic ignorance and the spiral of silence: A formal analysis." *Public Opinion Quarterly* 46:3, pp. 311–35.

Tufekci, Z. 2017: *Twitter and Tear Gas: The Power and Fragility of Networked Protests* (New Haven, CT: Yale University Press).

Tversky, A. and Kahneman, D. 1975: "Judgment under uncertainty: Heuristics and biases." In D. Wendt and C. Viek, eds., *Utility, Probability, and Human Decision Making* (Dordrecht: Springer), pp. 141–62.

Van Bavel, J. J. and Pereira, A. 2018: "The partisan brain: An identity-based model of political belief." *Trends in Cognitive Sciences* 22:3, pp. 213–24.

Vosoughi, S., Roy, D., and Aral, S. 2018: "The spread of true and false news online." *Science* 359: pp. 1146–51.

Williams, D. 2020: "Socially adaptive belief." *Mind and Language*. doi: 10.1111/mila.12294.

Zollman, K. 2007: "The communication structure of epistemic communities." *Philosophy of Science* 74:5, pp. 574–87.

Zollman, K. 2010: "Social structure and the effects of conformity." *Synthese* 172:3, pp. 317–40.

Zollman, K. 2013: "Network epistemology: Communication in epistemic communities." *Philosophy Compass* 8:1, pp. 15–27.

# 13

## An Epistemic Defense of News Abstinence

*Sven Bernecker*

The man who reads nothing at all is better educated than the man who reads nothing but newspapers.

(Thomas Jefferson)

The digital age has not only promoted the growth of knowledge but also the growth of fake news and other forms of mis- and disinformation. Although people differ in their views on what constitutes fake news and which news items qualify as fake, there is general agreement that fake news poses a threat to liberal democracies because it undermines people's ability to make informed decisions. This then raises the question of how we can defend ourselves from fake news. A possible defense strategy against fake news is to cut back on one's news consumption and even to go as far as temporarily ignoring the news on certain topics or from certain sources. I refer to the latter strategy as *news abstinence*.

Usually the reasons advanced in favor of news abstinence are psychological and moral in nature. It is pointed out, for instance, that news consumption wastes valuable time and that most news is bad news, which can cause stress, aggression, tunnel vision, and desensitization. In this chapter, I abstract from these non-epistemic reasons for news abstinence and instead develop an *epistemic* argument. I argue that if we have reason to believe that by following the news, we acquire more false beliefs than true ones or we acquire beliefs that are true but irrelevant, then we may be justified in taking a newsbreak. More precisely, the argument states that we are propositionally justified in temporarily ignoring the news in a certain domain and from a certain source if the following conditions are met: (i) we are in a fake news environment or are justified in believing that we are; and (ii) it is cognitively difficult or time consuming to discriminate genuine from fake news or to obtain genuine news. The goal of the argument is to show that taking a newsbreak can be epistemically permitted. The defense of news abstinence rests either on reliabilism about justification (which explicates justification in terms of the truth-conduciveness of the belief-forming process) or on the defeasibility framework of justification (according to which positive evidence for a belief can be defeated or overridden by additional information). Given epistemic consequentialism plus

reliabilism, ignoring the news in a fake news environment is rationally required, not just rationally permitted.

Section 1 defines fake news as news that leads, or is likely to lead, a large portion of the target audience to form false beliefs. Section 2 introduces the concept of a fake news environment to account for the fact that the audience can be misled not only by forming false beliefs but also by being kept in the dark about relevant truths. A fake news environment is an environment that either contains fake news or lacks coverage reliability with respect to a certain topic or news source. Section 3 discusses different strategies to deal with fake news and explains the appeal of news abstinence, which is a form of motivated ignorance. Motivated ignorance is compared and contrasted with willful ignorance, on the one hand, and self-deception, on the other. Section 4 argues for the possibility of motivated ignorance having epistemic value. According to Section 5, we are epistemically justified to ignore the news when we can secure epistemic value better elsewhere. News abstinence is propositionally justified if the consumption of the news leads us to acquire false or irrelevant beliefs and prevents us from gaining true and relevant ones. Section 6 defends this thesis against two objections. The first objection states that motivated ignorance is as much a defense strategy against fake news, as it is a catalyst for fake news and misinformation. According to the second objection, news abstainers are trapped in an echo chamber of their own making. Neither objection holds up well under scrutiny. Section 7 contains some concluding remarks.

## 1. Fake News

The term ‘fake news’ is a recent addition to the English language but the phenomenon referred to has been around for a long time. Cailin O’Connor and James Weatherall (2018) trace the phenomenon back to the fourteenth century, when some of the great minds of the era, relying on reports by trusted peers, became convinced that lambs could grow on trees. Even if fake news is not a new phenomenon, today’s information technology makes the problem worse. Digital media allow any kind of news, including fake news, to spread significantly farther, faster, deeper, and more broadly than analog media.

Despite the widespread usage of ‘fake news,’ it is not clear whether the term has a fixed meaning and, if it does, what it is.<sup>1</sup> Accounts of fake news can be categorized along two axes: the *truth* of the information and the *truthfulness* of

<sup>1</sup> Habgood-Coote (2019) maintains that the term ‘fake news’ is either nonsense, context-sensitive, or contested and that we should therefore stop using the term. Even if I agreed that ‘fake news’ lacks a fixed meaning, it strikes me as an unrealistic demand that the term be banned from public discourse.

the attitude of the person spreading the information (Jaster & Lanius 2018: 2). The former axis concerns the question of whether only false information qualifies as fake news or whether fake news can also contain true statements, provided the true statements are misleading or are intended to be misleading. Hunt Allcott and Matthew Gentzkow, for instance, define fake news in terms of falsity. They write that fake news is “news articles that are intentionally and verifiably false, and could mislead readers” (Allcott & Gentzkow 2017: 213). Axel Gelfert, by contrast defends a more liberal view whereby fake news is “the deliberate presentation of (*typically*) false or misleading claims as news, where the claims are misleading by design” (Gelfert 2018: 108, my emphasis).

Accounts of fake news differ not only with respect to the truth of the information but also with respect to the truthfulness of the attitude of the people spreading the information. The question is whether fake news requires the intention to deceive or whether it also allows for an attitude of indifference to the truth. Following Harry Frankfurt (2005), statements that are made without any concern for the truth are called ‘bullshit.’ Is bullshit a kind of fake news or a distinct phenomenon? According to Matthew Dentith, bullshit is distinct from fake news. He defines fake news as “an allegation that some story is misleading – it contains significant omissions – or even false – it is a lie – designed to deceive its intended audience” (Dentith 2017: 66). Regina Rini, too, distinguishes fake news from bullshit:

A fake news story is one that purports to describe events in the real world, typically by mimicking the conventions of traditional media reportage, yet is known by its creators to be significantly false, and is transmitted with the two goals of being widely re-transmitted and of deceiving at least some of its audience. (Rini 2017: 45)

Nikil Mukerji, on the other hand, maintains that fake news is “bullshit asserted in the form of a news publication” (Mukerji 2018: 929).

The crux with the definitions of fake news found in the literature is that they analyze the term from the point of view of the sender of information as opposed to the receiver (Grundmann 2020). Sender-based accounts of fake news tend to be too narrow, by requiring the news to be false or the sender to have deceitful intentions. And when sender-based accounts are sufficiently broad to include phenomena such as bullshit and veridical fake news, then they lack theoretical unity because they disjoin disparate properties from semantics (truth), epistemology (justification), and psychology (intention). The only thing that connects these properties is that they mislead (or are intended to mislead) the audience.

A more fruitful way to define ‘fake news’ is to focus only on the effect it has on the receiver. Receiver-based accounts define ‘fake news’ in terms of the actual or likely misleading effect on the audience. The advantage of receiver-based accounts

over sender-based accounts is that they are sufficiently broad while still being theoretical unified.<sup>2</sup>

I agree with Grundmann and Mukerji that fake news need not be based on deceptive intentions but can rest on an attitude of indifference to the truth. I go even further by claiming that fake news may involve neither an intention to deceive nor an attitude of bullshit. Consider, for instance, an incompetent programmer who accidentally changes the settings on a news feed so that it displays only false information. Intuitively, the incompetent programmer spreads fake news even though they may neither intend to mislead nor have an attitude of indifference to the truth. This suggests that the attitude of the sender of information is irrelevant for whether it is fake news.<sup>3</sup>

A news item that is so obviously false that it does not and could not mislead anyone, even though it is intended to do so, does not qualify as fake news. News that misleads only a few members of the audience due to their idiosyncratic background assumptions but has no negative impact on the majority of the audience does not qualify as fake news either. Moreover, for news to be fake, it has to mislead the audience it is targeted at. Thus, for example, a scientific article that misleads laypersons qualifies as fake news only if laypersons are the target audience of the article.

So far, we have characterized fake news in terms of the actual or likely misleading effect it has on a large portion of the target audience. The next question is how fake news manages to mislead. Everyone agrees that the audience can be misled by being fed false propositions. However, fake news need not be false. I agree with Dentith and Gelfert that there is veridical fake news. What is distinctive about fake news is not the falsity of the news itself but the fact that it leads the audience to form false beliefs or is likely to do so. Even a true statement can convey false information. Suppose, for instance, a news report states that the number of burglaries in a neighborhood increased by 20 percent after immigrants moved in. The report does not contain any further relevant information. Even if

<sup>2</sup> Grundmann (2020) defines fake news as “news that is produced or selected in general ways such that it has the disposition to lead to a significant amount of false beliefs in the addressed consumers.”

<sup>3</sup> ‘Knowledge-lies’ are a special case of false assertions without deceitful intentions. Sorensen (2010: 610) defines a knowledge-lie as ‘an assertion that  $p \dots$  if [it is] intended to prevent the addressee from knowing that  $p$  is untrue but is not intended to deceive the addressee into believing  $p$ .’ He considers the game show ‘To Tell the Truth’ (1956–1968, CBS) to be a paradigm case of knowledge-lies. The game show features three challengers. A noteworthy person and two imposters who also claim to be the noteworthy person. The noteworthy person is sworn to answer truthfully. The challengers are permitted to lie when answering questions from a panel of four celebrities. These challengers’ statements are not designed to produce false belief but are instead intended to prevent the celebrity panel from discovering which of them is the noteworthy person. The falser the guesses, the higher the prize money for the challengers. According to Sorensen, the challengers’ statements do not involve deception because they do not bring the celebrities to flat-out believe a false proposition. And this fact is said to make knowledge-lies less bad than ordinary lies.

what the news report says is literally true, it may still be misleading. It is misleading if the implicit suggestion that there is a causal connection between the number of the immigrants and the number of burglaries is false or ungrounded.

The misleading effect fake news has on the audience typically manifests itself in the audience forming beliefs in news-related propositions that are false. Yet this is not the only way in which the news can be misleading. Sometimes the news is misleading because it keeps the audience in the dark about some relevant truths. The omission of relevant truths does not make the audience form false beliefs in news-related propositions; instead, it makes the audience form the false belief that it has been adequately informed. In the following section, the receiver-based account of fake news will be broadened to account for the idea that the news can mislead not only by commission but also by omission.

## 2. Fake News Environments

Sandy Goldberg has coined the term *coverage reliability* to refer to the degree to which sources in our social environment keep us “(i) reliably apprised of the relevant facts in a certain domain and (ii) [are] disposed to offer reliable reports regarding the obtaining of these facts (when they are believed by the source to have obtained).”<sup>4</sup> Coverage reliability consists of two components: reliable information as to whether  $p$  is the case and exposure to different types of evidence as to whether  $p$  is the case. (The variable ‘ $p$ ’ is here a schematic placeholder that stands for ‘such-and-such is the case’ or for ‘such-and-such is not the case.’) An environment exhibits coverage reliability if, across a wide range of beliefs and given certain conditions, it supports the following conditional: if some relevant  $p$  were true, the average person would hear about it in a timely fashion.

Building on the receiver-based approach to fake news proposed in the previous section, I can now introduce the notion of a ‘fake news environment.’ Such an environment does not meet the condition of coverage reliability. In a fake news environment it is not the case that if some relevant  $p$  were true, the average person would hear about it in a timely fashion.

The news coverage is, of course, never complete. Every news source covers only a limited number of topics. This does not mean, however, that fake news environments are ubiquitous. What it means instead is that the notion of a fake news

<sup>4</sup> Goldberg (2010: 157). More precisely, if an agent forms a belief  $p$  by relying on their social environment for coverage in a domain of interest  $D$ , then they epistemically depend on there being a source  $A$  in the environment such that  $A$  “(i) will (investigate and) reliably determine whether  $p$ , (ii) will be reliable in reporting the outcome of that investigation, and (iii) will satisfy both of the previous two conditions in a timely fashion” (Goldberg 2010: 159).

environment has to be relativized to particular topics and news sources as well as to what the average member of the target audience considers relevant.

The absence of coverage reliability is a necessary but not a sufficient condition for fake news environments. There are environments that lack coverage reliability even though they are not fake news environments. Consider, for instance, someone stranded on an uninhabited island in the middle of the ocean and staying there for many years. The Robinson Crusoe-like person is cut off from news of the world beyond the island. While it is clear that the castaway does not enjoy coverage reliability, it does not seem right to say that they inhabit a fake news environment. What distinguishes the desert island from a fake news environment is that the castaway cannot reasonably expect that the news coverage be reliable. A fake news environment is characterized not only by the lack of coverage reliability but also by the reasonable expectation of the inhabitant of the environment that the coverage be reliable.<sup>5</sup>

Fake news is a kind of deception by commission. In the previous section, we saw that fake news is news that leads, or is likely to lead, a large portion of the target audience to form false beliefs. While fake news deceives by commission, fake news environments deceive by omission. A fake news environment contains fake news or lacks coverage reliability with respect to certain topics or news sources despite the fact that the audience can reasonably expect the news coverage to be reliable.

In sum, the notion of fake news proposed here has two advantages over the notions found in the literature. First, false statements do not automatically qualify as fake news and true statements do not automatically qualify as genuine news. Second, fake news requires neither an intention to deceive nor an attitude of bullshit. Moreover, the notion of a fake news environment captures not only *misinformation* (incorrect information) and *disinformation* (misinformation with the deliberative aim to mislead) but also *missing information* (the non-inclusion of information that should be known or present in order to understand facts and to make decisions). The non-inclusion of information can be due to negligence, incompetence, or the desire to mislead.

<sup>5</sup> The notion of a reasonable expectation can be understood predictively or normatively (Paakkunainen 2018: 165). According to the predictive notion, whether an act (event) is reasonable to expect depends on what has happened in the past and whether the conditions that were responsible for similar acts (events) happening in the past are still in place. Given the normative notion, whether an act (event) is reasonable to expect depends on what would be reasonable to do in the given situation, absent excusing conditions. To drive this point home, consider a country where the people have gotten used to being fed false or misleading news. Given the political situation in the country, the people have no reason to expect that the epistemic quality of the news will improve in the near future. Yet it would be wrong to conclude that these people do not inhabit a fake news environment. Hence, the notion of reasonable expectation employed in the account of a fake news environment must be understood normatively.

### 3. Motivated Ignorance

Democracies need independent, fact-based news to provide a voice for a diverse range of people, to watchdog the powerful, and to keep members of a society informed. Fake news undermines democracy by subverting peoples' ability to make informed decisions in elections and referenda and by preventing public officials from acting on behalf of the public. Contrary to genuine news, fake news does not serve as a mechanism for democratic accountability and does not socialize citizens into democratic attitudes and values. Given that fake news is bad news for democracy, we need to defend ourselves from fake news. Broadly speaking, there are three strategies to combat fake news: first, government regulation of the media as well as self-regulation by media companies, second, programs to promote media literacy, and third, non-reliance on dubious news or news from dubious sources. In this chapter, I focus on the third strategy.

It can be quite taxing not to rely on dubious news items or news from dubious sources. The reason is that it is often difficult to tell reliable from unreliable sources and genuine from fake news. More and more people get their news through social media. People often do not check the source of the material that they view online before they share it, which can lead to fake news spreading quickly or even 'going viral'.<sup>6</sup> Fake news tends to be more novel and that people are more likely to share novel information (Vosoughi et al. 2018). At the same time, it has become harder to identify the original source of news stories, which can make it difficult to assess their accuracy.

Another problem with trying to defend ourselves against fake news by not relying on it is that we cannot help but be affected by it once we learn about it. Psychological findings suggest that false claims continue to affect the beliefs and attitudes of those who were exposed to them even after being debunked. This is known as the *continued influence effect* (also known as the *continued influence of misinformation*).<sup>7</sup> An innocuous example of the continued influence effect is the widespread yet false belief that the Great Wall of China is visible from the moon. Once a claim like this has been internalized, it is not easily forgotten, no matter how often it is debunked. Moreover, even if *you* stop believing in this piece of misinformation, others may learn about it from you and they may start believing in it.

The factors underlying effective messages to counter attitudes and beliefs based on misinformation have been investigated in a large meta-analysis. (Chan et al.

<sup>6</sup> In 2016, the *Washington Post* published an article describing a study conducted by a group of computer scientists from Columbia University and the French National Institute to measure the percentage of shares on social media the people had not read before posting. The study found that 59 percent of links shared on social media have never actually been clicked: in other words, most people appear to re-tweet news without ever reading it.

<sup>7</sup> According to Nyhan and Reifler (2015), a causal explanation for an unexplained event is significantly more effective than a denial even when the denial is backed by unusually strong evidence.

2017). The effectiveness of debunking has been measured in two ways. First, there is a *debunking effect*. Does finding out the initial information was false decrease the strength of people's belief in the false information? Second, there is a *persistence effect*. Does the false information continue to have some impact on beliefs later? The findings suggest that giving an elaborate debunking message to some false information has mixed effects. As one might expect, the more elaborate the debunking message, the stronger the debunking effect. That is, an elaborate message decreased people's beliefs in the false information. Surprisingly, though, an elaborate message for why the false information was encountered in the first place actually increases the long-term persistence of the false information. It is not clear yet why exactly an elaborate debunking message seems to solidify the effect of the false information. In any case, the upshot is that false information is hard to eliminate once it has incorporated into a person's belief system.

Given how difficult it can be not to rely on potentially dubious news once you are aware of it, the most effective way of achieving non-reliance on this kind of news is never to consume it. Motivated ignorance of potentially dubious news does not give rise to a fake news environment since there is no reasonable expectation that the news coverage be reliable.

News abstinence is a form of self-imposed ignorance. In epistemology, there are two competing views of what ignorance is. According to the standard view, ignorance is the absence or lack of *knowledge*. The rival position has it that ignorance is the absence or lack of *true belief*<sup>8</sup> Nothing in this chapter hinges on the debate between these views. I assume that if someone lacks the true belief that *p*, then they are ignorant of *p*. This assumption is compatible with both accounts of ignorance.

The topic of the chapter is not ignorance per se but *motivated* ignorance. To understand the notion of motivated ignorance it is helpful to first discuss the related concept of *willful* ignorance. The notion of willful ignorance is used in criminal law and jurisprudence (Charlow 1992; Husak 2010), feminist epistemology (Pohlhaus 2012; Tuana 2006), critical race theory (Mills 2007), and in the literature on terrorism and counter-terrorism (Garrard-Burnett 2009; Rodriguez 2008). I focus on the most worked-out account of willful ignorance, which is the legal account.

As the name suggests, willful ignorance is ignorance that the agent brings about voluntarily. For an attitude to qualify as willful ignorance it may not be excessively difficult for the agent to acquire true beliefs about the issue in question. Thus, the reason the agent lacks true beliefs about the issue in question is that they do not

<sup>8</sup> Among the advocates of the standard view of ignorance are DeNicola (2017: 200–2), Fields (1994: 403), Haack (2011: 25), Le Morvan (2012), and Zimmermann (2008: ix). Proponents of the rival view of ignorance are Goldman (1986: 26), Goldman & Olsson (2009: 19–21), Guerrero (2007: 62–3), Peels (2010), and van Woudenberg (2009: 375).

want to acquire these beliefs even though it would be relatively easy to do so. Jan Wieland, for instance, defines the legal notion of willful ignorance as follows:

S's ignorance of  $p$  is willful if (i)  $p$  implies that  $A$ , an action of  $S$  or another agent  $S^*$ , is wrong; (ii)  $S$  should have considered  $p$ ; (iii)  $S$  could have considered  $p$ ; but  $S$  does not consider  $p$ ; (iv) because this is inconvenient for  $S$ . (Wieland 2017: 111)<sup>9</sup>

The first thing to notice about Wieland's account of willful ignorance is that ignorance is defined not in terms of the absence of true belief but in terms of the absence of consideration. According to Wieland, being ignorant of  $p$  involves having no attitude whatsoever towards  $p$ . Willful ignorance is the voluntary act of ignoring a proposition as opposed to the voluntary act of being ignorant of the truth of a proposition. *Prima facie*, willfully ignoring something is paradoxical since it both requires being sufficiently aware of that thing to ignore it and not being aware of that thing. On closer inspection, however, the paradox dissolves. Willfully ignoring that  $p$  simply involves failing to be in the doxastic state *vis-à-vis*  $p$  that one should be in given the available evidence.

The notion of *motivated ignorance* differs from the notion of willful ignorance in at least three respects. First, it is part of the legal notion of willful ignorance that the proposition the agent is ignorant of is relevant to some issue concerning morality and legality. This does not apply to motivated ignorance. For instance, a parent may choose not to know about her teenager's social life even if the teenager is not engaged in any unlawful or immoral behavior.

Second, by not characterizing ignorance as *willful*, we can sidestep the controversial issue of whether agents can make choices free from constraints such as causal determinism. Sidestepping this issue is particularly important in the case of news abstinence. If you have reason to think that by following the news, you acquire more false beliefs than true ones or you acquire beliefs that are true but irrelevant, you may feel as if you have no choice but to take a newsbreak. You might really want to know about world affairs, but given that you have reason to believe that you inhabit a fake news environment, the process of acquiring true beliefs about world affairs is simply too difficult. So, for all you know, you become even more ignorant if you do follow the news.

Third, Wieland's legal account of willful ignorance assumes that considering the proposition in question is *inconvenient* to the agent. In the case of motivated ignorance, inconvenience is not the only thing that matters. Suppose I suspect that

<sup>9</sup> Sarch (2018) summarizes different legal account of willful ignorance and claims that the following two conditions can be found in all of them. To be willfully ignorant of an inculpatory proposition  $p$  one must have sufficiently serious suspicions of  $p$  (i.e., believe there is a sufficiently high likelihood that  $p$  is true, short of practical certainty), and deliberately (as opposed to negligently or recklessly) fail to take reasonably available steps to learn with greater certainty whether  $p$  actually is true.

two students in my class are dating each other. Since I respect the privacy of my students, I decide to ignore any evidence regarding the relationship between those students. What motivates my ignorance, though, is not that it is inconvenient to consider whether the students are having a relationship but that it would be inappropriate to do so. Just as there are virtues of seeking knowledge acquisition, there are virtues of ignorance, that is, virtues of restraint with regard to seeking knowledge of certain kinds (cf., Driver 1989; Manson 2012; Matheson 2013).

Given that lack of true belief is sufficient for ignorance and given the points in connection with Wieland's notion of willful ignorance into account, *motivated ignorance* can be defined as follows: an agent's ignorance of  $p$  is motivated if they consider  $p$  but do not acquire a true belief about  $p$ , even though they could easily do so via some epistemically respectable route. The reason the agent does not acquire a true belief is that it serves their motives (desires, interests, needs, values, or goals).<sup>10</sup>

Motivated ignorance must be distinguished not only from willful ignorance but also from *self-deception*. On the standard view, an agent is self-deceived about  $p$ , if they encounter significant evidence indicating that  $p$  is true but nevertheless believe that not- $p$  (or suspend judgment as to whether  $p$  is the case) simply because they strongly desire that  $p$  be not true (Mele 1997). Motivated ignorance and self-deception have in common that there is some truth,  $p$ , which the agent is not willing to form a belief about. This explains why motivated ignorance is often treated as a form of self-deception. DeNicola, for instance, maintains that "willful ignorance involves self-deception" (2017: 84–5). And David Jones declares that

[t]he most prevalent form of self-deception... is purposeful evasion of unwanted truths or information.... If successful, [this] results in a state of willful ignorance, which allows the person to avoid subjective distress. (Jones 2001: 782)

Yet despite the fact that motivated ignorance and self-deception both involve the refusal to form some belief for which there is significant evidence, there are at least two crucial differences between these phenomena.

First, an important feature of self-deception is that the agent has a belief that is unwarranted by the evidence they possess or that the agent fails to have the belief which the evidence they possess warrants. This feature is not shared by motivated ignorance. In cases of motivated ignorance, there is no discrepancy between the evidence the agent possesses and their propositional attitude. The reason is that the motivated ignoramus refuses to take in the available evidence that supports the proposition they are ignorant about. Motivated ignorance is a matter of steering

<sup>10</sup> Woomer (2015: 3) assumes the standard view of ignorance and defines motivated ignorance as "a state of not-knowing that is cultivated or maintained by a person in order to serve their motives (i.e. their desires, interests, needs, or goals)."

clear of evidence. Self-deception, however, is a matter of maintaining a false belief in the face of contrary evidence.

Second, besides the epistemological differences between motivated ignorance and self-deception, there are moral differences. There need not be anything morally problematic about motivated ignorance but when there is then, everything else being equal, motivated ignorance is more blameworthy than self-deception. The reason is twofold. First, motivated ignorance is brought about intentionally while self-deception is usually not. And we are to a greater extent morally responsible for intentionally performed acts than for non-intentionally performed acts. Second, it is possible to be fully aware of the fact that one is motivated to ignore something. Yet one cannot be fully aware of the fact that one is currently self-deceived. If one is aware of any self-deception, it is another person's or one's own self-deception in the past. As Kevin Lynch (2016: 516) states: "[t]o become fully aware of one's self-deception (which would involve becoming aware that one's belief is false/unwarranted) is to dispel it." And an agent is to a greater extent morally responsible for acts performed with awareness than for acts performed without awareness.

#### 4. The Epistemic Value of Motivated Ignorance

Before we can deal with the question of whether an agent can be propositionally justified in intentionally ignoring the news (genuine or fake), we have to deal with a principled objection to motivated ignorance having positive epistemic status. If true belief (knowledge) is epistemically valuable and if absence or lack of true belief is sufficient for ignorance, then ignorance lacks epistemic value. Moreover, if ignorance is not epistemically valuable, then one cannot be justified in intentionally bringing it about that one is ignorant. It is then impossible to improve one's epistemic situation by intentionally ignoring the news; instead, one has to stay tuned and try to discriminate real from fake news. The point of this section is to argue for the possibility of motivated ignorance having positive epistemic status.

Prima facie, there are three kinds of cases where ignorance allows an agent to acquire more true beliefs and fewer false ones: misleading defeaters, trivial truths, and bias-inducing information. I argue that the second and third cases, but not the first one, support the idea that motivated ignorance can be epistemically valuable.

##### 4.1 Misleading Defeater

The basic idea of a propositional defeater  $d$  is that  $S$  knows  $p$  only if there is no true proposition,  $d$ , such that if  $S$  were to believe  $d$  (or  $d$  were added to  $S$ 's reasons

for  $p$ ),  $S$  would no longer be justified in believing  $p$ . The existence of certain unpossessed evidence prevents a person from knowing  $p$  if this unpossessed evidence would result in a loss of justification were the person to acquire the evidence, be aware of it, or recognize it. A misleading defeater is a true proposition that would undermine a person's knowledge if the person were aware of it, but does not undermine that knowledge when the person is not aware of it.<sup>11</sup>

A well-known example of a misleading defeater is the *Tom Grabit case* due to Keith Lehrer and Thomas Paxson (1969). I see Tom Grabit, whom I know well, steal a book from the library. I know that Tom stole the book. But suppose that unbeknownst to me, Tom's mother has testified that Tom's identical twin stole the book. The mother's testimony is a defeater that undermines my knowledge. But let us further assume that Tom's mother is demented and that Tom has no twin brother. Given this additional information, intuitively my knowledge that Tom stole the book is reinstated. *What* the demented mother said does not undermine my knowledge. The true statement—that the mother said these things—is a misleading defeater.

If I learn about Tom's mother's statement (not knowing that she is demented and that Tom has no twin brother) then my knowledge is undermined. But as long as I remain ignorant about the mother's statement my knowledge is preserved.<sup>12</sup> This shows that unintentional ignorance of misleading defeaters is epistemic advantageous because it promotes the goal of maximizing true belief and knowledge (Pritchard 2016: 136–7). But what is true of *unintentional* ignorance of misleading defeaters does not automatically apply to *intentional* ignorance. An agent can protect their knowledge by *intentionally* ignoring a misleading defeater only if they are justified in believing that the defeater in question is in fact misleading. And when the agent is justified in believing that the defeater in question is misleading, the defeater thereby loses its knowledge-undermining power and thus ceases to be a defeater. The upshot is that motivated ignorance of misleading defeaters qua misleading defeaters is either unjustified or impossible.

To drive this point home, consider a variant of the Tom Grabit case. The case is like before except that I am aware that Tom's mother testified about her son being accused of having stolen a book but I decide to refrain from seeking out the video of her testifying, because I know that Tom's mother has been unreliable in the past. In this case, my ignorance of the content of the mother's testimony is motivated. But notice that in this case the mother's testimony does not qualify

<sup>11</sup> This account of misleading defeater is compatible with Harman's (1973: 148–9) account. According to Harman, misleading evidence against  $p$  is evidence against  $p$  when  $p$  is true.

<sup>12</sup> It might be argued that a passively ignored misleading defeater automatically transforms into a normative defeater and still robs the agent of their knowledge. A normative defeater is a proposition that one would believe to be true, if one performed one's epistemic duty and that indicates that one's belief that  $p$  is either false or unreliably formed or sustained.

as a misleading defeater since I am not aware of its content. If I did hear her testimony, I might be misled (or at least confused) because I know that even unreliable people say true things a good portion of times. But without me knowing what the mother said, her testimony does not function as a misleading defeater. As said before, motivated ignorance of misleading defeaters qua misleading defeaters is either unjustified or impossible.

There is also a principled reason that stands in the way of the defeasibility theory being able to account for motivated ignorance being epistemically valuable. The defeasibility theory assumes that the agent has a justified belief with respect to some proposition prior to learning about the defeater. When the agent learns about the defeater, the justification for the proposition in question is invalidated or diminished. This situation is different from that of a motivated ignoramus who refuses to take in the available evidence in support of a proposition. Here there may be no justification to be invalidated or diminished by the defeater.

After having discussed an unsuccessful case for the idea that motivated ignorance can be epistemically valuable let us consider two successful cases: trivial truths and bias-inducing information.

## 4.2 Trivial Truth

A popular view in epistemology has it that true belief is the prime determinant of intellectual value and the appropriate criteria of epistemic rightness are all truth-linked. This view is called *veritism*.<sup>13</sup> Critics of veritism hold that truth is not the only fundamental epistemic value but that there are others, such as understanding, justification, and knowledge.

Even if one accepts the veritistic claim that truth is the only epistemic value it does not follow that all true beliefs are of equal epistemic value. Veritism is compatible with the plausible idea that we should not be concerned with how many true propositions we believe but rather with *how much truth* we believe, where the amount of truth is a function of the degree of informativeness and relevance (Ahlstrom-Vij & Grimm 2013; Treanor 2014). The paradigm example of an uninformative and irrelevant truth is the number of grains of sand on a beach (Sosa 2002: 156). Of course, we can come up with a scenario in which knowing the number of grains of sand is important. Ordinarily, however, it is a waste of time to acquire or store this kind of knowledge.<sup>14</sup>

<sup>13</sup> Among the proponents of veritism are Alston (2005), Goldman (2002), and Sosa (2001).

<sup>14</sup> Trivial truths need to be distinguished not only from significant truths but also from prudentially or morally bad truths, which are also not worth knowing about (Zagzebski 2003: 21). Among the things that are bad to know about are arguably private information acquired through snooping or data theft, techniques of criminality, and methods for starting civil wars. I say ‘arguable,’ because the opposition

Given that true beliefs can differ in epistemic value, an agent can be justified in refraining from acquiring a large body of true but trivial beliefs in order to acquire beliefs about a few weighty truths (Haas & Vogt 2015: 20; Manson 2012: 250–2; Pritchard 2016: 141–2). Intentionally ignoring a large set of true beliefs can thus be epistemically valuable, provided the agent knows of the triviality of these beliefs and attempts to acquire weighty beliefs instead.

### 4.3 Bias-Inducing Information

Sometimes true beliefs (knowledge) can be obtainable only if we do not avail ourselves of information that would trigger biases. In clinical trials with two or more study groups, for instance, it is common to withhold true belief (knowledge) about treatment assignment from subjects and/or investigators in order to reduce the risk of bias. The withholding of information in this context is called *blinding* or *masking*.<sup>15</sup>

Blinding is used not only in clinical trials in medicine and other empirical fields but also in the legal and political context. John Rawls (1971) uses the *veil of ignorance* to test ideas about distributive fairness. The idea is that we cannot agree on a just society if we are biased by our particular situation in society (gender, race, nationality, status, wealth, etc.). We should therefore imagine we sit behind a veil of ignorance that keeps us from knowing about our natural abilities and our position in society. Behind such a veil of ignorance all individuals are simply specified as rational, free, and morally equal beings. By being ignorant of our circumstances, we can objectively consider how society should operate.

## 5. The Justification of Motivated News Abstinence

In light of what has been established up to this point, I now want to argue that an agent can be epistemically justified in temporarily ignoring the news on a certain topic or news from a particular source. If an agent has reason to believe that by following the news they acquire more false beliefs than true ones or that they acquire true beliefs but only irrelevant ones, then they are rationally permitted to take a temporary newsbreak. More precisely, an agent is propositionally justified in temporarily ignoring the news in a certain domain or from a certain source if (i) they are in a fake news environment or are justified in believing that they are,

will say that it is not morally bad to know these things. If there is anything morally bad, it has to do with the way the knowledge is acquired, or with the motivation behind the knowledge acquisition.

<sup>15</sup> It is, of course, an open question whether blinding in clinical trials leads to more objectivity. One may worry that blinding makes us abstract from relevant differences between people and erroneously assumes that people are essentially alike.

and (ii) it is cognitively difficult or time consuming to discriminate genuine from fake news or to obtain genuine news. Thus, news abstinence is epistemically justified provided it leads to us to have fewer false beliefs or more significant true beliefs.

*Condition (i):* The way condition (i) is phrased one does not have to inhabit a fake news environment to be justified in ignoring the news; it suffices to have good reasons for thinking that one inhabits a fake news environment. The situation is similar to the case of misleading defeaters discussed in the previous section. We saw that one cannot ignore a misleading defeater if one has reasons to believe that it is a genuine defeater. Similarly, an agent might still be justified in ignoring the news even when there is in fact no fake news or when coverage reliability obtains. But this requires that the agent be justified in believing that there is an abundance of fake news or that the coverage reliability is defective. In other words, the possibility of justifiably ignoring the news depends on either of two things: the presence of a fake news environment or having good reasons to believe that the conditions for a fake news environment are met.

At this point, it is instructive to draw an analogy between a fake news environment and Goldman's (1976) fake barn country. In both cases, the discriminative powers of the average person are not good enough to tell the epistemically good case from the bad case. From the country road, fake barns look just like genuine ones. Likewise, for the ordinary news consumer, fake news is indistinguishable from genuine news. And there is another similarity. Barn facsimiles in the immediate vicinity defeat knowledge claims regarding barns. By contrast, facsimiles in far distant areas and merely possible facsimiles are irrelevant. Likewise, fake news in different domains, fake news broadcast by different media outlets, and the mere possibility of fake news do not defeat knowledge about world affairs and therefore do not justify news abstinence.

A critic might wonder how an agent could become justified in believing that fake news is prevalent if they do not pay attention to the news. What could be an indicator for the prevalence of fake news, which does not require the agent to rely on dubious news sources? Two comments in response. First, news abstinence is justified only to the extent that one has reason to believe that one inhabits a fake news environment. Since the media is constantly evolving, the news abstainer must, from time to time, tune into the news to check whether the fake news environment persists. Second, when fake news environments are source specific, we may be able to use one source (e.g., National Public Radio) to justify that another source (e.g., *National Enquirer*) creates a fake news environment, provided we have good reason to think the former source is reliable.

*Condition (ii).* The possibility of justified news abstinence rests on the fact that it is cognitively costly for the agent to distinguish genuine news from fake news. In the case of text-based news, being able to tell real from fake news requires the agent to compare a given news story with accounts of the same event by different

sources and to monitor so-called *fact checkers*. In the case of image-based news, it is very difficult to detect whether an image has been manipulated with the help of artificial intelligence. What is true of images also applies to videos.<sup>16</sup> In the past, it required elaborate equipment to alter videos; today anyone can download *deepfake* software and create convincing fake videos. This has led to a proliferation of deepfakes on the Internet. Since deepfakes cannot be easily detected they pose a serious challenge to the reliability of our testimonial practices (Fallis 2019; Rini 2020). And even if there were reliable deepfake detection technology, past experience with fake news suggests that corrections rarely travel as far as initial fakes and are often not as readily believed. This was the point about the continued influence effect mentioned in Section 3.

Let us step back. The basic idea of this section is that if an agent has reason to believe that by following the news they acquire more false beliefs than true ones or that they acquire true but irrelevant beliefs, then they are propositionally justified in (temporary) ignoring the news in a certain domain or from a certain source. The argument for news abstinence given so far assumes *reliabilism*. Reliabilism about justification says that a belief is justified if it is produced by a reliable process, that is, a process that produces a high proportion of true beliefs. A justified belief may itself be false, but its mode of acquisition (or the way it is subsequently sustained) must be of a kind that typically yields truths. Random guessing, for example, does not systematically yield truths. This is why beliefs acquired by guesswork are not justified. If a belief is justified only to the extent that it is reliably acquired or sustained and if it is the hallmark of fake news environments that they lack reliability with respect to the acquisition (conservation) of beliefs and/or with respect to the coverage of relevant evidence, then one can secure more epistemic justification by ignoring the news.

The presupposition of a reliabilist approach to justification is noticeable in condition (i). Epistemic internalists take issue with the first disjunct of condition (i). Suppose fake news is prevalent, but I should think that it is rare. Moreover, suppose that condition (ii) obtains, namely, that it is difficult to discriminate genuine from fake news, and that I wish to ignore the news to avoid false and true but irrelevant beliefs. In this case, it would be irrational to ignore the news. After all, by stipulation I think that the news is reliable.

The reliabilist defense of news abstinence interacts in interesting ways with *epistemic consequentialism* (cf., Ahlstrom-Vij & Dunn 2018). Consequentialist positions in philosophy spell out normative notions by recourse to the goal of achieving states of affairs that are said to have final value. In ethics, the final value

<sup>16</sup> Examples of multimedia manipulation by means of artificial intelligence: *Smile Vector* takes the picture of a face and makes the face smile; *ArXiv 2016 Paper Video* creates 3D face models from a single 2D image; *Face2Face* lets you change the facial expressions of a target on video in real-time using a human 'puppet'; *Two Minute Papers* allows you to change the light source and shadows in any picture; and *Visually-Indicated Sounds* from MIT's (CSAIL) generates sound effects based on mute video.

is (expected) moral rightness. If moral rightness is interpreted as individual happiness (with no regard to how the well-being is distributed), we end up with utilitarianism. In epistemology, the final value is (expected) epistemic rightness. If epistemic rightness is interpreted as the attainment of true belief and the avoidance of false belief, we end up with reliabilism. Given reliabilism and epistemic consequentialism, we should form beliefs in such a way as to maximize the good of believing the truth and avoiding error.<sup>17</sup> On this view, an agent should not pay attention to the news if doing so will result in them acquiring more false beliefs than true ones. Ignoring the news in a fake news environment is then rationally *required*, not just rationally *permitted*.

Reliabilism is not the only way to argue for the thesis that an agent is epistemically justified to ignore the news when they have reason to believe that by following the news they acquire more false beliefs than true ones or that they acquire true but irrelevant beliefs. Another argument for the justification of news abstinence rests on the defeasibility framework of justification, according to which positive evidence for a belief can be defeated or overridden by additional information.<sup>18</sup> The agent's conviction that they inhabit a fake news environment functions as a *potential undercutting defeater* for any belief they would acquire if they consumed the news. An undercutting defeater is a proposition that is believed by the agent to be true yet indicates that the target belief that  $p$  is unreliably formed (sustained) or that there is no coverage reliability with respect to the evidence as to whether  $p$  is the case.<sup>19</sup> A *potential* undercutting defeater is an undercutting defeater for a belief the agent does not actually hold but could very easily hold. The belief that one inhabits a fake news environment is an example of a potential undercutting defeater; it defeats beliefs the agent would hold if they consumed the news.<sup>20</sup>

The advantage of using the defeasibility framework to argue for the justification of news abstinence is that it is independent of externalism and reliabilism about justification. The concept of epistemic defeat is essential for any fallibilistic epistemology. Internalists and externalists alike hold that grounds of belief are undermined by undefeated information, that is, information that defeats the power of the original information to put one in a position to know that the claim in question is true. Being justified in believing  $p$  must exclude an agent's

<sup>17</sup> Some epistemologists are skeptical of reliabilist consequentialism. They think that the fact that my having some belief  $p$  is likely to lead to *other* true beliefs in the future does not make it more justified or rational to believe  $p$  itself (Berker 2013). The epistemic effects of beliefs do not bear on the epistemic justification of those beliefs themselves.

<sup>18</sup> I owe this suggestion to Thomas Grundmann.

<sup>19</sup> Undercutting defeaters are distinguished from rebutting defeaters, which are propositions that are believed by the agent to be true yet indicate that the target belief is false.

<sup>20</sup> Kvanvig (2014: 34–8) is, to my knowledge, the only one who talks about 'potential defeaters.' He uses this term to explain what happens to a defeater when a defeater-defeater is present—the original defeater is transformed into a potential defeater.

having sufficient reasons for supposing either that  $p$  is false (in the case of rebutting defeat) or that the belief that  $p$  is not grounded (produced) in a way that is sufficiently truth-indicating (in the case of undercutting defeat).

## 6. Objections and Replies

In this section, I defend the epistemic defense of news abstinence against two objections.

*Objection 1.* A critic might argue that news abstinence does not work as a defense strategy against fake news because people who know very little about a certain subject matter are, by virtue of their ignorance, *more* susceptible to misinformation. An ignorant person is more likely to fall for fake news than someone who is (mis)informed. Instead of alleviating the threat posed by fake news, news abstinenence actually makes the problem worse. Suppose, for instance, the news abstainer learns about some news story through her friend who follows the news. Because the news abstainer trusts their friend and because they are unaware of the dubious source of the news story reported by the friend, the news abstainer is a helpless victim of fake news.

*Reply.* It goes without saying that justified news abstinence does not provide *complete* protection against false beliefs due to fake news. If someone does not pay attention to dubious media sources but acquires news-related false beliefs through other channels (social media or face-to-face conversations with friends), there is no epistemic net gain. And arguably it is easier to detect fake news when it is spread by news channels than if it is passed on by trusted friends. Yet for the objection at hand to hit its mark it has to be shown that, everything else being equal, someone who justifiably ignores the news acquires fewer relevant true beliefs and more false beliefs than if the same person followed the news. I do not see that this is the case. Let me explain.

A person's justified ignorance of the news may play some role in their acceptance of news-related false beliefs, but it is neither the only, nor the main factor. Psychological findings suggests that people tend to fall for fake news not because they are ignorant about particular subject matters but because they fail to reason analytically and/or are engaged in motivated reasoning.

According to Daniel Kahneman's (2011) dual-process theory, human cognition can be characterized by a distinction between autonomous, intuitive processes (System 1) and deliberative, analytic processes (System 2). System 1 is the 'gut reaction' way of thinking and making decisions. System 2 is the 'critical thinking' way of making decisions. System 1 forms first impressions and is the reason why we jump to conclusions. System 2 does reflection, problem-solving, and analysis. According to the classical reasoning approach, susceptibility to fake news is often due to people not engaging in analytic thinking. Analytic reasoning supports

sound judgment and makes people less likely to erroneously take a fake news item to be accurate (Pennycook & Rand 2019). This strategy is, of course, of no help if fake news is the product of rational reasoning on the basis of misleading evidence.

According to the motivated reasoning account, belief in fake news is primarily due to partisanship. People fall for fake news because they think in a motivated or identity-protective way. In the case of political news, for instance, people are susceptible to news stories that are amenable to their political ideology. Findings by Dan Kahan (2013) suggest that proficiency in analytic reasoning associated with System 2 information processing *increases* rather than decreases susceptibility to fake news. On this view, people who are better analytic thinkers wind up more polarized, rather than more accurate, in their beliefs. This view flies in the face of the classical reasoning approach in which analytic thinking is thought to mitigate the threat posed by fake news.

Granted the motivated reasoning account, ignorance regarding particular subject matters and news sources is not a decisive factor in determining whether someone is susceptible to fall for fake news.

To drive this point home, it is helpful to distinguish between two questions. Is it rational reasoning or motivated reasoning that makes someone believe that they inhabit a fake news environment? Is the person correct in thinking that they inhabit a fake news environment? If it is (in part) motivated reasoning that makes someone believe that they inhabit a fake news environment, then it is not clear that taking a newsbreak improves their epistemic situation—not even if they in fact live in a fake news environment. They might as well stay tuned to the news. Contrast this situation with that of someone who takes a newsbreak because they correctly see that they live in a fake news environment. Presumably, this person acquires fewer news-related false beliefs by taking a newsbreak than if they were to stay tuned to the news. This suggests that an epistemically motivated newsbreak leads to more news-related true beliefs and fewer false only if the belief that one inhabits a fake news environment is both true and rationally supported. Moreover, it is not the case, as the objection at hand claims, that an ignorant person is automatically more likely to fall for fake news than someone who is (mis) informed.

*Objection 2.* A critic might worry that the information-vacuum created by news abstinence is structured to serve *confirmation bias*, that is, the tendency to seek and privilege information that confirms one's preconceptions, and to skew interpretations toward those preconceptions. The information-vacuum allows the news abstainer to cultivate their identity-protective thinking instead of putting this thinking to the fire. In other words, the worry is that the news abstainer is trapped in an *echo chamber* of their own making and that echo chambers are, epistemically speaking, just as bad as the consumption of dubious news.

*Reply.* Thi Nguyen (2020) distinguishes between echo chambers and epistemic bubbles. An *epistemic bubble* is a social epistemic structure, which has inadequate coverage reliability through a process of exclusion by omission. Epistemic bubbles can be formed accidentally and with no ill intent, through ordinary processes of social selection and community formation (e.g., gated communities, exclusive clubs, and professional associations). An *echo chamber*, by contrast, is an epistemic structure in which other relevant voices are actively undermined and discredited. Epistemic bubbles and echo chambers have in common that they are structures of exclusion that can reinforce ideological separation. But while epistemic bubbles exclude through omission, echo chambers exclude by manipulating trust and credence. The members of an echo chamber

are not just cut off, but are actively alienated from any of the usual sources of contrary argument, consideration, or evidence. Members have been prepared to discredit and distrust any outside sources; thus, mere exposure to relevant outside information will have no effect. (Nguyen 2020)

This is why echo chambers are much harder to escape than epistemic bubbles, which can be popped simply by exposing the person to the relevant information. A person trapped in an echo chamber would reject relevant information out of hand if they were exposed to it.

The problem with the objection at hand is that it lumps epistemic bubbles in with echo chambers as part of one unified phenomenon. News abstinence can indeed contribute to the creation of echo chambers, but it need not. If news abstinence necessarily created an echo chamber, then the news abstainer would not be able to occasionally tune into the news sources in question to check whether the quality of the source has changed. Yet this is required for news abstinence to be justified. The epistemically motivated news abstainer must keep an open mind about the news sources in question and be willing and able to abandon their epistemic bubble. Those trapped within an echo chamber, however, are unable to realize that that is their predicament. Moreover, it is difficult for them to escape the echo chamber because they fundamentally mistrust the excluded news sources.

By claiming that news abstinence does not necessarily give rise to echo chambers, I do not mean to down-play the problems with news abstinence. One of the problems is that because the news abstainer encounters disagreement less frequently, they are tempted to inflate their epistemic confidence. Epistemic overconfidence is an epistemic vice because it makes reasoning errors more likely. Yet while news abstinence is not without its challenges, it is a mistake to think that the news abstainer is necessarily trapped in an echo chamber of their own making.

## 7. Conclusion

I have argued that we are propositionally justified in temporarily ignoring the news in a certain domain and from a certain source if (i) we are in a fake news environment or are justified in believing that we are, and (ii) it is cognitively difficult or time consuming to discriminate genuine from fake news or to obtain genuine news.

Ignoring the news, even if only temporarily and selectively, is undoubtedly an extreme measure to deal with the threat posed by fake news. Whether news abstinence is justified depends on how many false beliefs one can reasonably expect to acquire if one consumes the news and how irrelevant the truths are that one gathers by following the news. The harm of not being informed must be weighed against the harm caused by the consumption of the (genuine or fake) news.

Consider an analogy. Sometimes one faces a situation that requires doing something illegal in order to prevent an even greater harm from occurring. The legal defense of such actions is called a *lesser harm defense*. The Model Penal Code (§ 3.02) provides a justification defense to an agent who engages in conduct that otherwise constitutes a crime provided they believe the conduct is necessary to avoid a harm or evil that is greater than that sought to be prevented by the law defining the crime.<sup>21</sup> The proposed epistemic defense of news abstinence is a kind of lesser harm defense. For news abstinence to be justified, the anticipated epistemic harm generated by the consumption of the news must be greater than the harm caused by the creation of an information vacuum.<sup>22</sup>

## References

Ahlstrom-Vij, K., Dunn, J., eds. (2018). *Epistemic Consequentialism*. Oxford: Oxford University Press.

Ahlstrom-Vij, K., Grimm, S.R. (2013). Getting it Right. *Philosophical Studies* 166: 329–47.

Alexander, L. (2005). Lesser Evils: A Closer Look at the Paradigmatic Justification. *Law and Philosophy* 24: 611–43.

<sup>21</sup> Alexander (2005). The Model Penal Code designed to stimulate and assist U.S. state legislatures to update and standardize the penal law.

<sup>22</sup> Drafts of this chapter were presented at the Cologne Center for Contemporary Epistemology and the Kantian Tradition (CONCEPT) in March 2019; a conference on ignorance in the age of information at Scripps College in March 2019; and a conference on archeology in a post-truth world at UC Irvine in February 2020. I am grateful to these audiences for challenging questions. For discussion of the ideas and for comments on previous drafts, I am grateful to Yuval Avnur, Thomas Grundmann, Jennifer Lackey, Giulia Napolitano, Steven Norris, Francesco Praolini, Maura Priest, Luis Rosa, Paul Silva, Eyal Tal, Julie Tannenbaum, and two anonymous referees. This work has been supported by an Alexander von Humboldt Professorship award.

Allcott, H., Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives* 31: 211–36.

Alston, W.P. (2005). *Beyond Justification: Dimensions of Epistemic Evaluation*. Ithaca, NY: Cornell University Press.

Berker, S. (2013). Epistemic Teleology and the Separateness of Propositions. *Philosophical Review* 122: 337–93.

Chan, M.S., Jones, C.R., Jamieson, K.H., Albarracin, D. (2017). Debunking: A Meta-Analysis of the Psychological Efficacy of Messages Countering Misinformation. *Psychological Science* 28: 1531–46.

Charlow, R. (1992). Willful Ignorance and Criminal Culpability. *Texas Law Review* 70: 1351–429.

DeNicola, D.R. (2017). *Understanding Ignorance: The Surprising Impact of What We Don't Know*. Cambridge, MA: MIT Press.

Dentith, M.R.X. (2017). The Problem of Fake News. *Public Reason* 8: 65–79.

Driver, J. (1989). The Virtues of Ignorance. *Journal of Philosophy* 86: 373–84.

Fallis, D. (2019). The Epistemic Threat of Deepfakes. *Philosophy & Technology*. <https://doi.org/10.1007/s13347-020-00419-2>.

Fields, L. (1994). Moral Beliefs and Blameworthiness. *Philosophy* 69: 397–415.

Frankfurt, H.G. (2005). *On Bullshit*. Princeton, NJ: Princeton University Press.

Garrard-Burnett, V. (2009). Blind Eyes and Willful Ignorance: U.S. Foreign Policy, Media, and Foreign Evangelicals. In V. Garrard-Burnett (ed.), *Terror in the Land of the Holy Spirit: Guatemala Under General Efrain Rios Montt, 1982–1983* (pp. 146–66). Oxford: Oxford University Press.

Gelfert, A. (2018). Fake News: A Definition. *Informal Logic* 38: 84–117.

Goldberg, S. (2010). *Relying on Others: An Essay in Epistemology*. Oxford: Oxford University Press.

Goldman, A. (1976). Discrimination and Perceptual Knowledge. *Journal of Philosophy* 73: 771–791.

Goldman, A. (1986). *Epistemology and Cognition*. Cambridge, MA: Harvard University Press.

Goldman, A. (2002). *Pathways to Knowledge: Private and Public*. New York: Oxford University Press.

Goldman, A., Olsson, E.J. (2009). Reliabilism and the Value of Knowledge. In A. Haddock, A. Millar, and D. Pritchard (eds.), *Epistemic Value* (pp. 19–41). Oxford: Oxford University Press.

Grundmann, T. (2020). Fake News: The Case for a Purely Consumer-Oriented Explication. *Inquiry*. DOI:10.1080/0020174X.2020.1813195.

Guerrero, A.A. (2007). Don't Know, Don't Kill: Moral Ignorance, Culpability, and Caution. *Philosophical Studies* 136: 59–97.

Haack, S. (2011). 'The Ethics of Belief' Reconsidered. In M. Steup (ed.), *Knowledge, Truth and Duty: Essays on Epistemic Justification, Responsibility, and Virtues* (pp. 21–33). Oxford: Oxford University Press.

Haas, J., Vogt, K.M. (2015). Ignorance and Investigation. In M. Gross and L. McGeoy (eds.), *Routledge International Handbook of Ignorance Studies* (pp. 17–25). London: Routledge.

Habgood-Coote, J. (2019). Stop Talking about Fake News! *Inquiry* 62: 1033–65.

Harman, G. (1973). *Thought*. Princeton, NJ: Princeton University Press.

Husak, D. (2010). Willful Ignorance, Knowledge, and the 'Equal Culpability' Thesis: A Study of the Deeper Significance of the Principle of Legality. In D. Husak (ed.), *The Philosophy of Criminal Law: Selected Essays* (pp. 200–32). Oxford: Oxford University Press.

Jaster, R., Lanius, D. (2018). What is Fake News? *Versus* 2: 207–27.

Jones, D.H. (2001). Holocaust. In L.C. Becker and C.B. Becker (eds.), *Encyclopedia of Ethics* (2nd edition, pp. 779–85). New York: Routledge.

Kahan, D.M. (2013). Ideology, Motivated Reasoning, and Cognitive Reflection. *Judgment and Decision Making* 8: 407–24.

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Kvanvig, J.L. (2014). *Rationality and Reflection: How to Think About What to Think*. Oxford: Oxford University Press.

Le Morvan, P. (2012). On Ignorance: A Vindication of the Standard View. *Philosophia* 40: 379–93.

Lehrer, K., Paxson, T. (1969). Knowledge: Undefeated Justified True Belief. *Journal of Philosophy* 66: 225–37.

Lynch, K. (2016). Willful Ignorance and Self-Deception. *Philosophical Studies* 173: 505–23.

Manson, N.C. (2012). Epistemic Restraint and the Vice of Curiosity. *Philosophy* 87: 239–59.

Matheson, D. (2013). A Duty of Ignorance. *Episteme* 10: 193–205.

Mele, A.R. (1997). Real Self-Deception. *Behavioral and Brain Sciences* 20: 91–102.

Mills, C.W. (2007). White Ignorance. In S. Sullivan and N. Tuana (eds.), *Race and Epistemologies of Ignorance* (pp. 13–38) Albany, NY: SUNY Press.

Mukerji, N. (2018). What is Fake News? *Ergo* 5: 923–45.

Nguyen, C. Thi. (2020). Echo Chambers and Epistemic Bubbles. *Episteme* 17: 141–61.

Nyhan, B., Reifler, J. (2015). Displacing Misinformation about Events: An Experimental Test of Causal Corrections. *Journal of Experimental Political Science* 2: 81–93.

O'Connor, C., Weatherall, J.O. (2018). *The Misinformation Age: How False Beliefs Spread*. New Haven, CT: Yale University Press.

Paakkunainen, H. (2018). Internalism and Externalism about Reasons. In D. Star (ed.), *The Oxford Handbook of Reasons and Normativity* (pp. 143–70). Oxford: Oxford University Press.

Peels, R. (2010). What is Ignorance? *Philosophia* 38: 57–67.

Pennycook, G., Rand, D.G. (2019). Lazy, not Biased: Susceptibility to Partisan Fake News is Better Explained by Lack of Reasoning than by Motivated Reasoning. *Cognition* 188: 39–50.

Pohlhaus, G. Jr. (2012). Relational Knowing and Epistemic Injustice: Toward a Theory of Willful Hermeneutical Ignorance. *Hypatia* 27: 715–35.

Pritchard, D. (2016). Ignorance and Epistemic Value. In R. Peels and M. Blaauw (eds.), *The Epistemic Dimensions of Ignorance* (pp. 132–43). Cambridge: Cambridge University Press.

Rawls, J. (1971). *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Rini, R. (2017). Fake News and Partisan Epistemology. *Kennedy Institute of Ethics Journal* 27: 43–64.

Rini, R. (2020). Deepfakes and the Epistemic Backstop. *Philosopher's Imprint* 20. <http://hdl.handle.net/2027/spo.3521354.0020.024>.

Rodriguez, S.D. (2008). Caging Careless Birds: Examining Dangers Posed by the Willful Blindness Doctrine in the War on Terror. *University of Pennsylvania Journal of International Law* 30: 691–750.

Sarch, A. (2018). Willful Ignorance in Law and Morality. *Philosophy Compass*. DOI:10.1111/phc3.12490.

Sorensen, R. (2010). Knowledge-Lies. *Analysis* 70: 608–615.

Sosa, E. (2001). The Place of Truth in Epistemology. In M. DePaul and L. Zagzebski (eds.), *Intellectual Virtue: Perspectives from Ethics and Epistemology* (pp. 155–80). Oxford: Oxford University Press.

Sosa, E. (2002). The Place of Truth in Epistemology. In M. DePaul and L. Zagzebski (eds.), *Intellectual Virtues: Perspectives from Ethics and Epistemology* (pp. 111–34), Oxford: Clarendon Press.

Treanor, N. (2014). Trivial Truths and the Aim of Inquiry. *Philosophy and Phenomenological Research* 89: 552–9.

Tuana, N. (2006). The Speculum of Ignorance: The Women's Health Movement and Epistemologies of Ignorance. *Hypatia* 21: 1–19.

van Woudenberg, R. (2009). Ignorance and Force: Two Excusing Conditions for False Beliefs. *American Philosophical Quarterly* 46: 373–86.

Vosoughi, S., Roy, D., Aral, S. (2018). The Spread of True and False News Online. *Science* 359: 1146–51.

Wieland, J.W. (2017). Willful Ignorance. *Ethical Theory and Moral Practice* 20: 105–19.

Woomer, L.M. (2015). A Unified Account of Motivated Ignorance. Dissertation, Michigan State University. <https://search.msu.edu/index.php?q=woomer#gsc.tab=0&gsc.q=woomer&gsc.page=1>.

Zagzebski, L. (2003). The Search for the Source of the Epistemic Good. In M.S. Brady and D. Pritchard (eds.), *Moral and Epistemic Virtues* (pp. 13–27). Oxford: Blackwell.

Zimmermann, M. (2008). *Living with Uncertainty: The Moral Significance of Ignorance*. Cambridge: Cambridge University Press.

# Fake News, False Beliefs, and the Fallible Art of Knowledge Maintenance

Axel Gelfert

## 1. Introduction

Public debate about ‘fake news’ alternates between moral panic and cynicism. On one side are those who worry that electoral outcomes are increasingly influenced by secretive fake news campaigns, on the other, those who argue that propaganda and rumour have always been around. Didn’t the Athenians, during the Peloponnesian war, already blame the outbreak of plague on the Spartans’ having poisoned the wells in Piraeus (when, most likely, it was an infectious hemorrhagic fever that simply spread from its point of arrival, the port city of Piraeus)? The philosophical response to the fake news scare has been equally divided. Whereas some side, almost uncritically, with the popular view that “fake news is a problem”—as Scott Aikin and Robert Talisse put it, that’s “one thing that most people can agree on, despite the expanding breadth of their various political disagreements” (Aikin & Talisse 2018)—others, like David Coady, insist that “the problem is not fake news, it is the term ‘fake news’” (Coady 2019: 40). ‘Fake news’, so the argument goes, has no determinate meaning, and is increasingly being hijacked as a political slur to discredit unfriendly reporting. As a result, recommendations concerning how to respond to the rise of fake news (if indeed there is one) differ greatly, ranging from homilies about the need for more ‘critical thinking’ to ambitious political regulatory frameworks.

In this chapter, I argue that the term ‘fake news’ captures—albeit in an exploratory way—a novel kind of social-epistemic dysfunction, arising from *systemic* distortions of established processes of creating and disseminating news-like content. As such, fake news not only is a source of false or unreliable information, but also threatens to undermine other forms of epistemic dependence on our informational and media environment, notably our dependence on epistemic coverage. At the same time, there is an underexplored realm of remedial strategies that we *can* employ in order to safeguard, however fallibly, the integrity of our socially acquired knowledge: namely, via well-chosen *epistemic routines* that vindicate trust in reliable sources, yet which need to be periodically reviewed and adjusted in response to changes in our informational environment.

Epistemic agents navigate informational environments both by actively seeking out (and assessing) sources of information, and by relying on their environment in more unspecific, routinized ways. If dysfunctions arise from the interplay between these different modes of epistemic dependence, it will be important to have a conceptual framework at hand for locating what has gone wrong. Section 2 develops such a framework by contrasting different modes of belief revision (inferential vs. routine) and epistemic dependence (testimonial dependence on individual interlocutors vs. dependence on the epistemic coverage of one's social-epistemic environment). With these distinctions in hand, Section 3 develops a (tentative) definition of the term 'fake news', defending both its value as an exploratory concept and its specific character as referring to *systemic* distortions in the processes of aggregating and presenting news (or news-like information). Some of the overall distorting influences, due to the convergence of social, political, and technological trends that have shaped how news is being produced and consumed, are then discussed in Section 4. While these overall trends are conducive to the emergence of fake news, they are not yet sufficient: for something to constitute fake news, it must manipulate an audience's well-formed expectations about the baseline processes of news aggregation and presentation in systematic and misleading ways. Finally, in Section 5, I outline how the cultivation of successful epistemic routines holds out the promise of reducing exposure to misleading and deceptive fake news, while still allowing us to partake in the collective growth of knowledge.

## 2. Acquiring Knowledge and Maintaining Coverage through Epistemic Routines

Dysfunctions in the relationship between epistemic agents and their social-informational environment can be expected to arise in at least two ways: (1) because the environment does not provide sufficiently reliable epistemic coverage to function as a valid basis for the continued updating of the agents' beliefs, or (2) because agents fail to access existing reliable sources (perhaps instead relying on unreliable or misleading sources), for example because their epistemic routines are maladapted to the agents' social-informational environment. At the heart of both worries lies a general concern for how epistemic agents update and maintain their belief system in the face of incoming information, including, more specifically, information received in the form of testimony and reports from others.

Any epistemic agent for whom the question of whether to believe a given report arises, is already operating on the basis of a—typically extensive—web of belief. From a veritistic perspective, there are exactly two compelling reasons why an epistemic agent should revise her beliefs: to eliminate a falsehood, or to add a new true belief. Both moves alter the corpus of the agent's beliefs, which "is modified

by expansion as well as by contraction and replacement" (Levi 1983: 34). Contraction occurs whenever a belief has been found wanting, either because it has turned out to be false or because its justification has been defeated in some way. Adding new beliefs—as we constantly do, whether we rely on other people or on our own 'on-board' epistemic resources (such as perception and reasoning)—is a riskier move, in that we may unwittingly import falsehoods into our belief system, which in turn may undermine otherwise perfectly justified true beliefs we already hold.<sup>1</sup>

As human beings we cannot, of course, simply refuse belief expansion in the interest of avoiding error. Practical goals and interests, not least the basic need to have a sufficiently broad and up-to-date doxastic base that allows us to act in ways that ensure our survival, compel us to adopt new beliefs all the time—which, as David Hume knew, is nothing we can bring ourselves to avoid. At best, then, we can strive to *manage* our epistemic risk by minimizing avoidable sources of error, while eliminating erroneous beliefs when they do slip in. Such error management is key to achieving what may be called *knowledge-maintenance*—that is, maintaining doxastic procedures and conditions that are conducive not only to the acquisition of new knowledge, but also to the preservation of knowledge we already have.

Just how much voluntary control we have over our doxastic states and procedures has been a topic of much philosophical controversy (see also Holton 1994). In the case of visual perception, whatever voluntary control we have over those perceptual beliefs it generates may well be limited to judgments overruling our perceptual seemings: when faced with a pair of Müller-Lyer lines, I cannot bring myself to *see* the lines as being of equal length—one line will always *seem* longer than the other one—but, having learnt my lesson in the past, I can stop myself from *believing* that the lines are of unequal length, for I know that this is one of those situations where, predictably, I cannot trust my eyes. Yet, this does not mean that, in the *absence* of such conditions (which, let us assume, I have learnt to identify reliably), I cannot routinely accept what I see. On the contrary: taking what I see at face value, across a wide range of contexts and interests, is a legitimate—and economical—way of acquiring reliable knowledge about the world.

Following Isaac Levi (1983), one may distinguish between 'inferential' and 'routine' belief expansion. Sometimes—as in the case of being faced with what I know to be a likely optical illusion—it may be necessary for me to actively weigh the evidence, consider different alternative scenarios, and deliberate on just how much I should trust the deliverances of, say, my perceptual system. Such inferential 'weighing of the evidence', however, is cognitively demanding—and may often

<sup>1</sup> True beliefs, of course, on occasion can also undermine other true beliefs, especially under conditions of incomplete background knowledge.

be quite unnecessary (for example in contexts where the chance of my being wrong is, in fact, very low). By contrast, in the case of *routine expansion*, I have already, ahead of time, committed myself to accepting the outcome of my chosen routine as final (all else being equal). This may be because I take myself to have enough background knowledge to warrant trust in the normal functioning of the corresponding belief-forming process in the case at hand—as when I trust my own eyes for the identification of medium-size, non-Müller-Lyer objects in the vicinity. Or it may be because I have been inductively conditioned to accepting new beliefs based on *past instances* of a certain epistemic routine, “but not depending on [my] finding out what that outcome is and [then] combining it with the rest of [my] knowledge to decide upon an optimum expansion strategy” (Levi 1983: 39).

We should not assume that strategies of routine belief expansion are limited to those epistemic sources that, like the perceptual system, are ‘hard-wired’ in some way. After all, consider what such a restriction would mean for our practice of giving and receiving testimony. Every time we receive a report, we would have to activate our inferential reasoning processes, search for independent evidence, combine it with the rest of our knowledge—only to often find that such an assessment would be as compatible with the report we did receive as with its opposite, in which case we would presumably need to withhold judgment. The infeasibility of such an approach has been amply discussed by anti-inferentialist critics of testimonial reductionism, i.e. of the attempt to reduce whatever justification testimonial beliefs might enjoy to factors that can be fully ascertained first-hand—so much so that even those with reductionist inclinations for the most part acknowledge that, at least on occasion, it must be possible to acquire testimonial knowledge by *trusting someone for the truth* (rather than by individually building up a compelling inferential argument in support of the claim in question).<sup>2</sup>

It is perhaps not surprising, then, that Levi, already in his 1983 book, mentioned human testimony and the testimony of the senses—i.e. perception—in the same breath as prime candidates for the deployment of *expansion routines*:

Neither the testimony of the senses nor of other witnesses added via an expansion routine are, once admitted to X’s corpus, distinguishable from theories, laws, statistical assumptions, predictions, or other singular hypotheses with respect to certainty or [subjective] infallibility. (Levi 1983: 41)

The idea that our overall epistemic outlook is shaped to a large extent by epistemic routines—notably, by a variety of prior commitments to accept what various types of informants tell us—gives substance and structure to the frequently encountered

<sup>2</sup> For a review, see Gelfert 2014: ch. 5.

claim that, in order to acquire anything resembling the extent of knowledge that we typically credit ourselves with, we must place trust in our epistemic environment.

As an example, consider Trudy Govier's statement that "we are committed by our own judgments, by consistency, and by social practice to the epistemic reliability (in favorable contexts) of other people" (Govier 1993: 22). When understood correctly, Govier's description—and the recommendation implicit in it—seems entirely apt: if we never trusted others to be reliable reporters, we could hardly expect to acquire the kind of knowledge that we do, in fact, regularly acquire. Yet the blanket recommendation to trust others is open to the—not necessarily disingenuous—worry that, as Elizabeth Fricker puts it, "we know too much about human nature to want to trust anyone, let alone everyone, uncritically" (Fricker 1995: 400). Once we recognize that our responses to other people's testimony are structured by recurring situational contexts, social roles, and epistemic routines associated with them, it is easy to see that this worry is misplaced, or at the very least overstated.<sup>3</sup> *Of course* we do not trust just any random stranger, on any random topic under the sun—and we do not have to. Instead, we habitually rely on epistemic routines, which we have found to be trustworthy and reliable in the past (at least to the required degree), and if such routines result in our being exposed to new pieces of information, or new informants, we happily—and quite reasonably—accept those as well, *except when we have reason to suspect that our epistemic routines are no longer well-adapted to the informational environment we are in*.

It is hard to overstate the importance of our social-informational environment for the successful acquisition and maintenance of knowledge. In the case of direct reliance on the testimony of others, our epistemic reliance on those around us is readily obvious: you know something I don't know, you tell me (or I ask you about it and you respond truthfully), and I thereby come to form a new belief and, if all goes well, acquire knowledge. But we continue to depend on others for things *we already know*. This is a much less recognized form of epistemic dependence on our social-informational environment, which has recently been analysed under the label "epistemic coverage" (Goldberg 2010). The term is intended to capture our continuous reliance on our social-informational environment for keeping us abreast of relevant ongoing developments and of changes to our background knowledge of the changing world around us. Much of what we take ourselves to know is about what John McDowell describes as "reasonably durable", but in fact "impermanent states of affairs to whose continued obtaining we have only intermittent epistemic access" (McDowell 1994: 422). Yet it would be disastrous—and

<sup>3</sup> On this point, see Gelfert 2019.

a sure-fire way to land ourselves in scepticism—if we were to disavow the corresponding beliefs, merely because they concern ultimately impermanent states of affairs. And there is no need to do so, as long as we have reason to think that, if things were to drastically change around us, we would be made aware of such changes in a reasonably timely way. Sandy Goldberg refers to this as “*the coverage-reliability* of one’s community” (Goldberg 2010: 154), which is crucial to keeping us informed in ways that ultimately help to safeguard knowledge we have already obtained. The existence of “coverage-supported beliefs”, Goldberg rightly argues, shows “that our dependence on others for what we know and justifiably believe outstrips our reliance on their testimony” (Goldberg 2010: 156).

How can we ensure that our system of coverage-supported beliefs is being maintained properly? Some factors are beyond our control: If no reliable source concerning a given subject matter  $S$  is present, or readily accessible, we will be unable to update beliefs concerning  $S$  in a timely and reliable fashion. So, whether our environment is, in fact, coverage-reliable is something we have very little control over. In other cases, reliable sources concerning  $S'$  may exist, yet individuals fail to access them on a regular basis—or, for various reasons that may include partisan affiliation or a desire for affirmation, instead turn to sources of misinformation about  $S'$ . Such cases, it is easy to see, constitute a problem if one’s primary goal is the prevalence of true beliefs in a given population. Furthermore, many of us deploy epistemic routines—such as reading the newspaper—at least in part because they naturally expose us to a broad range of facts, at least some of which will be of interest to us and will help us update our beliefs about the world.

### 3. ‘Fake News’ in the Post-Truth Regime

In an epistemically ideal world, perhaps we might expect epistemic agents to selflessly offer information to the best of their abilities, trust their interlocutors for knowledge, and thereby seamlessly acquire new knowledge while, at the same time, being kept abreast of ongoing developments by a social-informational environment that has only their best interests at heart. Such a picture, however, would seem hopelessly at odds with the realities of how we receive and consume information at the end of the second decade of the twenty-first century:

One of the greatest threats we face, simply put, is bullshit. We are drowning in it. We are drowning in partisan rhetoric that is just true enough not to be a lie; in industry-sponsored research; in social media’s imitation of human connection; in legalese and corporate double-speak. It infects every facet of public life, corrupting our discourse, wrecking our trust in major institutions, lowering our standards for the truth, making it harder to achieve anything. (Lovett 2013)

These remarks, made by former Democratic speechwriter and subsequent NBC comedy producer Jon Lovett during a commencement speech at Pitzer College in Southern California in 2013, employ the term ‘bullshit’ in a colloquial way (rather than, say, by adopting Harry Frankfurt’s 1986 much-discussed philosophical definition), so one could well imagine Lovett adding the term ‘fake news’ to his list, were he to give the speech again today.

There is, without a doubt, a growing sense—not least among academics, journalists, and other epistemic ‘gatekeepers’ of sorts—that ‘fake news’ poses a new and growing problem to the social and political arena. Political debate, so a standard narrative goes, no longer takes place via reasoned argument on the basis of mutually acknowledged facts, in carefully vetted venues such as the opinion pages of newspapers or highly regulated settings like parliamentary debates,<sup>4</sup> but instead utilizes direct-messaging techniques such as Twitter, which allows political leaders to shore up support by stoking partisan fervour among their most ardent supporters. This sharp dichotomy is, of course, a caricature. Newspapers and public broadcasters have not historically been the bastions of objectivity and truthfulness that they are now, in retrospect, made out to be; and even before Facebook and Twitter, people were susceptible to ‘filter bubbles’—for example, by socializing primarily with like-minded friends and acquaintances, or by consuming media they had prior reason to think would support their already held beliefs.

The term ‘fake news’ itself first became used in a more systematic way in the early 2000s, in connection with efforts to satirize the changes in patterns of news presentation and media consumption. In particular, it was applied to *fake news shows* such as the *Colbert Report*, which ran from 2004 until 2015 on the American cable TV channel Comedy Central. These shows became a staple of satirical commentary during George W. Bush’s presidency and were criticizing an increasingly ‘fact-free’ political discourse—with comedian Stephen Colbert, in 2005, coining the term ‘truthiness’ for “the belief in what you feel to be true rather than what the facts will support”—as well as firing away at the news media’s increasingly dumbed-down mode of presentation:

As fake news, it satirizes traditional news by reporting in a style similar to network and cable TV news, but it amplifies their biases, mistakes, and deficiencies to ensure that viewers hear them loud and clear. (Gettings 2007: 26–7)

The latest twist in the story of ‘fake news’ came in the run-up to the 2016 US presidential elections, when it became clear that, on social media sites such as Facebook, fake news claims had been actively peddled to US voters by groups and organizations linked to foreign governments (notably Russia). The partisan nature

<sup>4</sup> Spoiler alert: it never did.

of many of these claims—which were heavily biased in favour of Donald Trump (e.g. by suggesting that his candidacy had the support of Pope Francis) and against his opponent, Hillary Clinton (who was, *inter alia*, accused of being a criminal, suffering from health problems, and of corruption)—led to suggestions that the election had been ‘stolen’ by foreign meddling.

It was against the backdrop of the accusation that Trump owed his electoral success to fake news peddled by Russian operatives, that Trump himself, in late 2016, adopted the term on Twitter. Not only did he repeatedly claim (falsely) to have invented the term (in fact, the term had been used by political analysts in connection with electoral campaigns from around 2014 onwards; see Beaujon 2019), but he also quickly attempted to turn the tables by applying the term to established US news outlets, such as the *New York Times*, *Washington Post*, and CNN. This added yet another wrinkle to the whirlwind career of the term since, thus understood, the term ‘fake news’ no longer singles out particular claims for criticism, but instead extends the dismissal to whole *news organizations*, such as (in Trump’s words) “the failing *New York Times*”. Through this strategic re-interpretation of the term, accusations of ‘fake news’, it has been argued “serve as a power-shifting governance mechanism to delegitimize the institutional press as a whole” (Levi 2018: 234). The idea that Trump’s co-option of the label has permanently tainted the term ‘fake news’ as a descriptive category, is not uncommon amongst philosophers writing on the topic. Thus, Matthew Dentith suspects the label ‘fake news’ to be a “rhetorical device used by the powerful to crush dissent” (Dentith 2017: 65), while Coady argues that

although the term ‘fake news’ has no fixed meaning it does have a fixed function, that of restricting permissible public speech and opinion in ways that serve the interests of powerful people and institutions. (Coady 2019: 40)

By contrast, the sociologist Steve Fuller views Trump’s usage in a more neutral, and even potentially positive, light when he characterizes it as an unconventional tool for “question[ing] the conventional liberal vehicles by which the truth/false distinction is reproduced in the American mass media” (Fuller 2018: 3). What seems clear is that current usage of the term ‘fake news’ has its roots in the recognition that political polarization and the weaponization of social media for political purposes have added a novel dimension to the way people consume, interpret, and, on occasion, discredit stories presented to them as news. In other words, proponents of the term ‘fake news’ do not merely intend it as an umbrella term for false or misleading reports, but hold that the term meets a conceptual need—though perhaps one not universally felt—that is intricately tied up with systemic changes in how publics, experts, policymakers, and the media interact.

Whereas some social epistemologists have been quick to attempt definitions of the term ‘fake news’,<sup>5</sup> others have urged philosophers to “Stop talking about fake news!” (e.g. Habgood-Coote 2018)—and not only because of the weaponization of the term lamented by Coady. Two main complaints about the term have surfaced: first, the label ‘fake news’ is used in competing and divergent ways, raising the worry that it is too vague and unstable to allow for a proper conceptual explication; second, there already exists a rich terminology for labelling communicative pathologies—from outright *lying* to (Frankfurtian) *bullshit*, all the way to political *propaganda*—such that it is quite unnecessary to introduce yet another term. Regarding the first complaint, Habgood-Coote (2018: 1039–40) writes:

I suspect that if we were to carry out a proper study of linguistic usage, we would find speakers applying it in various incompatible ways. In Tandoc, Lim, and Ling’s (2018) survey of academic usage, we see “fake news” being applied to news satire, news parody, fabricated claims, photo manipulation, and to advertising. [...] The picture is complicated when we explore the history of the term. It seems to have originally meant just “news that is fake” (Gelfert 2018 cites (Montgomery-McGovern 1898)), before coming to be associated with satirical news shows (such as the Daily Show, and the Colbert Report), before coming to be associated with profit-driven clickbait producers (Silverman and Alexander 2016), finally acquiring a use as a catch-all for bad information, and a connection with journalistic bias. Each usage has a radically different extension, going some way toward explaining the current confusion around the term.

This polysemy, so the suggestion goes, should make us pessimistic about the potential for identifying a stable conceptual core that is shared by the various attempted characterizations. The second complaint points in a similar direction:

We already have plenty of words for talking about deceit, miscommunication, and epistemic dysfunction. We can talk about lies, misleading, bullshitting, false assertion, false implicature, being unreliable, distorting the facts, being biased, propaganda, and so on. (Habgood-Coote 2018: 1047)

Why muddy the waters by introducing yet another term—especially when (as the first worry would have it) the term can mean just about anything?

It seems to me that these worries, while raising *prima facie* legitimate questions that any prospective definition should address, are nonetheless overstated. The fact that a survey of the actual usage of a term turns up a multiplicity of (partly incompatible) meanings hardly invalidates efforts to come up with a more

<sup>5</sup> Notably Levy 2017, Rini 2017, Dentith 2018, Gelfert 2018, Jaster & Lanius 2018, Mukerji 2018, Fallis and Mathiesen 2019, and Pepp et al. 2019; for a survey, see Jaster and Lanius, this volume, ch. 1.

principled definition; similarly, if the complicated history of a term were sufficient for dismissing the concept behind it, much of philosophy would be in dire straits. Rather than despairing at the first signs of empirical ‘messiness’, philosophers should take inspiration from the sciences and recognize that concepts can also serve an *exploratory* function. In the social sciences, in particular, new concepts are often needed because scientists are still in the process of figuring out whether a coherent, genuinely new phenomenon has indeed been identified. As Uljana Feest has noted with respect to the psychological sciences, “concepts are often imported from everyday language, which typically inform the ways in which scientists think about the concept’s referent”, where this exploratory process of importing concepts “can be understood as one in which these everyday intuitions are put on a more rigorous and explicit footing” (Feest 2012: 172). Seen in this light, the term ‘fake news’ can be expected to play a productive role in exploring what is novel about the ongoing shifts in the social and political world around us—and, with a bit of luck, may even be given a stipulative meaning that captures certain aspects of reality that are *not* adequately captured by the extensive list of epistemic dysfunctions given above.

Practising journalists and media scholars, perhaps more so than philosophers interested in ‘timeless’ definitions, have for some time recognized the systemic nature of the ongoing changes in the relationship between news providers, opinion makers, politicians, and the general public. These, in the words of journalist and BBC *Newsnight* presenter Evan Davis, represent “genuine changes in the way public discourse was conducted”, and Lili Levi, a communications and media law specialist, concurs when she gives the following succinct summary of a view that is shared widely among media practitioners:

We now find ourselves in an informational environment where technology enables psychometric targeting, information floods, and filter bubbles; a political environment typified by escalating polarization, extremism, and distrust; a commercial environment in which financial markets depend on high-speed trading by bots; and a journalistic environment marked by economic pressure, declining shared norms, a resurgent partisan media, harassment of journalists, and increasing uncertainty about the degree of remaining legal and non-legal protection for the press. (Levi 2018: 236)

Many of these points are reflected in extant definitions of ‘fake news’ by communications practitioners, even if no single factor suffices to adequately characterize the novelty of the phenomenon. First, there is the medium of the Internet, which has become so dominant that ‘fake news’ is sometimes defined simply as “the online publication of intentionally or knowingly false statements of fact” (Klein & Wueller 2017). Second, fake news is often equated with “false news” or as “invented entirely from thin air”, “completely fabricated”, with “no factual

basis”.<sup>6</sup> On the face of it, this might be taken to be the most natural interpretation of ‘fake news’—reflected also in more philosophical definitions of ‘fake news’ as “bullshit in the form of a news publication” (Mukerji 2018: 929), which signals a lack of concern for the truth of what is being asserted as news. Third, an element of intent is usually attributed to the creators and purveyors of fake news, which in turn is then equated with “intentionally or knowingly false statements of fact”, along with the “intention to deceive”. Yet, attributing intentions is tricky, in that it is not always the case that purveyors of fake news intend their audiences to believe the claims being made. Sometimes, as in the well-known case of pro-Trump fake news sites run by Macedonian teenagers during the 2016 election campaign, the goal may be as simple as enticing users to click on certain links, thereby generating advertising revenue, rather than changing their minds.

Elsewhere (Gelfert 2018), I argued for a definition of ‘fake news’ that does not focus on any single combination of factors at the individual level, but instead highlights its *systemic dimension*. On this account, fake news need not be deliberately misleading *per se*, but needs to be deliberately presented as news in a way that is, as a matter of fact (foreseeably, but not necessarily intentionally), likely to mislead its target audience. There is also no reason to limit our definition of fake news to linguistic data such as statements or claims, since visual content, too, can be subject to systemic distortions. While not all fake news is intended to be believed—the Macedonian clickbait farmers being a case in point—it seems reasonable to require *some* action-guiding impetus, whether this means getting people to vote for a political candidate, discriminate against a particular ethnic group, or simply click and share a particular report. In other words, fake news is not only *misleading* (in a purely cognitive sense) but also *manipulates* us in some specifiable way. False news that bears no relation whatsoever to the interest and motivations of its target audience, on this view, would not constitute ‘fake news’ in any relevant sense. Tweaking my 2018 definition accordingly, I wish to characterize ‘fake news’ as follows:

(FN\*) Fake news is the deliberate presentation of manipulative and misleading content as news, where the content is manipulative and misleading by design.

Central for the purposes of the present chapter is the phrase “by design”, which is meant to draw attention to what is novel about fake news *as it is currently being discussed*—that is, as something which manifests itself as the result of a specific convergence of ongoing social, political, and technological developments. In the spirit of conceptual exploration advocated earlier, I am not especially wedded to any single ingredient of (FN\*), yet I do think that the *systemic dimension* of the

<sup>6</sup> All quoted snippets in this paragraph are from extant ‘working definitions’ of the term ‘fake news’; see Gelfert 2018: 96.

current fake news phenomenon ultimately needs to be reflected by any prospective definition, for, ultimately it is systemic features inherent in the processes of creation and dissemination of fake news that give the phenomenon of fake news its novel significance and require the introduction of a new term. This is what the phrase “by design” is meant to capture. For a certain content presented as news to count as fake news, it must not merely be likely to mislead *simpliciter*, but must be misleading *by design*. An isolated one-off lie, told in private by one person to another, even when it appears to communicate newsworthy content, would hardly constitute fake news—not least since casually ‘quoting’ a news item hardly counts as presenting oneself as a news source. At the same time, mild distorting influences that no news source—real or fake—could escape, would hardly suffice to render reports produced in such environments (which, given the scenario, would encompass *all* news reports!) fake news. Fake news differs from genuine news in that it systematically exploits an audience’s well-formed expectations about the baseline processes of news aggregation and presentation, seeking to manipulate its audience and shape its responses in a way that can reasonably be expected to result in a preponderance of false or misleading content.

#### **4. News Production and Consumption under Conditions of the ‘Attention Economy’**

The systemic dimension of the processes and conditions that give rise to fake news becomes readily obvious, once we reflect on some of the pervasive changes that have taken place in Western media landscapes in recent years. While this is not the place for a detailed historical survey, a few observations may help appreciate how processes of news production and consumption have gradually become subject to systemic distortions. As such, they also provide relevant context for the “by design” component of my definition (FN\*). While these overall trends are fairly unspecific, and are therefore not themselves sufficient for deeming their products ‘fake news’, they indicate, by way of example, weaknesses in the overall structure of news production that may then be exploited in order to implement mechanisms and substructures that are designed to mislead and manipulate.

One of the most distinctive features of contemporary news production and consumption is the significant acceleration of the news cycle. News no longer gets aggregated once a day, in the form of the ‘evening news’ on TV or in the morning edition of a newspaper; instead, it has become more pervasive and more fast-paced. It is now available—and updated—24/7. On the face of it, this might seem like a good thing. After all, did we not earlier find that *epistemic coverage*—that is, the condition of being updated by one’s informational environment about any relevant changes—is a necessary precondition for maintaining knowledge that we already possess? In theory that is true; but in practice, it may well be that, for

average consumers, constant bombardment with news (and *with talk about the news*) creates a feeling of uncertainty that itself undermines self-ascriptions of knowledge. An inflated sense of the changeability of those “impermanent states of affairs” that are the object of our knowledge may lead us to doubt whether our knowledge is, in fact, still accurate. Perhaps I missed a vital piece of information because, for a few hours, I didn’t tune into my news feed! It is also worth noting that, due to the 24/7 nature of online news in a competitive media environment, “online news sites today are emphasizing *recent* news stories over *relevant* or *important* news stories” (Chakraborty et al. 2015: 1); to compound matters further, updates are not necessarily reported—or prominently displayed—*when* they happen, but rather when they are likely to generate user engagement. As a result, “the topical composition of the information consumed by a user is effected by the user’s browsing patterns on online news media sites” (Chakraborty et al. 2015: 3), including by their temporal distribution. Instead of keeping us better informed, the 24/7 nature of online news may thus put us at the mercy of algorithms that optimize the ‘trending’ patterns of news stories, using their potential for user-engagement as the main metric, not their value in ensuring adequate epistemic coverage.

A second characteristic development—though one that varies considerably from country to country—has been the changing nature of media markets. Taking television news in the United States as an example, the decline of big nationwide (free-to-air) broadcasters, notably the ‘Big Three’—ABC, NBC, and CBS—was accompanied by fragmentation, due to the rise of cable channels, including (but not limited to) news channels such as CNN (which was founded in 1980 as the original ‘Cable News Network’):

In 1980, more than 90% of television viewers were tuned in to one of these three networks during prime time. By 2005, the season ending average prime-time share of the Big Three [= ABC, NBC, CBS] networks had fallen to 32%.

(Hindman & Wiegand 2008: 119)

Broadcasters quickly realized that the closest they could get to a monopoly would be to corner political segments: most famously, Fox News became the channel for right-wing conservatives, making it the most-viewed cable TV channel in the USA—thereby beating CNN—for the first time just five years after it was started in 1996. Much later, coinciding roughly with Barack Obama’s presidential campaign, MSNBC attempted, much less successfully, to replicate such a business model for liberal viewers. The US example is an especially extreme case among Western democracies, yet similar dynamics have unfolded elsewhere, not only in anglo-phone countries such as the UK and Australia, but also in countries such as Brazil and India. There is some evidence that, in countries with a strong presence of public broadcasters, the share of partisan news stories in circulation is smaller

(Humprecht 2019: 1982), but this effect is not uniform, and public broadcasters in many countries have lost ground and are facing stiff opposition (both from commercial competitors and, increasingly, from populist political movements).

Third, and this is perhaps the most prominently discussed novel trend, online social media and the tracking of users across different platforms has led to news not only being tailored to a certain political *segment* of society—which, after all, may still have considerable diversity within—but to *individuals*. When what we see online has been ‘curated’ by algorithms in such a way as to match our individual preferences as inferred from our online activity, we risk ending up in a ‘filter bubble’ (Pariser 2011) that only confirms our basic outlook and prior preferences. Further self-segregation of like-minded groups online, for example through closed groups on social media such as Facebook, can give rise to *echo chambers*, which are characterized not only by a lack of diverse viewpoints, but which encourage self-policing of groups for dissenting voices, which are then actively discredited and marginalized (see Nguyen 2020).

These are but three examples of *systemic changes* in the processes and mechanisms by which news content is created, collated, disseminated, and consumed. They are pervasive and affect anyone who uses the corresponding technologies, virtually regardless of one’s own (or other individuals’) prior intentions concerning their possible uses. While some of the observations above emphasize the role of online social media, it would be quite erroneous to think that the traditional news media are unaffected by these changes. Both print and broadcast media increasingly source stories from the online world: what is ‘trending’ on Twitter or Facebook may well become a newspaper headline the next day; TV news programmes now often include segments that summarize discussions from their corresponding social media channels; and news organizations increasingly merge their online and offline editorial staff into a single team. As the online and offline business models converge, one should expect spill-over effects, making it likely that no part of the media and information ecosystem we rely on will remain entirely immune from the various dynamics outlined above. This does not, of course, mean that every news source that operates against the backdrop of these changes—that is, *all* media—is thereby rendered suspect; rather, any news source wishing to maintain its integrity and trustworthiness in such an environment needs to take active steps to mitigate exploitation and to minimize the corresponding vulnerabilities.

Taking a step back, many of the systemic changes we are witnessing, including those changes in the production and presentation of news content outlined above, may be related to the—admittedly overused—term *attention economy*. As Herbert Simon noted, presciently, in 1971:

[I]n an information-rich world, the wealth of information means a dearth of something else: a scarcity of whatever it is that information consumes.

What information consumes is rather obvious: it consumes the attention of its recipients. (Simon 1971: 40)

In an environment where content providers are a dime a dozen (and, indeed, most information can be accessed for free), providers need to attract the audience's attention afresh every time. Alternatively, they may aim to instil new forms of allegiance and loyalty, e.g. by becoming hyper-partisan sources which not only offer information but, first and foremost, a sense of ideological belonging. On the part of consumers, the overabundance of information creates pressure to utilize 'quick and dirty' cognitive shortcuts, in order to keep the overall amount of information manageable. This, too, is readily obvious in such cases as 'clickbait' and online social media—which, in a memorable phrase, have been dubbed "cognitive biases on steroids" by the magazine *Psychology Today* (Braucher 2016).<sup>7</sup> If news providers, in an effort to monopolize attention, have begun catering to the very heuristics and cognitive biases that have been shown to modulate our reasoning, then this is all the more the case for fake news websites which, after all, need to compensate for their lack of an established readership.

As these examples make clear, social and technological trends have converged in ways that have led to systemic changes in the processes involved in creating and disseminating news and information. These create weaknesses and vulnerabilities—at both the structural and the individual level—that can be exploited by the kind of fake news that, in recent years, have become the focus of the public debate: deceptive, distorted, or dubious information, which often percolates through ideologically charged online echo chambers, before being sometimes picked up by more mainstream news outlets. Certain design features, especially of the main social network sites—such as the decontextualized way in which posted information appears on one's Twitter feed, or the ease with which it is possible, on Facebook, to confine oneself to 'Closed Groups' of like-minded individuals—feed into, and at the same time encourage, these trends. In the most egregious cases, these are then actively exploited in order to spread manipulative pseudo-information; yet even when there is no specific intention to deceive, mere reliance on processes of content production that are so shoddily designed as to foreseeably result in systemically misleading output, may well cross the boundary into 'fake news' territory. The relentless race for attention, and the systemic exploitation of the very biases and heuristics that consumers utilize in order to manage the onslaught of information they are faced with, create conditions that render individual reasoners vulnerable to accepting false or misleading claims—all the more, when these confirm deeply held convictions or help reduce cognitive dissonance (see also Gelfert 2013). Even the political weaponization of the term 'fake news' as a general put-down for critical

<sup>7</sup> For a more extensive discussion of cognitive biases and heuristics in relation to fake news, see Gelfert 2018.

news sources fits this pattern, since it is designed to systematically shift audience's patterns of trust, e.g. through simple repetition of such phrases as 'mainstream media' or the 'lying *New York Times*', which actively utilize priming and repetition effects in an effort to undermine trust in sources that, by most standards, would normally have a valuable role to play in ensuring epistemic coverage.

## 5. Knowledge Maintenance in an Age of Fake News

If, as I have argued, the convergence of current political, social and technological trends renders our information environment more susceptible to *systemic* distortions—and does so in novel ways that merit the application of the new(ish) label 'fake news' to a subset of cases enabled by them—one might very well wonder what can be done about it. Specifically, one might ask: what can *I do*, as an individual, to safeguard my own capacity for knowledge-maintenance—that is, to minimize exposure to falsehoods and to guarantee sufficient coverage so as to ensure that I can continue to trust what I once knew? And what should we, collectively, aim for in order to stem the tide of fake news in our respective communities?

Existing proposals so far have tended to focus either solely on the individual responsibilities of consumers (at the expense of the systemic aspects of our informational environments) or on technological tweaks to existing platforms (thereby largely ignoring the perspective of the individual epistemic agent). While there is a lively debate about whether, and how, online media markets should be shaped and regulated—e.g. by forcing online providers such as Facebook to take down false claims, or by strengthening public broadcasters—I shall leave these more global questions aside in this chapter. Many of the more general recommendations, while eminently plausible, suffer from a lack of specificity, which makes it difficult to estimate their overall effect on the production and circulation of fake news. For example, diversity of sources is often held up as a gold standard, yet, as the example of fringe websites disseminating fake news demonstrates, a mere multiplicity of online sources run by various groups and individuals does not ensure high-quality information. Similarly, proposals to encourage 'fact-checking' as a service provided by public broadcasters have largely fallen flat, since "the consumption of fact-checks is concentrated among non-fake news consumers, so they often do not reach their intended audience" (Brown 2018: 211).

One proposal—frequently voiced (perhaps self-interestedly?) by philosophers—has been to stem the tide of fake news by training consumers to be less gullible and to engage in more 'critical thinking'. On the one hand this reflects the sentiment, encountered earlier in our discussion of testimonial reductionism, that "we know too much about human nature to want to trust anyone, let alone everyone, uncritically" (Fricker 1995: 400); on the other hand, it is often accompanied by

complaints about the decline of critical thinking in civic education, which is said to leave “young people [...] not only less informed than we might have expected, but [...] also less interested in applying what little they might have learned to their responsibilities as citizens” (Nichols 2017: 138). By contrast, critical thinking is said to “confer significant benefits on students, benefits that could prove to be powerful tools against Truth Decay” (Kavanagh & Rich 2018: 135)—after all, as Lee McIntyre succinctly puts it, “when we are looking for the truth, critical thinking, skepticism, and subjecting our ideas to the scrutiny of others works better than anything else” (McIntyre 2018: 59–60).

Notwithstanding their initial plausibility, such homilies in support of critical thinking hardly constitute an adequate response to the current problem of fake news—not least since it would seem to require waiting for generational change, once curricula have been revamped accordingly. They also seem to get things backwards: as has been shown empirically for the 2016 US elections, older users—those over 65—were more likely to share fake news on Facebook than younger users, even when other characteristics, such as education, ideology, and partisanship, were held constant (see Guess et al. 2019: 1). What is more, one of the psychological drivers of fake news acceptance—motivated cognition (i.e., the reverse-engineering of assessments of evidence, such that one’s deeply held convictions come out intact)—is heightened, at least in some contexts, by a high level of education (see e.g. Hamilton 2011). As it turns out, well-educated individuals are especially prone to using their argumentative skills to convince themselves that they—and their in-group—are right.

One of the main problems in addressing fake news consists in the fact that, as ample research in social psychology has shown, misinformation—once encountered and processed—is not easily corrected. Even when factual corrections have been received, and have been acknowledged as such by experimental subjects, the original misinformation continues to linger and exert an effect, since people tend to keep relying on debunked falsehoods (Nyhan & Reifler 2010). At the same time, fake news often spreads faster through social networks and penetrates more deeply than competing truthful information (Vosoughi et al. 2018), making it unlikely that ‘fact-checking’ and ‘debunking’ alone will be effective in halting the spread of fake news. This has led to efforts to explore the possibility of interventions aimed at pre-emptively refuting anticipated fake news, a process that is sometimes called “prebunking” (Cook et al. 2017: 4). The idea is (1) to issue an explicit warning *that* a false or misleading argument or claim is about to be encountered, and (2) to offer an explanation as to *how* the technique that is employed is misleading or fallacious. The idea is “that people can be ‘inoculated’ against misinformation by being exposed to a refuted version of the message beforehand” (Cook et al. 2017: 4). Furthermore, by repeatedly exposing people to weakened versions of problematic claims, along with their pre-emptive refutations, it appears that “attitudinal resistance can be conferred against future

deception attempts" (Roozenbeek & van der Linden 2019: 2). The immunological metaphor is quite deliberate: "just as injections containing a weakened dose of a virus can trigger antibodies in the immune system to confer resistance against future infections, the same can be achieved with information by cultivating mental antibodies against misinformation" (Roozenbeek & van der Linden 2019: 2).

Likening our (desired) response to fake news to how our immune system responds to antigens is, of course, a wild analogy. We have no reason to believe that our belief-forming mechanisms operate—causally or functionally—anything like the B-cells and T-cells of our immune system. It would also be entirely conceivable that, with frequent exposure to fake news, we simply become *habituated* to such cases and pay less attention to them—similar to how we adapt to our auditory environments by screening out background noises that we have become used to. As Davis puts it, "messages that are potent in some contexts become normalised and then impotent in others—it all depends on whether they are externally validated" (Davis 2017: 226); as a result, the credibility of fake news sources can be expected to diminish over time. *If*, indeed, "...the empirical result holds up that properly framed..." 'prebunking' can confer some degree of resistance to subsequent fake news, then this is certainly something worth exploring. As things stand, however, we can only say this much: once encountered in an uncontrolled way, fake news can have a lasting detrimental effect on our belief system, which cannot easily be reversed later on. *Mutatis mutandis, not encountering* fake news—e.g. by avoiding conditions where one is likely to encounter, or be influenced by, it—is perhaps the best defence mechanism.

Interestingly, it is again medicine—specifically, the history of infection prevention and control—which provides an interesting analogy, though this time not for any alleged causal similarity, but because the historical example shows how risks can be mitigated even in situations of uncertainty. In the nineteenth century, many young mothers, even while in hospital, were afflicted by puerperal fever which often struck shortly after childbirth: "A woman could be delivered on Monday, happy and well with her newborn baby on Tuesday, feverish and ill by Wednesday evening, delirious and in agony [...] on Thursday, and dead on Friday or Saturday" (Loudon 1992: 54). Ignaz Semmelweis, who, in 1846, had assumed the position of assistant at one of the two maternity clinics that were part of the Vienna General Hospital, developed an interest in this phenomenon, for which no specific causes were known, and noted a curious pattern: one of the wards, where deliveries were carried out by highly trained medical students, had a much higher incidence of puerperal fever than the other one, which was run by midwives. Through a series of contrastive questions, he was able to rule out certain causes—such as the miasmatic hypothesis which blamed the illness on influences communicated via the air (since both wards were in the same location)—eventually arriving at the hypothesis "that the medical students, who would perform their examinations [of the young mothers] having come directly

from working on cadavers, carried some sort of ‘cadaveric matter’ on their hands, and that this was the source of infection” (Gregg 1993: 287). In order to test his hypothesis, from some time in 1847 onwards, Semmelweis required students and doctors to disinfect their hands with chlorinated lime before examining patients—a move that promptly reduced the relative number of cases of puerperal fever significantly (by about two-thirds, when averaged over several years). Semmelweis’s recommendation preceded any detailed causal understanding of what brought on puerperal fever; rather, by asking a series of contrastive questions in the absence of any detailed theory about the phenomenon, Semmelweis was nonetheless able to recommend procedural changes that saved lives.<sup>8</sup> Outside Vienna, Semmelweis’s recommendations, sadly, were not adopted until much later. Yet, once they were accepted, they quickly became *stable routines* that reduced the risk of transmission and infection, regardless of the specific nature of the pathogen and of the patient’s condition.

It is this *routinization* of a preventive measure which holds an important lesson also for how we can hope to change our epistemic habits for the better. Recall Levi’s description of routine belief expansion above, which had us accept the outcome of a belief-forming process without having to resort to “finding out what that outcome is and combining it with the rest of our knowledge to decide upon an optimum” (Levi 1983: 39). The same, of course, goes for withholding trust—and even more so for steering clear of certain situations that likely expose us to biased, deceptive, or manipulative information. Committing to a routine—whether it is disinfecting one’s hands before seeing another patient, regardless of any specific risk assessment, or trusting a reliable source without scouring the web for potentially dissenting information of dubious origins—is a way of *reducing complexity*. In this sense, epistemic routines achieve what, notoriously, epistemic biases and heuristics also achieve: they reduce the mental effort required to make choices and decisions. Advocating reliance on epistemic routines might make proponents of testimonial reductionism and ‘critical thinking’ shriek in horror: would this not amount to a defence of ‘cognitive laziness’, which, in turn, would result in a blank cheque for mental stagnation? Isn’t the worry precisely that, when it comes to fake news, individuals may become stuck in filter bubbles, such that we should demand they actively seek out alternative sources, rather than mindlessly follow their bad habits?

While these worries are all too understandable, they ignore an important point: whereas cognitive biases and heuristics operate largely automatically, epistemic routines can often be deliberately chosen—not anew every time we form a belief, of course, but at the point of adoption—and can also be periodically reviewed and

<sup>8</sup> I am grateful to Katherine Furman for highlighting the fact that Semmelweis proceeded in the absence of an established theory, showing that routines do not always require full theoretical understanding.

adjusted in response to changes in our informational environment. This is why a focus on epistemic routines also has the potential of overcoming the dichotomy found in current proposals, which either emphasize the individual responsibility of reasoners or the implementation of technological tweaks. Epistemic agents may well be held responsible for the epistemic routines they commit themselves to; however, in order for them to be able to make responsible choices, predictable effects of their choices must be made transparent wherever this is technologically possible. Arguing in favour of epistemic routines, including the habitual reliance on trusted sources, is thus not the same as advocating blind trust in any one source. On the contrary: sticking with epistemic routines that one has committed to may be thought of as a precondition for acquiring any meaningful sort of inductive track record, which may then be used to reassess the very choices one initially made. In the past, the need for such choices was partly created by systemic features of how news was delivered, e.g. by the subscription model of newspaper delivery, which, on the one hand, forced readers to stick with a (potentially one-sided) ‘news diet’, but on the other hand allowed them to acquire substantial knowledge of their preferred newspaper, its columnists, biases, and even a sense of its target readership. Just as readers are free to cancel their newspaper subscription, for example because over time their assessment of the newspaper’s biases has changed or its coverage has deteriorated, an agent who is following a certain epistemic routine can, on occasion, choose to revise it.<sup>9</sup> The present media environment, with its abundance of available sources and a very low threshold for accessing them, discourages the pre-selection of sources, encouraging instead a non-committal ‘browsing’ of a variety of news outlets, whose quality and track record often evade us.

It is worth emphasizing that commitment to epistemic routines goes beyond one’s choice of trusted news sources. If, for example, it becomes clear—and the empirical evidence appears to point in that direction—that being exposed to prejudiced views in the comments section of an online article influences one’s own interpretation in prejudiced ways (see Hsueh et al. 2015), one might resolve to make it a habit to ignore those sections (or only seek them out on rare occasions). News sites can facilitate this by displaying such sections less prominently. If I find myself becoming more gullible when tired or stressed, I might restrict my news intake to times of the day when I am alert and not in a rush. I might even adopt routines that *counteract* known cognitive biases, e.g. by separating emotionally charged contexts from information-gathering activities, or that may help bypass known flaws in the algorithmic ‘curation’ of online information. Thus, Boaz Miller and Isaac Record have suggested that we should actively seek out individuals outside of our ‘bubble’ on social media, “casually visit

<sup>9</sup> My usage here differs from Levi (1983), who explicitly subsumes ‘hard-wired’ belief-forming mechanisms, e.g. sense perception, under ‘routine belief expansion’.

their Facebook profiles and see whether they have posted an interesting story that the automatically generated news feed missed" (Miller & Record 2013: 130). This way, we might even succeed in partially 'retraining' algorithms in such a way as to improve our *own* epistemic coverage.

Whatever the merits of individual suggestions like these, identifying successful epistemic routines—and analysing their dependence on contingent features of the informational and media environments that epistemic agents find themselves in—should be a core goal of an applied social epistemology that actively seeks to improve our collective epistemic position. In a broadly empiricist spirit, we may take inspiration from Semmelweis and encourage others to adopt epistemic routines that have been found to render us less vulnerable to the dysfunctions identified above—even before we have a full understanding, or a complete theory, of how they work. It is only by simultaneously *analysing* and *navigating* the interplay between, on the one hand, the systemic features of our informational and environment and, on the other hand, our individual actions and responses to them, that a full appreciation of our current collective epistemic predicament—as well as any genuine prospect of improving it—can emerge.

## References

Aikin, Scott F., and Talisse, Robert B., 2018: "On 'Fake News'", 21 May, *3QuarksDaily.com*, <https://www.3quarksdaily.com/3quarksdaily/2018/05/on-fake-news.html>, accessed 14 January 2020.

Beaujon, Andrew, 2019: "Trump Claims He Invented the Term 'Fake News'—Here's an Interview With the Guy Who Actually Helped Popularize It", 2 October, *Washingtonian*, <https://www.washingtonian.com/2019/10/02/trump-claims-he-invented-the-term-fake-news-an-interview-with-the-guy-who-actually-helped-popularize-it/>, accessed 14 January 2020.

Braucher, David, 2016: "Fake News: Why We Fall for It", 28 December, *PsychologyToday.com*, <https://www.psychologytoday.com/blog/contemporary-psychology-analysis-in-action/201612/fake-news-why-we-fall-it>, accessed 18 January 2020.

Brown, Étienne, 2018: "Propaganda, Misinformation, and the Epistemic Value of Democracy", *Critical Review*, vol. 30, no. 3–4, pp. 194–218.

Chakraborty, Abhijnan, Ghosh, Saptarshi, Ganguly, Niloy, and Gummadi, Krishna P., 2015, *Can Trending News Stories Create Coverage Bias? On the Impact of High Content Churn in Online News Media*, Computation and Journalism Symposium, <http://cj2015.brown.columbia.edu/papers/trending-news.pdf>, pp. 1–5.

Coady, David, 2019: "The Trouble With 'Fake News'", *Social Epistemology Review and Reply Collective*, vol. 8, no. 10, pp. 40–52.

*Colbert Report, The*, 2005: “The Word – Truthiness” (video clip), 17 October, <http://www.cc.com/video-clips/63ite2/the-colbert-report-the-word—truthiness>, accessed 14 January 2020.

Cook, John, Lewandowsky, Stephan, and Ecker, Ullrich, 2017: “Neutralizing Misinformation Through Inoculation: Exposing Misleading Argumentation Techniques Reduces Their Influence”, *PLoS ONE*, vol. 12, no. 5, pp. 1–21.

Davis, Evan, 2017: *Post-Truth: Peak Bullshit and What We Can Do About It*. London: Abacus.

Dentith, M. R. X., 2017: “The Problem of Fake News”, *Public Reason*, vol. 8, no. 1–2, pp. 65–79.

Dentith, M. R. X., 2018: “What is Fake News?”, *University of Bucharest Review*, vol. 8, no. 1 (new series), pp. 24–34.

Fallis, Don, and Mathiesen, Kay, 2019: “Fake News is Counterfeit News”, *Inquiry*, <https://doi.org/10.1080/0020174X.2019.1688179>.

Feest, Uljana, 2012: “Exploratory Experiments, Concept Formation, and Theory Construction in Psychology”, in Uljana Feest and Friedrich Steinle (eds), *Scientific Concepts and Investigative Practice*. Berlin: de Gruyter, pp. 167–89.

Frankfurt, Harry, 1995: “On Bullshit”, *Raritan Quarterly Review*, vol. 6, no. 2, 1986, pp. 81–100.

Fricker, Elizabeth, 1995: “Telling and Trusting: Reductionism and Anti-Reductionism in the Epistemology of Testimony”, *Mind*, vol. 104, no. 414, pp. 393–411.

Fuller, Steve, 2018: *Post-Truth: Knowledge as a Power Game*. London: Anthem.

Gelfert, Axel, 2013: “Climate Scepticism, Epistemic Dissonance, and the Ethics of Uncertainty”, *Philosophy and Public Issues* vol. 3, no. 1 (new series), pp. 167–208.

Gelfert, Axel, 2014: *A Critical Introduction to Testimony*. London: Bloomsbury.

Gelfert, Axel, 2018: “Fake News: A Definition”, *Informal Logic*, vol. 38, no. 2, pp. 84–117.

Gelfert, Axel, 2019: “Beyond the ‘Null Setting’: The Method of Cases in the Epistemology of Testimony”, *Epistemology & Philosophy of Science*, vol. 56, no. 2, pp. 60–76.

Gettings, Michael, 2007: “The Fake, the False, and the Fictional: *The Daily Show* as News Source”, in Jason Holt (ed.), *The Daily Show and Philosophy*. Oxford: Blackwell, pp. 16–27.

Goldberg, Sanford, 2010: *Relying on Others: An Essay in Epistemology*. Oxford: Oxford University Press.

Govier, Trudy, 1993: “Needing Each Other for Knowledge: Reflections on Trust and Testimony”, in Eric Krabbe, Renée Dalitz, and Pier Smit (eds), *Empirical Logic and Public Debate: Essays in Honour of Else M. Barth*. Amsterdam: Rodopi, pp. 13–26.

Gregg, Kevin, 1993: “Taking Explanation Seriously; or, Let a Couple of Flowers Bloom”, *Applied Linguistics*, vol. 14, no. 3, pp. 276–94.

Guess, Andrew, Nagler, Jonathan, and Tucker, Joshua, 2019: "Less Than You Think: Prevalence and Predictors of Fake News Dissemination on Facebook", *Science Advances*, vol. 5, no. 1, pp. 1–8.

Habgood-Coote, Joshua, 2018: "Stop Talking About Fake News!", *Inquiry*, vol. 62, no. 9–10, pp. 1033–65.

Hamilton, Lawrence, 2011: "Education, Politics and Opinions about Climate Change: Evidence for Interaction Effects", *Climatic Change*, vol. 104, pp. 231–42.

Hindman, Douglas B., and Wiegand, Kenneth, 2008: "The Big Three's Prime-Time Decline: A Technological and Social Context", *Journal of Broadcasting & Electronic Media*, vol. 52, no. 1, pp. 119–35.

Holton, Richard, 1994: "Deciding to Trust, Coming to Believe", *Australasian Journal of Philosophy*, vol. 72, no. 1, pp. 63–76.

Hsueh, Mark, Yogeeswaran, Kumar, and Malinen, Sanna, 2015: "Leave Your Comment Below: Can Biased Online Comments Influence Our Own Prejudicial Attitudes and Behaviors?" *Human Communication Research*, vol. 41, no. 4, pp. 557–76.

Humprecht, Edda, 2019: "Where 'Fake News' Flourishes: A Comparison Across Four Western Democracies", *Information, Communication & Society*, vol. 22, no. 13, pp. 1973–88.

Jaster, Romy, and Lanius, David, 2018: "What is Fake News?", *Versus*, vol. 127, no. 2, pp. 207–24.

Kavanagh, Jennifer and Rich, Michael D. 2018, *Truth Decay: An Initial Exploration of the Diminishing Role of Facts and Analysis in American Public Life*. Santa Monica, CA: RAND Corporation.

Klein, David O., and Wueller, Joshua R., 2017: "Fake News: A Legal Perspective", *Journal of Internet Law*, vol. 20, no. 10, pp. 5–13.

Levi, Isaac, 1983: *The Enterprise of Knowledge: An Essay on Knowledge, Credal Probability, and Chance*. Cambridge, MA: MIT Press.

Levi, Lili, 2018: "Real 'Fake News' and Fake 'Fake News'", *First Amendment Law Review*, vol. 232, pp. 232–327.

Levy, Neil, 2017: "The Bad News About Fake News", *Social Epistemology Review and Reply Collective*, vol. 6, no. 8, pp. 20–36.

Loudon, Irvine, 1992: *Death in Childbed*. Oxford: Oxford University Press.

Lovett, Jon, 2013: Graduation Speech at Pitzer College, 18 May, <https://whatrocks.github.io/commencement-db/2013-jon-lovett-pitzer-college/>, accessed 14 January 2020.

McDowell, John, 1994: "Knowledge by Hearsay", in Bimal K. Matilal and Arindam Chakrabarti (eds), *Knowing from Words: Western and Indian Philosophical Analysis of Understanding and Testimony*. Dordrecht: Kluwer, pp. 195–224.

McIntyre, Lee C., 2018: *Post-Truth*. Cambridge, MA: MIT Press.

Miller, Boaz, and Record, Isaac, 2013: "Justified Belief in a Digital Age: On the Epistemic Implications of Secret Internet Technologies", *Episteme*, vol. 10, no. 2, pp. 117–34.

Montgomery-McGovern, J.B, 1898: "An Important Phase of Gutter Journalism: Faking", *Arena*, vol. 19, no. 99, pp. 240–53.

Mukerji, Nikil, 2018: "What is Fake News?", *Ergo: An Open Access Journal of Philosophy*, vol. 5, pp. 923–46.

Nguyen, C. Thi, 2020: "Echo Chambers and Epistemic Bubbles", *Episteme*, vol. 17, pp. 141–61.

Nichols, Tom, 2017: *The Death of Expertise: The Campaign Against Established Knowledge and Why It Matters*. New York: Oxford University Press.

Nyhan, Brendan, and Reifler, Jason, 2010: "When Corrections Fail: The Persistence of Political Misperceptions", *Political Behavior*, vol. 32, no. 2, pp. 303–30.

Pariser, Eli, 2011: *The Filter Bubble: What the Internet is Hiding From You*. London: Penguin.

Pepp, Jessica, Michaelson, Eliot, and Sterken, Rachel K., 2019: "What's New About Fake News?", *Journal of Ethics and Social Philosophy*, vol. 16, no. 2, pp. 67–94.

Rini, Regina, 2017: "Fake News and Partisan Epistemology", *The Kennedy Institute of Ethics Journal*, vol. 27, no. 2S, pp. E-43–E-64.

Roozenbeek, Jon, and van der Linden, Sander, 2019: "Fake News Game Confers Psychological Resistance Against Online Misinformation", *Palgrave Communications*, vol. 5, no. 65, pp. 1–10.

Silverman, Craig, and Lawrence Alexander, 2016: "How Teens in the Balkans are Duping Trump Supporters with Fake News", *BuzzFeed*, 3 November, <https://www.buzzfeed.com/craigslister/maiden-macedonia-became-a-global-hub-for-pro-trump-misinfo>, accessed 18 January 2020.

Simon, Herbert, 1971: "Designing Organizations for an Information-Rich World", in Martin Greenberger (ed.), *Computers, Communications, and the Public Interest*. Baltimore, MD: Johns Hopkins University Press, pp. 37–52.

Tandoc, Edson, Lim, Zheng Wei, and Ling, Richard, 2018: "Defining 'Fake News': A Typology of Scholarly Definitions", *Digital Journalism*, vol. 6, no. 2, pp. 137–53.

Vosoughi, Soroush, Roy, Deb, and Aral, Sinan, 2018: "The Spread of True and False News Online", *Science*, vol. 359, no. 6380, pp. 1146–51.

# 15

## Trust No One?

### The (Social) Epistemological Consequences of Belief in Conspiracy Theories

*Michael Baurmann and Daniel Cohnitz*

#### 0. Introduction

The National Security Agency (NSA) spies on us. Before Edward Snowden leaked classified information in 2013, which confirmed this claim, many would probably have shrugged it off as a “mere” conspiracy theory. What about now? Is the theory that the NSA spies on us still a conspiracy theory, now that it is a widely held (and apparently well-evidenced) belief?

It seems common to think that it’s not. Accordingly, that Caesar was murdered by a conspiracy of Roman senators, or that 9/11 was the outcome of a conspiracy among members of al-Qaeda does not make these historical accounts conspiracy theories. For many, the latter requires that there is an element of speculation, perhaps paranoia in the belief of such theory.

However, most philosophers who work on conspiracy theories disagree with that common understanding of the term. They find it hard to identify features that make conspiracy theories an intrinsically bad explanation type, in part because some initially suspicious conspiracy theories (like, perhaps, the theory that the NSA is spying on us) later turned out to be true, in part because the deficient features of some stereotypical conspiracy theories are not shared by other stereotypical conspiracy theories. Instead, these philosophers argue that “conspiracy theory” should be defined widely: *a conspiracy theory is the explanation of an event that cites conspiring agents as a salient cause* (Dentith 2014). Consequently, we are all conspiracy theorists. Everyone who believes that some historical event came about thanks to the successful secret collaboration of several individuals believes in a conspiracy theory and thus is a conspiracy theorist, and surely everyone believes this of some event.

Since some of these accounts are true and known to be true (e.g. that the assassination of Caesar was due to a conspiracy), believing in a conspiracy theory as such can’t be irrational or misguided. In principle, then, there is nothing wrong with conspiracy theories or belief in such theories. Of course, sometimes

conspiracy theories are mistaken and sometimes they are believed on the basis of insufficient evidence, but that is the possible fate of every theory (Dentith 2017). There is nothing that makes conspiracy theories particularly irrational, doubtful, or fishy, just because they are conspiracy theories.

Accordingly, attempts by psychologists and sociologists to investigate the psychological and social profile of conspiracy believers might be seen as nothing but a witch-hunt. In a recent public statement, a group of social epistemologists and sociologists even argues that such witch-hunt endangers our (development towards an) open society (Basham and Dentith 2016, 13):<sup>1</sup>

[W]e believe that it is not conspiracy theorizing that is the danger, but rather the pathologizing response to conspiracy theories.

The antidote to whatever problems conspiracy theories present is vigilance, not some faux intellectual sophistication which dismisses conspiracy theories out of hand. It's really quite simple when you think about it: conspiracy theorizing is essential to the functioning of any democracy, or indeed any ethically responsible society.

The argument behind it is that conspiracy theorizing keeps the public in critical control of the people in power and might prevent the latter from doing serious harm. Such critically minded citizens should be interested in developing an even more open society with institutions that exercise mutual control, one might add, because that's what makes conspiring much harder.

Consider our opening paragraph again. Some years ago people who'd have claimed that the NSA spies on us would have been ridiculed as conspiracy theorists, while in fact they were right. We shouldn't be critical of conspiracy theorists, because if we had taken their scepticism seriously, we might have learned much sooner that the NSA is spying on us. Perhaps more of that conspiratorial scepticism would have been better for our society, because it could have strengthened democratic institutions (for example, institutions that control the NSA). An argument of that kind is suggested in (Clarke 2002, 148):

The prevalence of conspiracy theories confers a third benefit upon us, which is that it helps to maintain openness in society. Government agencies have a tendency to be less than forthcoming with information that might prove embarrassing to them but that the public would prefer to have made available. The information gathering activities of conspiracy theorists can help to prevent such secretiveness.

<sup>1</sup> The cited paper is signed at the end by Matthew Dentith, Lee Basham, David Coady, Ginna Husting, Martin Orr, Kurtis Hagen, and Marius Raab.

So, is conspiracy theorizing not actually a danger to our current political system, but rather a force for good?

We believe that these philosophers and sociologists are right in thinking that the problem with (certain) conspiracy theories is not their explanation *type*, and that the fault of conspiracy theories needs to be identified on a case by case basis in the many ways in which people make mistakes when theorizing. But from that it doesn't follow that unleashed conspiracy theorizing in a society and a general conspiratorial scepticism are forces for the good or that we should welcome them in the interest of an open society and its institutions.

On the contrary, indiscriminate and pervasive conspiracy theorizing is a danger to the institutions of an open society, and this can be shown on the basis of social epistemological considerations alone.

In any case, it can already be made plausible on the basis of empirical evidence. We just need to take a look around at countries that were on a path to open, democratic societies with separation of power, freedom of speech, etc. and in which conspiracy theories have played a significant role in political campaigns that led to political change. The examples we have in mind are Turkey, Hungary, Poland, and the USA. In all these cases, the political change induced was then not at all towards a general strengthening of the institutions of open societies so that these could better exercise mutual control. On the contrary, the change was towards a mutilation of these institutions and a development away from an open society towards a closed society that displays elements of an autocracy.

Now, obviously, in all these cases there is a variety of factors that came together and led to the particular political development. We don't want to argue that it is only or even primarily due to conspiracy theorizing that these countries got off the path to an open society. But we do want to argue that conspiracy theorizing has been a causal factor in this process. There is a social epistemological explanation for the turn these societies took.

In Section 1 of this chapter, we will briefly revisit the discussion over the nature of conspiracy theories as such. We will argue that even if there is no simple definition of conspiracy theories as an explanation type that would entail that conspiracy theories are always deficient theories (and thus irrational to believe), this is still a far cry from having in any sense vindicated conspiratorial thought, let alone the belief in conspiracy theories in Western democracies.

In Section 2, we will characterize the social-epistemological predicament that individuals in a modern, complex society find themselves in, and how they depend on relatively stable trust-networks in order to benefit from the knowledge that is generated by the institutions of these societies.

In Section 3, we will show that belief in false conspiracy theories disrupts such trust networks and detaches the conspiracy theorist effectively from the knowledge sources in her social environment. Instead of a wide trust-network, she will, typically, be left with only a few personal trust relations, relations that can—and

very often are—exploited by enemies of open societies. Hence, in contrast to what many philosophers and sociologists who work on conspiracy theories seem to believe, conspiracy theories as such do not serve an important and in any sense positive function in open societies.

In Section 4, we will consider objections to our analysis. Is it really the case that conspiracy thinking is only of possible negative impact for open societies? Aren't there also benefits that we have overlooked? We will argue that it is difficult to contain the scepticism that conspiracy theories encourage. Once that scepticism begins evolving, spreading false conspiracy theories become a serious threat to the institutions of open societies and we need to find strategies to diminish their destructive influence. To conclude, in Section 5, we will discuss some options of what such strategies could look like.

## 1. Are Conspiracy Theories always Irrational?

As we said above, ordinary usage of the term 'conspiracy theory' and also much of academic usage of the term (at least beginning with Hofstadter 1965) implies that conspiracy theories are false and irrational to believe. Labelling a belief a conspiracy theory expresses that the belief is not worth being taken seriously and only held on irrational, presumably paranoid grounds. Accordingly, people who try to defend the view that a certain event or phenomenon (say, 9/11 or the frequency and duration of contrails) is due to a conspiracy often make their point by emphasizing that their belief is *not* a conspiracy theory.

Furthermore, conspiracy theories are widely considered to be a fringe phenomenon and why people believe such theories is a matter for psychologists to find out. Many psychologists who work on conspiracy thinking seem to agree that conspiracy theories can only be held irrationally, since they almost never inquire into the reasons for why their participants have conspiracy beliefs and instead immediately look for psychological profiles which anyone with a conspiracy belief would share (Cohnitz 2018).

But this attitude seems ill-founded. First of all, for a long time conspiracy theories used to be widely held and were considered a legitimate way of making sense of the social world. As Butter (2018) argues, it is only since roughly the 1960s that conspiracy theories became marginalized and disappeared from mainstream social discourse in Western societies.<sup>2</sup> Thus, unless one is prepared to

<sup>2</sup> According to Butter, conspiracy theories presuppose certain assumptions about the effectiveness of human action, a certain understanding of time, and a public in which they can be disseminated via texts or other media. We can then find early conspiracy theories in ancient Athens and Rome and a conspiracy culture starting from the sixteenth century onwards (Butter 2018). At around that time the terms "conspiratio" and "conspiracy" become important elements of political discourse (Zwierlein and de Graaf 2013). Conspiracy theories remained influential until far into the twentieth century.

defend that quite recently a major cognitive development took place, conspiracy theories can't be *that* irrational. And, indeed, typical conspiracy theories don't seem to be internally inconsistent or incoherent.

This would still allow that conspiracy theories are now only irrationally believable, because they rest on an assumption that everyone (in Western societies) since the 1960s knows to be false. In that way, conspiracy theories could be like miracles: for a long time, miracles (interventions into the course of nature) were rationally believable. But, as Hume argued,<sup>3</sup> they ceased to be believable with the advent of modern science and our knowledge of the empirical support that the (exceptionless) laws of nature enjoy. But what is it that we all learned around the 1960s that made belief in conspiracy theories irrational?

Michael Butter (2018) picks up an idea—often attributed to Karl Popper—that we learned from modern sociology that social events can't be the result of successful (large-scale) conspiracies. The social world is highly complex and difficult to control. Our plans seldom come out as intended and most larger social events or phenomena—even if they look as if they were designed and intended—are typically just unintended consequences of intentional actions. Thus, for Friedrich Hayek (1967) and Karl Popper (1966), a central explanation type of the social sciences are *invisible hand explanations*. Invisible hand explanations explain macro-level events and phenomena that seem intended and planned (perhaps due to their stability or their apparent optimality) as the result of intentional action at the micro level that did not aim at bringing the phenomenon or event in question about (Ullmann-Margalit 1978).

For Michael Butter, conspiracy theories (properly so called) always involve complexities (several groups of conspirators and the interaction between them) that are just practically impossible to control. Hence all conspiracy theories are false. “Real” conspiracies, in contrast, have a limited amount of conspirators and are short-lived.

But although there is certainly a substantial grain of truth in the idea that conspiracy theories get the more implausible the more they require a great deal of coordination and loyalty among a large and diverse group of individual agents, this seems to be a matter of degree and thus unfit to serve as a defining feature of ‘conspiracy theory’ that would allow the conclusion that conspiracy theories are likely to be wrong, regardless of the specific circumstances.

Under which conditions exactly postulates a theory “too much complexity” in order to count as a conspiracy theory? This seems to depend on a variety of

A prominent example is the idea that the *Illuminati* or the Freemasons orchestrated the French Revolution.

<sup>3</sup> What Hume precisely argued is a matter of debate. For the range of alternative interpretations, see McGrew (2019). We merely use Hume's argument as an illustration here; we are neither committed to the correctness of our interpretation of Hume, nor to whether this is actually a good argument to establish the irrationality of belief in miracles.

contextual factors. For example, we are quite confident that a conspiracy between more than ten first-graders would be very short-lived (and collapse as soon as you bring in the candy), while a conspiracy between, say, fifty well-selected CIA agents can probably last considerably longer. Since the factors that would require specification are too numerous, a definition that would guarantee that all conspiracy theories are likely to be false, would have to read as follows:

A *conspiracy theory* is an explanation that cites the secret collaboration of a group of agents as a salient cause, where the complexity of social coordination that is required in order to bring the explanandum about is too great to be plausible.

However,—as we just explained—what counts as “too great to be plausible” depends on a variety of empirical factors. What is the loyalty of the conspirators supposed to be grounded in? Can the conspirators be controlled by means other than appeal to their self-interest? How many of the conspirators need to know the “whole picture” and for how many conspirators is it sufficient that they only know their part of the plan? To what extent can the number of conspirators be kept to a small circle by using new (and not yet widely known) technology? How easy is it to “corrupt” controlling institutions?

Answers to these questions depend on the details of the conspiracy theory and the structure of the society in which it is supposed to apply. Some conspiracy theories will emerge as outright implausible for the reasons that Butter identifies, but for some conspiracy theories opinions might diverge about their plausibility. In these cases, whether an explanation should count as a “conspiracy theory”, defined in the way above, would depend on matters other than the explanation type; it will depend on empirical questions that are perhaps not widely known and that are independent of the insights of Hayek and Popper.

Hence the definition above is perhaps a good approximation to the ordinary language meaning of ‘conspiracy theory’ but it is of little use for theoretical and empirical purposes, especially if the aim of the empirical work is to find out why people believe conspiracy theories and whether anything can be done about it. It seems, then, indeed more fruitful to use a wide notion of conspiracy theory (i.e. *an explanation that cites the secret collaboration of a group of agents as a salient cause*) and look at the details of the theory and the evidence provided for it, in order to assess its plausibility and likelihood.

If such a wider definition is used, then it is clear that not all conspiracy theories are irrationally believed. Of course, conspiracy theories can be unsubstantiated because they may not consider the plausibility of alternative causes for an outcome like invisible hand explanations, the possibility of coincidental relations, or unintentional failures of institutional processes. Deficient conspiracy theories may also be immunized against contradictory evidence by ad hoc assumptions or extensions of the scope of the alleged conspiracy. In short, conspiracy theories may fail

the criterion to suggest a theory that delivers the best explanation available in light of the known facts.

But, as in the case of other theories, whether a conspiracy theory is in this sense unsubstantiated is open to critical examination and consideration. And, as we have seen in the case of Snowden, the result of this check can be that a conspiracy theory is indeed proved to be the best explanation for certain events. So there seems to be a real chance that not only can conspiracy theories be true but that we also can successfully differentiate between true and false conspiracy theories.

What implications does that have for how we should deal with conspiracy theories and their believers? Does that mean that conspiracy theories are vindicated? Does that mean that we don't need to worry about these theories and their believers because a general conspiratorial scepticism raises the level of attentiveness and we can easily get rid of false theories?

As we have seen, philosophers such as Basham and Dentith argue that this indeed shows that the current attention that conspiracy theories receive is in fact a witch hunt. If conspiracy theories can be rationally believed and can sometimes be true, then they shouldn't deserve special attention *qua* being conspiracy theories. False theories should be debunked, but that holds for all types of theories. Even Michael Butter, who—as we have seen—operates with a definition of conspiracy theories on which these theories are always false, doesn't think that conspiracy theories pose a special danger to society; conspiracy theorists don't seem to be particularly violent, and some of their scepticism with regard to the establishment and elites is even healthy.<sup>4</sup>

In what follows, we will argue for a very different view. We will arrive at that conclusion not by contesting the view of Basham and Dentith that conspiracy theories are rationally believable, but by proposing that belief in unsubstantiated and false conspiracy theories can indeed be dangerous for open societies and their institutions. Our result is, in fact, the result of a rational reconstruction of the epistemic situation of people who believe in (false) conspiracy theories.<sup>5</sup> This reconstruction not only explains the consequences of beliefs in false conspiracy theories, but it also outlines why it is difficult to successfully dissuade people from adhering to objectively wrong views in this respect.

Before we get to this, it's worth reflecting, though, on the epistemic situation that citizens of a modern society find themselves in if they *don't* believe that they are targets of a conspiracy.

<sup>4</sup> Butter (2018) sees the danger to democracies in the polarization of political positions and understands conspiracy theories as an expression of that polarization. As we will argue below, the polarization of modern democracies is partly *caused* by conspiracy theories.

<sup>5</sup> Of course, the fact that conspiracy theories are rationally believable does not mean that they are sometimes, let alone typically, rationally believed. To what extent conspiracy theorists are, in fact, nutcases is a matter of empirical research.

## 2. The Division of Cognitive Labour and the Role of Trust Networks

Let's reflect for a minute on the things we know and why we know them. Most of that knowledge stems from testimony. A lot of it stems from the testimony of people that we do not know personally but that we have trusted because we realized that they have the relevant expertise on the matter. We get knowledge from reading the news, watching TV, reading books, attending classes in college or school, or talking to a physician or a lawyer. The fact that we attain knowledge on the basis of what we read, see, and hear there is due to the fact that the people that certain institutions (like the media, universities, colleges, and schools) *present* as experts actually *are* experts.

Now, unless we have intimate insight into these institutions ourselves and know how journalists, scientists, lawyers, physicians, etc. work, how they are trained and selected, and what track record they have of getting things right, we are typically not in a position ourselves to evaluate whether trust in these experts is justified. But then how do we realize their expertise?

Well, typically we do that on testimony as well. We picked it up from people that we already trusted on a *personal* level—like our parents and others in our close vicinity—who told us that we can also trust these institutions and their experts. Our parents, or those others in our close vicinity had themselves, then, either direct personal reasons to trust specific experts (perhaps based on personal acquaintance) or also indirect reasons for such trust, based on the testimony of yet others. That is how we typically form our beliefs. Is it also *rational* to form beliefs like that? Fortunately, that is the case.

The trustworthiness of an informant is a matter of the interplay of at least the following factors (Baurmann 2007b):

- (1) *Competence*: reliable and useful information from informants is dependent on their appropriate cognitive and intellectual abilities as well as on their external resources to identify the truth in the relevant area.
- (2) *Extrinsic incentives*: benefits and costs, rewards and sanctions, recognition and contempt can motivate informants to exhaust their cognitive potential and utilize their resources to discover reliable information and transmit their knowledge to recipients. Extrinsic incentives can also tempt informants to behave opportunistically, to underachieve, to misuse their resources and to manipulate and deceive recipients with wrong, misleading, or useless information.
- (3) *Intrinsic incentives*: emotional bonds of solidarity, sympathy and benevolence, the internalisation of common social values and norms, moral virtues, and personal integrity can motivate informants to transmit valuable knowledge and reliable information to a recipient. Emotional aversion

and hatred, the internalization of deviant values and norms, moral vices, and malignance are potential reasons to deceive and cheat a recipient and to give false and deceptive testimony.

In certain situations (e.g. asking a stranger on the street for the time of the day), assessing the reliability of an informant might be relatively easy and not require deep insight into the factors just mentioned. In other areas, the situation may be far more complex. Gathering evidence about competence, and extrinsic and intrinsic incentives is far too costly for most cases of information transfer. In these situations it is rational to use *heuristic rules* in order to assess the trustworthiness of your information sources. For example, we rely on certifications from approved educational institutions or from employment in professional institutions as indicators of scientific competence and academic expertise.

How do we know that these heuristic rules are reliable? Certainly not on the basis of our own experience alone. For some sciences that are in direct contact with technology, we can, to some extent, assess the trustworthiness of that science's expertise. Airplanes mostly fly, ill persons are often cured, etc. But laypeople are already not able to assess on the basis of their individual experience whether, say, being a certified practitioner of homeopathy promises a better track record in curing diseases than a university degree in standard medicine. Even if the sciences make sometimes *exoteric* claims that can in principle be assessed without expert knowledge (in contrast to *esoteric* claims that cannot be so assessed), individual experience of the track record of a science or discipline with respect to these exoteric claims comes typically nowhere near a sufficient empirical basis for assessing the reliability of that science. If a society has knowledge of that track record via exoteric claims, then this knowledge is *distributed* knowledge. Again, it would be irrational to try to gather that evidence that justifies our reliance on heuristic rules ourselves.

But this seems to put us in a dilemma: on the one hand, we are extremely and unavoidably dependent on the testimony and the knowledge of experts in our society; on the other hand, the same unavoidable and irreducible dependence on testimony reoccurs for knowing which experts we can so trust. How can we break out of this predicament?

In the real world, we do this via relations of *personal trust*. We learn, on the basis of our own experience, that we can trust our parents and they inform us that we can trust our school teacher and our family doctor. They can provide us with this information because they stand in personal trust relations to others that have made the relevant collective experience. From their own, and the personal experience of others in their network, they assess whether these potential epistemic authorities "know what they are talking about" and which heuristic rules are reliable. The wider the network on which this assessment is based, the more

accurate this assessment will be.<sup>6</sup> Over the years, we learn ourselves whom to personally trust and widen our trust-network. We also develop heuristics for assessing the trustworthiness of people with which we only have short-term interactions.

The more individuals I trust personally, the broader the potential reservoir of independent information and knowledge from which I can draw to judge the validity of social rules and criteria for the credibility and trustworthiness of people, institutions, and authorities. This judgement would also involve reference to testimony to a large extent—but it is testimony from sources whose quality I can evaluate myself. Therefore, I can ascribe a high *trust-value* to the testified information.

I will also be inclined to ascribe a high trust-value to information which stems from sources whose trustworthiness is not approved by myself, but by the testimony of people I personally trust. In this way it is possible to profit from a more or less widespread network of personal trust relations which is linked together by people who trust each other personally and thus simultaneously function as mutual trustintermediaries (Coleman 1990, 180). Such trust-networks pool information and knowledge and make them available for the individual at low costs or even for free. Thus they represent important instances of “social capital” (Baurmann 2007a).

The efficiency of personal trust-networks as information pools is enhanced if they transgress the borders of families, groups, communities, classes, or nations. The more widespread and the larger the scope of trust networks, the more diverse and detailed the information they aggregate. The possibility of individuals getting from their trust-networks the quality and quantity of information they need to form a realistic and balanced picture of their world is, therefore, largely dependent on the coverage their trust-networks provide.

Trustnetworks can remain latent and silent about the established social criteria for epistemic credibility and authority for a long period. Their special importance becomes evident when, for example, under a despotic regime a general mistrust towards all official information prevails. But personal trust-networks also provide fallback resources in well-ordered societies with usually highly generalized trust in the socially and formally certified epistemic sources. Under normal circumstances in our societies we consult books, read newspapers, listen to the news, and pay

<sup>6</sup> The success of this mechanism requires (amongst other things) that reliable knowledge is indeed generated in the society in question and that there is a recognizable and substantial track record of that knowledge. For example, in a small tribe the local shaman is an epistemic authority even though he or she is a charlatan—simply because there is no competition with other, more trustworthy and reliable epistemic authorities. Likewise, it may be hard for outsiders to identify who is right in phases of scientific revolution. The new and better paradigm might not yet have track record that would allow non-experts to recognize its superiority over the old.

attention to our experts and authorities if we want to learn something about the world. And even when we develop mistrust in some of our authorities or institutions, we normally do so because we hear suspicious facts from other authorities or institutions.

However, from the subject's point of view, the ultimate touchstone of my belief in testimony can only be my own judgement. Even if I'm ready to defer my judgement to an epistemic authority, *I* must recognize that authority. And it makes a great difference for the reliability of that judgement whether I can base this judgement only on my own very limited personal information or if I can fall back on the information pool of a widely spread network which is independent of socially predetermined criteria for epistemic credibility and authority.

So, on the one hand, our society with its division of cognitive labour and its institutions that train and systematically educate highly specialized and knowledgeable experts, and that provide incentive structures and selection processes which lead to reliable and trustworthy performance of these experts, generates a lot of knowledge. However, on the other hand, this does not by itself guarantee that everyone can automatically benefit from the generated knowledge. One needs to happen to stand in a number of stable enough personal trust relations of the right kind in order to be able to get oneself to trust in the output of these knowledge-generating institutions.

Ultimately, for all that most people directly know about academia, the media, and schools, and for all knowledge of facts they observe themselves and that they can use in order to verify claims made by members of these institutions, this "generated knowledge" could just be a major scam. Which brings us back to our conspiracy theorists.

### 3. Epistemological Effects of Belief in Conspiracy Theories

In many contemporary prevalent conspiracy theories, the relevant conspirators are many, if not all of the institutions that, in open societies, are supposed to exercise mutual supervision and control. Big pharma lobbies politicians and pays scientists and the media to convince everyone else that vaccinations are beneficial and pretty harmless to the recipient, in order to make a profit.

For most people, the reason to believe such a conspiracy theory originates from the testimony of some opinion leaders and alleged experts whom they trust as epistemic authorities in this matter—maybe because they are able to fake a special competence and personal integrity in social media or group meetings. The basis for this trust may be irrational; to believe the information from a trusted source is not.

But belief in a false conspiracy theory of this kind has repercussions for your epistemic situation. Let us assume that you believe in a factually wrong

conspiracy theory that vaccination is harmful for the recipient but that this is covered up in the way and for the reasons described above.

The heuristic rules that the relevant institutions provide for the identification of expertise (e.g. having a scientific degree, being employed at such an institution) will then become useless to you—unless, of course, you'd see that the institutions react appropriately to the alleged fraud by firing corrupt scientists or journalists, which, of course, they don't, since your theory is false. It will also impact the way you view the rest of your trust-network. Those members of your family, or your immediate circle of friends, who initially provided a pathway to benefit from the knowledge produced by the institutions of your society are now unreliable. You don't need to think that they tried to mislead you; it is sufficient to think that they, too, have been misled. And indeed, if pressed on details of your new vaccination conspiracy theory, they don't have direct evidence that they can provide against it, right? So, they naively believed on hearsay, and you can now “enlighten” them.

Therefore, the initial and seemingly quite harmless entry into the world of conspiracy theories can trigger a dynamic mechanism that leads to a process of ongoing epistemic reinforcement of a deficient world view and, as a final result, to a cut-off from the knowledge generated in a society. The core of this social mechanism is constituted by a process of mutual influence and adaptation in which individual experiences and deliberations are continuously compared and adjusted in accordance with the experience and deliberations of other persons who are considered relevant and reliable (Baurmann et al. 2014, 2018; Betz et al. 2013).

It is crucial for an understanding of this mechanism that opinion formation involves *first-order opinions* about the issues that are relevant in a certain field—big pharma lobbies politicians and pays scientists and the media—and *second-order opinions* about the epistemic trustworthiness of persons who express their opinions about these issues—for example, opinion leaders in a peer group. Second-order opinions refer to characteristics of persons that are relevant for their quality as epistemic sources. It is essential to note that persons influence each other mutually both in the formation of their first-order opinions and their second-order opinions. They consider the opinions of other trustworthy persons with regard to the explanation of, for example, political processes and developments, as well as with regard to their estimation of who is competent and reliable to pass considered judgements over these issues.

It is an important feature of this social mechanism that it not only works in the development of first- and second-order opinions but that it also entails dynamic relations between these different layers of opinion formation. On account of this structure, persons will be influenced by other persons not only in regard to their opinions about political options, societal connections, or ideological world views. This adaptation process itself will, in turn, be intertwined with the mutual adaptation of the second-order opinions about who has sufficient or special competence to understand and judge such options, connections, or world views. These two-layer

dynamics could result in far-reaching transitions of the initial convictions of persons so that they ultimately may adopt extremist opinions which were originally not within their opinion space and may well have appeared absurd to them.

How may this mechanism work in our example? As one result of your “new” belief in the pharma conspiracy your epistemic trust in people who don’t believe in the truth of your conspiracy theory will be weakened. Simultaneously, you will develop new trust in the epistemic competence of persons who share your opinion of the deceitful schemes of the pharma industry. In consequence, in the future the influence of your new epistemic friends on your opinion formation will grow and that of your old friends will decline.

This will affect your first-order opinions about the world: the new authorities in your social-epistemic peer group may strengthen your convictions about the conspiracy of the pharma industry and may transfer it to other areas of society, maybe in regard to a conspiracy between politics and the media.

But your second-order opinions will also be infected by the new influences: they may further erode your epistemic trust in your old circle of friends and present, in addition, new authorities and special experts who can “enlighten” your world view even more.

Next steps of this vicious spiral may follow: the growing circle of your new epistemic trustees will also produce a further growing influence on your first- and second-order opinions. Your conspiracy theories may get more and more radical and wide-ranging, undermining your confidence in all relevant institutions of your society. And you may terminate all your former epistemic trust-relations, beginning with your social environment and ending with a break with all the “official” epistemic authorities and sources of your society—leaving you with a close and exclusive network of a special group of believers who are confirming themselves mutually in their opinion of the factual world and other persons.

Therefore, false conspiracy theories are dangerous levers to start a dynamic downward spiral in the (epistemic) trust-relations of persons, because they evoke an initial mistrust towards societal institutions *and* towards persons who deny reasons for this mistrust. In consequence, believers of conspiracy theories will often generalize their institutional mistrust and simultaneously restrict their epistemic trust to persons who are enforcing and stabilizing this mistrust.

As a final result, you are indeed cut off from the knowledge generated in your society. Presumably you have a residual core of personal trust relations left; at least those relations with your fellow “truthers”, the people who put you initially in the know about the purported large-scale conspiracy that is going on in your society. Your interest will be that none of the institutions that have failed you will get between you and those you personally trust. It will be rational for you to prefer an information-flow architecture that gives you unfiltered and immediate access to information, coming from persons to which you (believe you) stand in a direct trust relation.

This is, indeed, rational for someone who believes a false conspiracy theory, because for her it seems that the institutions that are meant to filter, mediate, or cross-check information, are all corrupt or broken. As we noted above, even though personal trust is necessary to participate in the knowledge generated in your society, your trust in its institutions is not exclusively based on testimony. For one thing, you may have direct evidence that the experts in your society can't be completely incompetent. Technology typically works and makes progress; occasionally things turn out the way that politicians promised such that you experience the consequences of that improvement yourself. But normally you also observe that when things go wrong, there are correcting mechanisms: journalists report, say, that scientists falsified their data, and politics and academia react properly. Studies are retracted; perhaps laws are implemented in order to ensure higher standards; policies that were based on the misinformation are changed; the scientists get punished or fired. Thus, in order to have trust in the institutions of your society, you don't need to believe that everything is always going well. But you need to believe that *when* things go wrong, there is a good chance that the mutual control mechanisms of these institutions will detect and correct the mistakes, and you have occasionally evidence that this indeed happens.

Now, as we already noted, in a case when you believe a *false* conspiracy theory, you'll think you have evidence that *none* of this happens. The vaccination programme doesn't stop; scientists just deny the allegations; politicians even discuss the introduction of a formal duty to vaccinate in order to force vaccination sceptics like you to comply. You can directly observe that the system is broken. Why should you want corrupt institutions to become even stronger?

If you get someone who you personally trust into power—perhaps even into a presidency—you will, therefore, not be interested in having that person's actions controlled by corrupt institutions. The influence of these institutions would need to be reduced, their political power limited, the “swamp” must be “drained”. It will be rational to prefer the destruction of (what actually are) institutions of an open society. That is precisely what we can empirically observe when open societies take an autocratic turn based on unleashed and self-reinforcing conspiracy theorizing.

As we discussed above, a network of personal trust relations is your entrance ticket to the knowledge society. It is also your fallback option if the institutions of that society let you down. In this case, you will want to side-step these institutions and establish a tight network of people to whom you think you have reliable personal relations. This seems to be the empirical phenomenon we observe: generalized social trust—as we find it in open societies—is replaced by *particularist trust*.

Individuals adhere to a particularistic trust if they only trust members of a clearly demarcated group and generally mistrust members of all other groups. Particularistic trust is supported by heuristic rules which are the exact mirror image of those heuristic rules which embody a generalized trust: while rules of

generalized trust state that one should *trust* everybody unless exceptional circumstances obtain, rules which constitute a particularistic trust state that one should *mistrust* everybody with the exception of some specified cases.

This is an epistemically limited and dangerous position. People who are thrown back on particularistic trust can easily be manipulated and controlled. The result is the exact opposite of the truth-generating epistemic dynamic in an open society.

These are all almost purely epistemological considerations that make dismantling the structures of open societies rational, if you believe a sufficiently wide false conspiracy theory about these structures. Thus, instead of strengthening an open democracy and its institutions, these beliefs lead to their erosion and destruction.

## 4. But Is Conspiracy-Thinking Always Bad?

One objection to our discussion could be that it is too naive and one-dimensional. Granted, there are these negative effects that are to be predicted on a rational reconstruction of the epistemic situation of conspiracy believers, but (a) perhaps conspiracy believers aren't *fully* rational after all and don't draw the proper consequences that their belief should have for their generalized trust, and (b) perhaps there are still other benefits that conspiracy-thinking may have for open democracies. We'll briefly address both of these objections.

### 4.1 Are Conspiracy Believers Consistent?

(a) is indeed somewhat plausible. Conspiracy theorists are often internally inconsistent in their world view and may thus not see that their belief that all institutions massively fail when it comes to *X* (say, vaccination), should also imply the untrustworthiness of those same institutions when it comes to *Y* (say, whether you can believe any other medical advice).

As Lewandowsky et al. (2018) show, climate science deniers often hold inconsistent views. In what Lewandowsky et al. call "contrarian discourse", one can find over one hundred incoherent pairs of arguments (Lewandowsky et al. 2018, 184) claiming that "future climate cannot be predicted", as well as that "we are heading into an ice age", or that the observed CO<sub>2</sub> rise is actually caused by warming, as well as that there is no correlation between CO<sub>2</sub> and temperature. Most of the incoherent arguments identified are not actually endorsed by one and the same individual, but Lewandowsky et al. can also show that some individuals endorsed incoherent pairs of arguments at different times and different places.

Doesn't that suggest that conspiracy theorists will most likely fail to draw the epistemological conclusions of their views (just as they often fail to draw also other conclusions from their views)? They may simply choose to believe a convenient

(though incoherent) set of views on which they can still trust the testimony of experts and other relevant institutions as long as that information is unconnected to the (often) politically charged topics for which those same experts and institutions are thought to be corrupt and complicit in a conspiracy.

This is clearly an empirical question, and we can't answer it in the context of this chapter. However, there is some empirical evidence that seems to speak against this objection.

First of all, even though Lewandowsky et al. show that climate change deniers endorse incoherent arguments at different places or times, this does not show that these climate change deniers have—individually—an inconsistent or even incoherent set of beliefs. As Lewandowsky et al. also say in their paper, it is often precisely their conspiracy belief which provides coherence of their beliefs at a higher level:

[A] known attribute of conspiracist thought is that it can appear incoherent by conventional evidentiary criteria. To illustrate, when people reject an official account of an event, they may simultaneously believe in mutually contradictory theories—e.g., that Princess Diana was murdered but also faked her own death. The incoherence does not matter to the person rejecting the official account because it is resolved at a higher level of abstraction; there is an unshakable belief that the official account of an event is wrong. (Lewandowsky et al. 2018, 179)

Thus, the fact that some conspiracy theorists hold incoherent beliefs at some level does not, by itself, establish that conspiracy theorists are, in general, incoherent, and that a rational reconstruction of their epistemological situation is inapplicable.

There are two further empirical findings which suggest that our analysis is on the right track. According to our analysis, we should expect that conspiracy theorists will not only distrust the government or other epistemic authorities when it comes to one specific issue, but will show general distrust for such information-providing institutions. Thus, someone who believes one conspiracy theory should then be more likely to believe other conspiracy theories also on unrelated issues. It is a relatively stable finding in social psychology that this is indeed the case:

One of the main research findings on this phenomenon [i.e. in belief in conspiracy theories] is that conspiracy beliefs are monological in nature: one conspiracy theory reinforces other conspiratorial ideas, making individuals who believe in one conspiracy theory more likely to also believe in other conspiracy theories.

(van Prooijan and van Lange 2014, 237)

Social psychologists find that result very surprising and speculate for its explanation over a “conspiracist mindset”, a particular psychological disposition to

believe conspiracy theories. However, on a rational reconstruction of the epistemic situation of conspiracy adherents, this outcome can be expected without assuming any kind of irrational disposition (cf. also Hagen 2018 for a similar result).

If you believe that a certain institution is not trustworthy because it has been corrupted, then you have reason to assign a relatively low credence to any piece of information it provides and higher credence to alternative information, stemming from presumably more trustworthy sources. This can explain why conspiracy theorists believe also other, non-related conspiracy theories and might assign a relatively high plausibility to several mutually inconsistent claims in contrast to the “official” account (cf. Bruder et al. 2013).

The relevant mechanism behind this “surprising” result is simply that you will, in general, assign a low trust value to all of the official information-providing institutions. This very mechanism can also be observed directly. In a recent study, Katherine Levine Einstein and David M. Glick (2015) exposed test subjects to a conspiracy theory by having them read an article that reported claims by Jack Welch, former CEO of General Electric, suggesting that the US Bureau of Labor Statistics had manipulated recently reported unemployment data for political reasons. Mere exposure to these claims affected the amount of trust that the test subjects afterwards reported for a range of governmental institutions, such as the US Census Bureau, the Food and Drug Administration, the Presidency, the local police, and local schools. Needless to say, Welch’s allegations in the report didn’t implicate these other institutions. Thus, it seems, institutional distrust spreads rather quickly.

There certainly need to be more studies of this kind before one can say anything definitive, but these findings support the social epistemic dynamics that we describe above and suggest that the undermining effects of conspiracy theories are, indeed, to be expected empirically.

## 4.2 Are False Conspiracy Theories Always Bad?

At the beginning of this chapter we said that we will oppose the view that propagating conspiracy theories is good for open societies, because they induce scepticism of the government, which will ultimately lead to a strengthening of the institutions of open societies that exercise mutual control. Then we argued that, *yes*, conspiracy theories lead to scepticism of government institutions, but *no*, rampant conspiratorial speculations do not strengthen institutions of open societies, and, over time, will lead to scepticism about all of them and completely undermine the kind of trust that is necessary in order to keep them functioning.

This argument leaves open whether belief in false conspiracy theories could still have *other* positive (epistemic) effects. This question is too broad to be discussed thoroughly in this chapter. However, we can give a few pointers here.

First of all, our analysis is a *rational* reconstruction. Hence, the epistemic situation that conspiracy theorists end up in, the low trust that they assign to the institutions of open societies, the comparatively high trust they assign to a small set of personal acquaintances, etc. is an appropriate response in circumstances in which their beliefs are true. Now, even belief in *false* conspiracy theories can lead to an adequate limited trust-network, if the degree of trustworthiness assigned is the same that a properly informed agent would assign in that same situation.

Let us assume that you live in a society in which the institutions that should provide information and exercise mutual control are, in fact, broken and corrupt. Let us also assume that you hold a complex conspiracy theory about these institutions: you believe that all these institutions are controlled by the New World Order (NWO). Consequently, you have little trust in these institutions and epistemically navigate on the basis of a network of personal trust relations. Let us further assume that your conspiracy theory has it all wrong; it's not the NWO that controls everything, but another organization with intentions that are very different from those that you suppose the NWO to have. In this case, you believe a false conspiracy theory, but your epistemic reaction to that is still adequate and, moreover, *objectively* adequate. From an internalist point of view you are rational with respect to your background beliefs, and you are also objectively justified when reducing your trust-network to the actual reliable core. Hence, under certain conditions, believing false conspiracy theories can indeed be epistemically beneficial.

Perhaps some of the disagreement between our overall estimate of the value of conspiratorial reasoning and that of our colleagues that we cited at the beginning of the chapter, has to do with differences in judgement about the kind of open society we actually live in. We assumed for our argument here, an open society in which the relevant institutions are largely functional and the division of cognitive labour is overall reliable. Under these circumstances, belief in unsubstantiated and false conspiracy theories and the suspicion they promote has only bad consequences. Any kind of serious conspiracy theory has, in a well-ordered open society, to accept a reversal of the burden of proof and has to present salient indicators as evidence for a malfunction of *prima facie* efficient and trustworthy institutions. Otherwise, a plain conspiratorial scepticism only undermines a virtuous equilibrium of institutional stability and institutional trust. Of course, one may have a less optimistic picture of our current society, but then our disagreement is ultimately not about the positive or negative role that conspiracy theories can play in open societies, but rather about the type of society we are in.

A relevant disagreement may concern potential other epistemic benefits that belief in false conspiracy theories may have. One might argue that false conspiracy theories, just like any other false beliefs, are to some extent epistemically beneficial, because they allow us to challenge our true beliefs and thus to arrive at a better and deeper understanding of these truths.

This argument needs more elaboration than we can provide here, hence two brief comments must suffice: (i) it is not clear that a better and deeper understanding is always preferable. We said above that belief in false conspiracy theories will cut you off from knowledge via testimony. It is in everyone's interest to know certain things merely on the basis of testimony, since we lack the time and resources to know them in any better way, and we just don't care enough to understand them fully. For example: given that building 7 of the World Trade Center collapsed as a result of the events nearby, we couldn't care less why exactly it collapsed and how the events nearby precisely caused this, and that's so for most people.

(ii) This fact doesn't change much if, instead of focusing on the first-order belief that the conspiracy theory targets, we move to the higher-order belief about the trustworthiness of our institutions. Although it might sound more plausible that it would be good if most people more deeply understood how peer review, consensus formation, and other quality control mechanisms in the sciences will lead to reliable expert opinions, most people don't care about a deep understanding of these things either, and it would be a waste of their cognitive resources to develop any expertise in these matters.

In a complex knowledge society with a cognitive division of labour, it is not necessary for everyone to know these things to profit from the knowledge generated. In order for such a society to produce reliable knowledge, some people need to exercise specific control and be sceptical and alert with respect to the institutions for which they are responsible, but this responsibility is distributed. Nobody needs to know all the details of it for the system to work, and unless someone is specifically interested in the details of the system people are not generally in an epistemically better position by knowing many details about it.

## 5. How to Confront Conspiracy Thinking

If the social-epistemic mechanisms that we have described in this chapter adequately represent the empirical belief and trust formation of conspiracy theorists, then we should be able to also say a thing or two about the ways in which unsubstantiated and false conspiracy theories can be successfully debunked or confronted.

Debunking deficient conspiracy theories is often seen as impossible, or at least very difficult, due to the alleged unfalsifiability of dogmatic conspiracy theories. It is often thought that deeply believed conspiracy theories are immune to falsification because any counter-evidence is automatically explained away by the conspiracy theory as evidence planted by the conspirators. But although there may be conspiracy theories which indeed assume an all-powerful group of conspirators (maybe if the conspirators are an alien race that have the power of Descartes' evil demon and can make all kinds of circumstances appear as counter-evidence for

the conspiracy theory to us), most conspiracy theories are not that all-encompassing and—at least logically—allow for falsification.

A different matter is, of course, if and when conspiracy theorists actually *change* their theories in response to counter-evidence. Very often, counter-evidence is explained away by a conspiracy theory as being planted or disseminated by members of the conspiracy. This by itself is not yet a problematic move by the conspiracy theorists. After all, her theory states that there is an ongoing conspiracy of people who don't want to be exposed. Thus, it makes a lot of sense to assume that these conspirators will do what they can in order to hide their tracks and mislead the public. However, this move becomes problematic when this strategy of explaining away *prima facie* counter-evidence leads to an *ad hoc* extension of the assumed group of conspirators.

As an example, assume that a conspiracy theorist believes that big pharma lobbied politicians and physicians into a nationwide vaccination programme which is in fact harmful for the citizens, but makes big pharma a lot of money. In this case, official denials from the big pharma corporations that such conspiracy theories are false, will plausibly not be of much evidential weight for the conspiracy theorist as proof that she is wrong. On her theory, such denials are to be expected from the corporations that are implicated in the conspiracy.

Now, let's assume further that a group of seemingly independent journalists start investigating the matter, but come back empty-handed. As far as they could find out, there is no ongoing conspiracy and they publish articles that reject the conspiracy theories to the contrary as an unfounded witch-hunt. What, indeed, often happens is that conspiracy theorists will, in reaction to such reports, extend the group of conspirators (which thus far only included physicians and some politicians) to also include at least the journalists who claim to have investigated the matter. Such a move is typically *ad hoc*, in the sense that there is no independent reason to believe that the journalists are part of the conspiracy (independent from the fact that these journalists have produced this apparent counter-evidence). Such a move—*ad hoc* extending the group of conspirators—is not rationally warranted or licenced by the conspiracy theory as such.

As we have seen above, such moves, however, may be rational reactions in cases in which the conspiracy theory has already destroyed the foundations for generalized trust. If trust in the functioning of institutions is generally low, then not putting much trust into the institution *journalism* is not irrational or unmotivated.

But this makes *debunking* conspiracy theories especially difficult. Official pronouncements that a conspiracy theory is mistaken or crazy will not carry much weight for someone who already assigns a low trust value to the institution making that announcement. What can be done?

Sunstein and Vermeule (2009) offer a strategy that is aimed at breaking up the trust networks of the conspiracy theorists. The idea is to infiltrate—either openly or anonymously—their networks, for example via government agents

participating in the relevant chat rooms or newsgroups. Open infiltration may prove to be less promising in light of that fact that these government agents will be perceived as members of the conspiracy. Anonymous infiltration may seem more promising:

The risk with tactics of anonymous participation is that those tactics may be discovered or disclosed, with possibly perverse results. If the tactic becomes known, the conspiracy theory may become further entrenched, and any genuine member of the relevant groups who raises doubts may be suspected of government connections. And as we have emphasized throughout, in an open society it is difficult to conceal government conspiracies, even the sort of conspiratorial tactic we have suggested, whose aim is to undermine false and harmful conspiracy theorizing.

If disclosure of the tactic does occur, however, the perverse results are just a possible cost, whose risk and magnitude is unclear. Another possibility is that disclosure of the government's tactics will sow uncertainty and distrust within conspiratorial groups and among their members; new recruits will be suspect and participants in the group's virtual networks will doubt each other's bona fides. To the extent that these effects raise the costs of organization and communication for, and within, conspiratorial groups, the effects are desirable, not perverse.

(Sunstein and Vermeule 2009, 225–6)

Indeed, detected anonymous infiltration will lead to further distrust within the remaining trust-network of the conspiracy theorists. However, it is not clear how that will be a remedy for the problem. Knowing that the government, which I didn't trust in the first place, anonymously infiltrated my peer network maybe destroys my peer network or the trust I put into it, but it certainly doesn't reinstate my trust in the government (on the contrary!). Thus, the anonymous infiltration tactic primarily promises to further destroy and diminish the remaining trust-networks of people believing a conspiracy theory, which we identified as the primary problem to begin with.

What would our account suggest as a more promising strategy? As we argued, the fall-back option and default basis for wider trust-networks are personal trust-relations. These need to be strengthened and developed in order to reintegrate conspiracy theorists back into the "knowledge society". In terms of general strategies, that means that debunking conspiracy theories at the level of a big, abstract, and anonymous institution will have less impact than questioning these theories that takes place at the level of personal relations. As an example: a debunking campaign against anti-vax conspiracy theories will have probably less impact if it primarily consists of pronouncements from, say, the World Health Organization. The debunking campaign will be a lot more successful if the trusted family doctor is the one who carefully explains the value of vaccination campaigns.

Secondly, personal relations with people who believe conspiracy theories should be kept alive—not inviting Uncle Bernd over for Christmas anymore because he has developed funky views on the causes for the refugee crisis might have short-term benefits for the general atmosphere at the Christmas dinner, but it will have devastating long-term effects for the epistemic trust-network of Uncle Bernd.

In general, it would be important not to further alienate or ostracize conspiracy theory believers from those social networks that still provide a link to the knowledge generated by the reliable epistemic institutions of an open society. Already labelling someone as a “conspiracy theorist” is, of course, a first move towards ostracizing that person. Perhaps it would be better not to use this label in discussions with conspiracy theorists, and instead address conspiracy theories as what they ultimately often are: false theories that are based on misleading evidence.

We started our chapter with the observation that conspiracies sometimes happen and that, therefore, belief in a conspiracy theory can't be irrational just because you believe that certain events are orchestrated by a conspiracy. Indeed, uncovering actual conspiracies in our society is important. Conspiracy theorizing might occasionally be onto something, and in this case we need to know. So, shouldn't one conclude that conspiracy theorizing is an important force for the good in our society? Shouldn't we tolerate the growth of false conspiracy theories as a harmless (and sometimes even somewhat entertaining) side-effect of an important control mechanism?

We have argued that this would be naive. False conspiracy theories are dangerous for the institutions of open societies. They undermine and eventually destroy the trust network that is necessary for these institutions to perform their primary functions. As a consequence, their very existence may be put in question. It is thus necessary that we understand why (some) people are prone to believe false conspiracy theories, even though the evidential situations for these theories seems objectively bad. This will require epistemological, sociological, and psychological research on conspiracy theories and their believers. This is not a witch-hunt.

## References

Basham, Lee and Matthew R. X. Dentith, 2016: “Social Science's Conspiracy-Theory Panic: Now They Want to Cure Everyone”, *Social Epistemology Review and Reply Collective* 5: 12–19.

Baurmann, Michael, 2007a: “Political Norms, Markets and Social Capital”, in J. Kuhnelt (ed.), *Political Legitimization without Morality*, Vienna and New York: Springer.

Baurmann, Michael, 2007b: "Rational Fundamentalism? An Explanatory Model of Fundamentalist Beliefs", *Episteme* 4: 150–166.

Baurmann, Michael, Gregor Betz and Rainer Cramm, 2018: "Führer befiehl, wir folgen dir! Charismatic Leaders in Extremist Groups", in Thomas Christiano, Ingrid Creppell and Jack Knight (eds.), *Morality, Governance, and Social Institutions. Reflections on Russell Hardin*, Basingstoke: Palgrave Macmillan, Basingstoke, 259–287.

Baurmann, Michael, Gregor Betz and Rainer Cramm, 2014: „Meinungsdynamiken in fundamentalistischen Gruppen. Erklärungshypothesen auf der Basis von Simulationsmodellen“, *Analyse & Kritik* 36: 61–102.

Betz, Georg, Michael Baurmann and Rainer Cramm, 2013: "Is Epistemic Trust of Veritistic Value?", *Etica & Politica* 15: 25–41.

Bruder, M. et al., 2013: "Measuring Individual Differences in Generic Beliefs in Conspiracy Theories Across Cultures: Conspiracy Mentality Questionnaire", *Frontiers in Psychology* 4: 225.

Butter, Michael, 2018: "Nichts ist wie es scheint": *Über Verschwörungstheorien*, Berlin: Suhrkamp.

Clarke, S. 2002: "Conspiracy Theories and Conspiracy Theorizing", *Philosophy of the Social Sciences*, 32: 131–50.

Cohnitz, Daniel, 2018: "On the Rationality of Conspiracy Theories", *Croatian Journal of Philosophy* 28: 351–65.

Coleman, J. S., 1990: *Foundations of Social Theory*, Cambridge, MA: The Belknap Press of Harvard University Press.

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

Dentith, M. R. X., 2017: "Conspiracy Theories on the Basis of the Evidence", *Synthese* 196: 2243–61.

Einstein, Katherine Levine and David M. Glick, 2015: "Do I Think BLS Data are BS? The Consequences of Conspiracy Theories", *Political Behavior* 37: 679–701.

Hagen, K., 2018: "Conspiracy Theorists and Monological Belief Systems", *Argumenta* 3: 303–26.

Hayek, F. A., 1967: *Studies in Philosophy, Politics and Economics*, London: Routledge.

Hofstadter, Richard, 1965: *The Paranoid Style in American Politics*, New York: Random House.

Lewandowsky, Stephan, John Cook, and Elisabeth Lloyd, 2018: "The 'Alice in Wonderland' Mechanics of the Rejection of (Climate) Science: Simulating Coherence by Conspiracism", *Synthese* 195: 175–96.

McGrew, Timothy, 2019: "Miracles", *The Stanford Encyclopedia of Philosophy* (Spring 2019 Edition), Edward N. Zalta (ed.), <https://plato.stanford.edu/archives/spr2019/entries/miracles/>.

Popper, K., 1966: *The Open Society and Its Enemies*, Vol. 2, London: Routledge.

Sunstein, Cass and Adrian Vermeule, 2009: "Conspiracy Theories: Causes and Cures", *Journal of Political Philosophy* 17: 202–27.

Ullmann-Margalit, Edna, 1978: "Invisible-Hand Explanations", *Synthese* 39, 263–91.

van Prooijen, J. and P. A. M. van Lange, 2014: "The Social Dimension of Belief in Conspiracy Theories", in J. van Prooijen and P. A. M. van Lange (eds), *Power, Politics, and Paranoia: Why People are Suspicious of their Leaders*, Cambridge: Cambridge University Press, 237–53.

Zwierlein, C. and B. de Graaf, 2013: "Security and Conspiracy in Modern History", *Historical Social Research* 38: 7–45.



# Index

*Note:* Tables, figures, and boxes are indicated by an italic ‘*t*’, ‘*f*’, and ‘*b*’, respectively, following the page number.

Aikin, Scott 57, 310

algorithms 2, 11, 277, 322

- content over quality of content 7
- echo chambers 2
- Facebook 170, 172
- fake news: mechanisms fostering the
  - production and spread of 7
  - filter bubbles 2, 7, 135, 323
  - Google 11–12, 72, 229–30, 239, 240
  - platforms 172
  - propagating discredited science 11–12
  - ‘retraining’ algorithms 330
- see also* Google and retracted articles; Internet
- ambient news environment 12, 277–80
- background gardening 12, 277–80
- epistemic rot 12, 281–2

anti-reductionism 3

anti-vaxxers 112, 124, 128–31

Arendt, Hannah 163, 164, 166

bad news 56

- bad news/fake news distinction 56, 60
- bad news/ineffective fake news distinction 60
- bad news sources 55
- good news/bad news distinction 59
- poor epistemic pedigree 9, 46, 58, 65

BBC (British Broadcasting Corporation) 228–9, 230, 247, 319

Bezos, Jeff 72

bias 182–3

- bias-inducing information 296, 298, 299
- cognitive bias 30, 57–8, 324, 328, 329
- confirmation bias 186
- experts: liberal bias 182–4, 200
- Fox News as right-biased 207–8, 217
- individual/personal bias 135, 186
- reflection on, as antidote to fake news 8
- social bias 135

Breitbart News 21, 22, 23*t*, 25, 35, 171, 268, 269

bullshit 315–16

- bullshit dimension of fake news 26, 29, 30, 31, 32, 32, 33, 38, 265, 288, 291, 320
- bullshit fake news 22, 31, 35
- bullshit/lying distinction 49, 50n.10, 74
- bullshitter, definition 22, 29
- definition 22, 29, 249
- fake news/bullshit distinction 288
- lack of truthfulness 22, 29, 49–50
- Macedonian teenagers 22
- misleading information and 50
- Mukerji’s definition of fake news 27, 33–5, 37*t*, 288, 320
- as post-truth phenomenon 134

Bush, George W. 76, 316

Butter, Michael 337, 338, 339, 340

Cambridge Analytica (UK) 1, 169

Cassam, Quassim 124–5

censorship 8

- by corporations 65
- by democratic institutions 65
- by governments 72
- by non-governmental bodies 65, 72
- fake news 65
- ‘fake news’ as instrument of corporate censorship 71–2
- ‘fake news’ as instrument of state censorship 70–1

Google 72

public opinion manipulation: Propaganda and Censorship model 11, 162–4, 173, 174

China 162, 173, 174

clickbait 248, 249, 324

- fake news and 31, 48–9, 320
- Macedonian teenagers as clickbait farmers 6, 22, 31, 320

climate change 216, 224, 237, 257

- climate change deniers 165, 216, 348–9

Clinton, Hillary 21, 73, 317

CNN (Cable News Network) 206, 208, 212, 219, 224, 317, 322

Coady, David 157, 310, 317, 318

Colbert, Stephen 316

*Collins Dictionary* 68

conspiracy theories 2, 7, 10, 102  
 attitude of the believer 10, 82, 87, 95  
 bad epistemic standards and 10  
 as background assumption 123, 125, 128  
 Bayesian framework 10, 90–1  
 benefits of 335–6, 340, 350–2, 355  
 C-features 124–5, 128  
 change in meaning as change in topic 95–7  
 consistency/inconsistency of conspiracy  
 believers 348–50  
 conspiracy belief 10, 86–7, 95, 96, 101, 102  
 conspiracy claim 92–3, 98, 100  
 conspiratorial scepticism 335, 336, 337, 340,  
 350, 351  
 counter-evidence, resistance to 10, 83, 86–8,  
 96–7, 101, 102, 125, 352–3  
 cover-up claim 92–3, 94, 98, 100, 101  
 dangers of 336–7, 345–52, 355  
 division of cognitive labour and role of trust  
 networks 341–4  
 epistemically negative connotation 83,  
 84–5, 96  
 epistemic authorities and 344, 346, 349–50  
 epistemic filters 123, 124, 125, 126  
 epistemological effects of belief in conspiracy  
 theories 344–8  
 evidential insulation 83, 88n.13, 91, 94,  
 95, 97  
 examples of 82, 123–4, 336  
 fake news/conspiracy theories distinction 25  
 fake news/conspiracy theories relation 25, 126  
 institutional distrust 349–50, 351, 353, 355  
 as irrational/epistemically irrational 10, 83,  
 85, 88–95, 97, 101–2, 335, 336, 337–40, 355  
 as legitimate way of making sense of the social  
 world 337  
 particularistic trust 347–8  
 PDR (Principle of Democratic Reason)  
 and 138, 146–9, 152  
 a pejorative expression 82, 84, 146–7, 337  
 philosophical methodology and 83–6, 95–7  
 post-enquiry 124, 125  
 predictions, reflexivity, and ad hoc-ness in  
 conspiratorial explanations 92–3, 97–100  
 science denialism and 123, 124–5, 126,  
 128–9, 131  
 as self-isolating belief 10, 83, 86, 87–8, 95,  
 97, 102  
 testimonial insulation 101–2  
 as threat to democratic institutions 13,  
 347, 348  
 trust-networks and 13, 345–6, 347, 351,  
 353–4, 355  
 witch-hunt 335, 340, 353, 355

conspiracy theories: therapies (remedial  
 strategies) 13, 352–5  
 cultivating relationships of trust with  
 conspiracy thinkers 13, 354–5  
 debunking conspiracy theories 13, 352–4  
 infiltrating and undermining conspiracy  
 theories from the inside 13, 353–4  
 vigilance 335

conspiracy theory: definitions 25, 82, 84–5, 86–8,  
 90, 102, 123, 146–7, 334, 338–40  
 conceptual re-engineering of ‘conspiracy  
 theory’ 84–6, 95–6  
 as explanation of an event that cites a  
 conspiracy 84–5, 86, 102  
 as explanation of an event that cites conspiring  
 agents as salient cause 334, 339  
 as holding a conspiratorial belief 86, 87, 102  
 a term without a fixed meaning 69

content reliability 1

corporations  
 censorship by 65, 71–2  
 corporate media 76–7

coverage reliability 13, 300, 302, 305  
 definition 1, 290, 315  
 fake news environment and 287, 290–1  
 lack of 13, 311

critical thinking 10, 190, 303  
 fake news: therapies 8, 310, 325–6  
*see also* PUCT

Cronkite, Walter 77

deception  
 fake news, deception by commission and  
 omission 291  
 fake news, deception dimension of 4, 6, 20, 22,  
 25–6, 27, 28, 29, 31, 32, 33, 34, 37, 41, 126,  
 248, 288, 291, 320  
 fake news distributors 37

deepfakes 301

defeater 209, 266  
 doxastic defeater 210  
 misleading defeater 296–8, 300  
 normative defeater 210  
 rebutting defeater 143, 210  
 science denialism 111, 115, 116, 119,  
 122, 125  
 undercutting defeater 144, 145, 149, 210, 302

democracy  
 conspiracy theories as threat to democratic  
 institutions 13, 347, 348  
 democratic institutions, monitoring and  
 flagging fake news 9  
 democratization of news production 2  
 disinformation as threat to democracy 174

echo chambers and degradation of democracy 206, 214

fake news as label used to discredit democratic disagreement 9

fake news as threat to democracy 68, 156, 286, 292

Internet 170–1

misinformation as threat to democracy 156

news and 1, 39

news as mechanism for democratic accountability 1, 292

persuasiveness rather than truthfulness as guiding principle in democratic systems 161

propaganda as threat to democracy 156

public opinion manipulation: *Pleasing and Seducing the Audience* model 11, 159–62, 173, 174

virality dimension of fake news and 26

*see also* elections; referenda

democratic reasoning *see* PDR

Dentith, Matthew 289, 317, 335, 340

definition of fake news 27, 32–3, 37*t*, 38, 288

*Denver Guardian* 73

Dewey, John 214

digital age 286

- acceleration of the news cycle 321–2
- distinctive features of contemporary news production 321–2, 323–4
- epistemology of testimony 3, 223–4, 225
- fake news and digital media 287
- lack of diversity of news 2
- social epistemology 222–4
- spread of fake news 26–7, 286
- see also* Internet; social media; website

disinformation 1, 51, 286

- deceptive intention 35
- definition 35
- disinformation/misinformation distinction 35
- fake news and 6, 35, 157
- fake news environment 291
- as old phenomenon 156–7, 173
- public opinion manipulation: Disinformation by Epistemic Pollution model 11, 164–7, 173–4
- regulation of 174
- as threat to democracy 174
- tobacco industry 165
- US 79

Zimmermann and Kohring's fake news definition: recent disinformation 27, 35–7, 37*t*

distrust 181–2

- benefits of 257
- conservatives (self-identified as, US) 180, 182

conspiracy theories and institutional distrust 349–50, 351, 353, 355

of elites 180, 188–9

epistemic insensitivity and epistemic obstruction 184, 188, 200

of experts 11, 181, 184–5, 188

of expert testimony 188, 200

fake news and 38, 325

of liberals 180, 182, 200

of news 180, 200

of trust-networks 13, 345

*see also* expert/expertise

Dortmund fire (Germany) 21

echo chambers 2, 208, 225, 324

- algorithms 2
- autonomous dependence 209–11, 219, 220, 224
- conditionally independent 212
- content-driven problems 216, 217–19, 225
- degradation of democracy and 206, 214
- diverse viewpoints, lack of 214–16, 221, 322
- echo chambers/epistemic bubbles distinction 305
- epistemic problems of 207, 208–16, 219, 220–1, 225
- Facebook 208–9, 215, 216, 324
- fake news 217–19, 225
- main features of 207
- negative epistemic consequences 215
- news abstinence and 287, 304–5
- non-independence, lack of awareness of 209
- non-independence (lack of independence) 208–9, 211–14, 216, 221
- re-posting and 258–9
- self-policing of groups for dissenting voices 323
- social conditions and 7–8
- social epistemology 222–4
- social media/bots 2, 214, 219–22, 225
- solutions to 216
- Trump/Fox News relation 206, 207–8, 216, 217, 225
- unreliable echo chambers 11, 207
- US right-wing media 164, 171

elections 1, 292, 310

*see also* US 2016 elections

enquiry (scientific enquiry) 110

- alethic criterion* 110–11
- alethic telos* 110–11, 112
- background assumptions 113–14, 116–19
- epistemic filters 113, 114–16, 119–23, 120*f*, 121*f*
- epistemic normativity model 110–12

enquiry (scientific enquiry) (*cont.*)  
 evidence 111  
 normative deviance 112, 116, 117, 119, 131  
 normative deviance, varieties of 119–26  
 truth 110–11  
*see also* post-enquiry; science denialism

epistemic authorities 152, 344  
 blind trust in 140, 149  
 conspiracy theories 344, 346  
 definition 139  
 epistemic superiority 138  
 need of one's own judgment 137  
 PV (Preemption View) 140  
 as source of reasons, arguments, data 140  
 TEV (Total Evidence View) 140  
*see also* expert/expertise; PDR; PUCT; PV

epistemic autonomy 11, 159, 161–2, 164, 166–7

epistemic dependence 151, 159, 255, 311, 314  
 echo chambers, autonomous  
 dependence 209–11, 219, 220, 224  
 echo chambers, non-independence 208–9,  
 211–14, 216, 221

epistemic filters 120*f*, 121*f*  
 conspiracy theories 123, 124, 125, 126  
 definition 113  
 normative deviance and 119–25  
 omitting and discrediting 114–16, 123  
 post-enquiry 124–6  
 science denialism 10, 119, 126, 128–9, 130, 131  
*see also* enquiry; post-enquiry; science  
 denialism

epistemic heteronomy 159, 164, 167

epistemic investigation 185, 186*b*  
 direct and indirect epistemic factors 186*b*

epistemic responsibility 197, 276, 279, 282

epistemic trustworthiness *see* re-posting;  
 trustworthiness

epistemic vices 184–5  
 epistemic insensitivity 11, 184, 185–91,  
 193, 200  
 epistemic obstruction 11, 184, 191–7, 200

epistemic overconfidence 305

interpersonal vice 185  
*see also* expert/expertise

epistemic virtues 94, 245, 251  
 community-regarding epistemic  
 virtues 254–5  
 epistemic virtues and re-posting 251–5  
 other-regarding virtues 253–4  
 self-regarding virtues 253  
 testimonial justice 252, 260  
 trustworthiness 253, 254–5  
 virtue epistemology 252, 253, 255  
*see also* re-posting

epistemology of fake news 3–8, 265  
 approaches to 6–8  
 fake news, meaning of 6–7, 9  
 mechanisms that foster the production and  
 spread of fake news 7–8, 10–12  
 need for a new epistemology of fake news 2–3  
 scientific use of the term 3–5  
 therapies available as antidote to fake news 8,  
 12–13  
*see also* fake news: arguments against usage of  
 the term; fake news: definitions; fake news:  
 mechanics of production and spread; fake  
 news: therapies

epistemology of testimony 224, 265, 341–4  
 definition 2  
 dependence on testimony 341–2  
 digital age and 3, 223–4, 225  
 evaluating the epistemic credentials of  
 testimony 59–61, 62–3  
 fake news and 3  
 online testimonial exchanges 223–4  
 testimonial exchanges 222–4  
 testimonial reductionism 313, 325, 328  
 testimonial smothering 280–1  
 testimony and false belief 245  
 using testimony to deceive 59  
 vertical/horizontal epistemic evaluation 61, 63  
*see also* echo chambers; social epistemology

European Union 70–1, 174

expert/expertise 341  
 2016 US election 181  
 accessibility efforts are epistemically  
 valuable 198  
 conservatives (self-identified as, US) 180, 181,  
 182, 200  
 distrust of 11, 181, 184–5, 188  
 distrust, politics, and fake news 181–2  
 as epistemic authorities 139  
 epistemic insensitivity 11, 184, 185–91,  
 193, 200  
 epistemic insensitivity, negative impact  
 of 189–91, 200  
 epistemic insensitivity and  
 defensiveness 188–9, 200  
 epistemic obstruction 11, 184, 191–7, 200  
 epistemic obstruction and talking past each  
 other 193–5  
 epistemic sensitivity 187, 189  
 experts as 'liberals'/'liberal elites' (US) 180–1,  
 182, 200  
 expert testimony 10, 188, 200  
 feigned disagreement and epistemic  
 excommunication 195–7  
 interpersonal vice 185

liberal bias 182–4, 200  
 objections to claims of epistemic insensitivity  
     and epistemic obstruction 197  
 polarization and 184, 190, 200  
 reliance on 136–7, 341, 342, 344  
 why experts cannot circumscribe the epistemic  
     scope 199  
 why experts carry the burden of  
     clarification 199  
*see also* epistemic authorities

Facebook 2, 269  
     2016 US elections and fake news 2, 316, 326  
     censorship by 65  
     echo chambers 208–9, 215, 216, 324  
     news feed algorithm 170, 172  
*see also* social media

fact-checking 77–8, 216, 301, 325, 326

fake news  
     audience 265, 276, 288–9, 320  
     background assumptions and 126  
     novel information and 2, 292  
     as political tool 2, 9, 35  
     post-enquiry and 131  
     power and 9, 69, 317  
     public opinion manipulation 159, 165–6, 173  
     science denialism and 10, 109–10, 126–9,  
         130, 131  
     veridical fake news 288, 289  
     weaponization of the term 69, 166, 318, 324–5  
*see also* epistemology of fake news

fake news: arguments against usage of the  
     term 68–70, 79–81, 287n.1, 310, 317, 318

academia and the ‘science of “fake news”’ 74–8  
 epistemic panic about ‘fake news’ 68, 71,  
     79–80

‘fake news’ as objectionable term 69, 81

‘fake news’ as instrument of corporate  
     censorship 71–2

‘fake news’ as instrument of state  
     censorship 70–1

‘fake news’ as label used to discredit  
     democratic disagreement 9

fake news and the Russian threat 79

objections from semantic instability,  
     redundancy, and propagandistic usage  
     4–5, 318

polysemy of the term 318

unsatisfactory nature of extant definitions  
     70, 157

*Washington Post* and ‘fake news’ 72–4

fake news: definitions 9, 41, 58, 75, 78, 157, 182,  
     265, 287–8, 291, 315–21  
     consumer-centered view 6–7

Dentith’s definition 27, 32–3, 37t, 38, 288  
 effective/ineffective fake news 56–8, 60–1  
 as epistemically corrupt news 3  
 fake news landscape 23, 24f  
 Gelfert’s definition 27, 29–31, 37t, 38, 47–9,  
     248, 288, 320–1

Google’s definition 72

hybrid view 6

importance of defining fake news 38–41, 58

Jaster and Lanius’ definition 9, 19–23, 23t,  
     37t, 41

Lazer’s definition 74–5

as mimicking genuine news 6, 9, 26, 38, 41, 74,  
     158, 166, 173, 246–7

as misleading news 4, 5, 46, 47, 58,  
     289–90, 291

Mukerji’s definition 27, 33–5, 37, 37t, 38, 288,  
     289, 320

as news lacking truth 6, 9, 20, 21–2, 23, 23t,  
     25, 27, 38, 287–8

as news lacking truthfulness (truth  
     indifference) 6, 9, 20, 22–3, 23t, 26, 27, 38,  
     287–8, 289

privative view 6

receiver-based definitions 288–9

related phenomena 23–5

Rini’s definition 27, 28–9, 37t, 38, 48n.6, 126,  
     248, 288

sender-based definitions 288–9

unsatisfactory nature of extant definitions  
     70, 157

Zimmermann and Kohring’s definition 27,  
     35–7, 37t

*see also* fake news: dimensions of

fake news: dimensions of 25–7, 27t, 37t, 41, 75  
 appearance dimension 26, 28, 29, 33, 38, 41,  
     57, 58

bullshit dimension 26, 29, 30, 31, 32, 32, 33,  
     38, 265, 288, 291, 320

deception dimension 4, 6, 20, 22, 25–6, 27, 28,  
     29, 31, 32, 33, 34, 37, 41, 126, 248, 288,  
     291, 320

effect dimension 26, 29–30, 31, 38

media dimension 26–7, 28, 29, 31, 38

truth dimension 25, 28, 29, 31, 32, 33–4,  
     37, 41

virality dimension 26, 38

fake news: effects and harms 292  
     distrust 38, 325  
     epistemic harms 248, 265–7  
     epistemic harms: Modified Revised Model of  
         communication 277–81  
     epistemic harms: Revised Model of  
         communication 272–6

fake news: effects and harms (*cont.*)

- epistemic harms: Simple Model of communication 267–71
- epistemic rot 12, 281–2
- erosion of truth and truthfulness 39
- failing to acquire a true belief 266
- false beliefs 22, 30, 31, 34, 38, 248, 266, 287, 289, 290, 291
- illegitimate collective decisions 39
- misguided trust 248–9
- political debate 316
- undermining the epistemic standards of news 39
- uninformed decisions 38, 286, 292
- violence 38

fake news: mechanics of production and spread 2, 10–12, 228, 281

- agents of misinformation 135
- communication technology features 7
- digital media 287
- epistemic ideologies 7, 8
- falsehood diffuses farther, faster, deeper, and more broadly than the truth 218, 326
- local monitoring 273–4, 275, 276
- social conditions and 7–8

*see also* fake news distributors; public opinion manipulation

fake news: therapies (remedial strategies) 8, 12–13, 41, 310

- background gardening 12, 277–80
- boundary work 9, 40–1
- critical thinking 8, 310, 325–6
- curtailment of fake news 64–5
- debunking, effectiveness of 257, 293, 326
- democratic freedoms 65
- epistemic responsibility 276, 279, 282
- identifying fake news 9, 60–2
- individual responses 9, 12, 59–64, 65, 282, 292, 325
- non-reliance on dubious news or news from dubious sources 292
- paradigm repair 9, 40–1
- ‘prebunking’ 326–7
- promoting media literacy 292
- regulation of fake news and media markets 8, 64–5, 70–1, 174, 292, 310, 325
- self-regulation by media companies 292
- structural responses 9, 64–5, 292
- technical expertise 63n.21, 325

*see also* fact-checking; intellectual virtues; news abstinence

fake news distributors 23, 41

- deceptive intention 37
- ‘Firehose of Falsehood’ 39
- lack of truthfulness 31

fake news as novelty 158, 173–4, 319

- fake news production, distribution, consumption 156, 169–73, 174, 287
- fake news as social-epistemic dysfunction 13, 310

Internet 156, 169, 170–3, 174

- public opinion manipulation models 158–9
- putative novelty of fake news 167–73
- social media 156, 170–3, 174

*see also* public opinion manipulation

fake news as old phenomenon 2, 11, 156–7, 167–8, 174, 247, 310

- fake news content 156, 168, 173
- fake news proliferation, circulation, influence 156, 169, 173, 321
- fake news/propaganda relation 158
- fourteenth century 287
- public opinion manipulation 168

*see also* public opinion manipulation

fake science 228

- fake news/fake science distinction 25n.17, 228

*see also* Google and retracted articles

false belief 80, 117–18, 302

- apparent news sources as sources of false belief 246–9
- correction of 257, 260–1

fake news and 22, 30, 31, 34, 38, 248, 266, 287, 289, 290, 291

news abstinence: acquiring false/irrelevant beliefs as condition for 13, 286, 287, 294, 299–300, 301–2, 306

- re-posting and 245, 251, 255, 259
- self-deception and 296
- testimony and 245

filter bubbles 316

- algorithms 2, 7, 135, 323
- Internet 2, 7
- re-posting and 258
- social media 2, 7

flat earth 83, 112, 124, 126–8, 131

Fox News 166, 171, 193n.15

- 2016 US elections and 208, 217
- conspiracy theories 207
- echo chamber: Trump/Fox News relation 206, 207–8, 216, 217, 225
- fake news 217–18
- as right-biased 207–8, 217, 322
- truth of statements made by 217
- US Conservatives and 208, 217–18

Frankfurt, Harry 74

- bullshit, definition 22, 29, 49, 249, 288

*see also* bullshit

free press 9, 64

free speech 71, 161, 215

Gelfert, Axel 289  
 clickbait and fake news 31, 48–9, 320  
 fake news, definition 27, 29–31, 37t, 38, 47–9, 248, 288, 320–1  
 fake news: ‘misleading by design’/intention to mislead 27, 29–31, 37t, 38, 47–51, 56, 288, 311, 320–1, 324  
 fake news as genuine form of news 51–2, 55  
 fake news and manipulation 320, 321, 324  
 problematic features of Gelfert’s definition of fake news, 50–2, 56  
*systemic* dimension of fake news 13, 310, 311, 320–1, 323–4  
*see also* knowledge maintenance

German Democratic Republic: potato beetle plague 24–5, 35

Goldman, Alvin 211–13, 214, 300

good news (genuine news)  
 accurate information 1, 58  
 good epistemic pedigree 54, 60  
 good news/bad news distinction 59  
 good news/effective fake news distinction 60–1, 64  
 good news/fake news distinction 55–6, 58, 59, 64, 78, 321  
 good news sources 53, 54–5, 60

Google 248  
 censorship 72  
 definition of fake news 72  
 PageRank 172  
 ranking 229–30  
 search results 170, 172, 230  
 secretive search algorithm 72  
*see also* search engines

Google and retracted articles 239–40  
 algorithms 11–12, 229, 240  
 Google Scholar 12, 229, 232–7, 235f, 236f, 237f, 240  
 Google Search 229, 232–3, 234–7, 235f, 240

Law of Retraction 12, 230–1, 239  
 outsourcing the handling of retractions 237–8, 240

PageRank algorithm 229–30, 239  
 propagating discredited scientific data/fake science 11–12, 228–9, 231, 238, 239, 240  
 ranking based on popularity 229, 238, 239  
 ranking of retracted articles 228–9, 233–7, 234f, 236f, 239–40  
 research methodology 231–3, 239–40  
 research results and discussion 233–9  
 retraction due to scientific fraud 229, 238  
 Retraction Watch public database 229, 231, 232, 239–40  
 solving the problem 238–9, 240

top link heuristics 236–7  
 why retracted articles may be a deep problem 229–31, 238

Greene, David 75–6

Habermas, Jürgen 39, 137

Habgood-Coote, Joshua 4, 5, 157, 287n.1, 318

heresy 69

heuristics 256, 271, 324, 328, 343  
 heuristic rules 342, 345, 347–8  
 ‘top link heuristics’ 236–7

Hume, David 312, 338

information cascades 135, 276

information technology 2, 287

information warfare 165, 171, 174

Instagram 7  
*see also* social media

intellectual virtues 62–4, 65  
 cultivation of 9, 62  
 fake news and 78  
 ‘golden mean’ 62, 63  
 honesty 9, 62  
 intellectual conscientiousness 9, 62  
 intellectual humility 9, 62, 78  
 love and desire for the truth 62  
 non-intellectual/moral virtues and 63–4

Internet  
 fake news and 2, 156, 169, 170–3, 174  
 fake news: mechanisms fostering the production and spread of 7  
 filter bubbles 2, 7  
 freedom, equality, public discourse, and democracy 170–1  
 public opinion manipulation 156, 169, 170–3, 174  
*see also* algorithms; platforms; website

Internet Research Agency 247

James, William 80

Jaster, Romy: definition of fake news 9, 19–23, 23t, 37t, 41

journalism 319  
 bad journalism 6  
 boundary work 40–1  
 decline of professional journalism 2, 170  
 good journalistic practices 54  
 journalistic balance 8, 135–6  
 journalistic objectivity and balance 39, 75–7  
 paradigm repair 40–1  
 partisan/non-partisan journalism 76–7  
 political neutrality 76–7  
 professionalisation of 77

journalism (*cont.*)  
 quest for truth as cornerstone of 34  
 truth and truthfulness as central values of 39, 41

Kant, Immanuel 136–7, 198

Keeley, Brian L. 89–90, 91, 98, 99, 100

knowledge maintenance 312  
 acquiring knowledge and maintaining coverage through epistemic routines 311–15  
 belief expansion/routine belief expansion 312–13, 328  
 coverage reliability, lack of 13, 311  
 epistemic coverage 310, 311, 314, 321–2, 325, 330  
 epistemic routines to vindicate trust in reliable sources 13, 310, 311, 328–30  
 ‘fake news’ in the post-truth regime 315–21  
 fake news as social-epistemic dysfunction 13, 310–11  
 knowledge maintenance in a fake news age 325–30  
 news production and consumption under conditions of the ‘attention economy’ 321–5  
 social-informational environment 311, 314, 315  
*systemic* distortions of creating and disseminating news-like content 13, 310, 311, 320–1, 323–4

Kohring, Matthias: definition of fake news 27, 35–7, 37t

Lanius, David: definition of fake news 9, 19–23, 23t, 37t, 41

Lazer, David M. J. 69, 74–5, 76, 80

left-wing politics *see* expert/expertise

Levi, Isaac 312–13, 317, 328

lying 34  
 bullshit/lying distinction 49, 50n.10, 74  
 fake news/lies distinction 34  
 ‘knowledge-lies’ 289n.3  
 as weapon of the weak 74

Lynch, Michael 113, 136–7

Macedonian teenagers 32, 35, 38  
 2016 US elections 6, 22, 23t, 248, 320  
 bullshit fake news 22  
 as clickbait farmers 6, 22, 31, 320  
 lack of political agenda 24

Macron, Emmanuel 71

Mathathir, Mohamad 71

Mass, Heiko 71

McIntyre, Lee 73, 326

media channel 1, 2

media deregulation 76–7

media literacy 8, 292

media markets 322, 325

Meyer, Marco 78

Mill, John Stuart 170, 214–15

Miller, Judith 74

Mills, Charles 222, 223

misguided trust  
 apparent news sources as objects of misguided trust 246–9  
 fake news and 248–9  
 re-posting 246, 251–2, 260  
 resilience to correction 260

misinformation 1, 156, 286  
 agents of 135  
 continued influence of 292, 301, 326  
 definition 35  
 disinformation/misinformation distinction 35  
 fake news 9, 11, 55–6, 255–6, 291  
 fake news environment 291  
 news abstinence and 287, 303–4  
 as old phenomenon 174  
 platforms and 172  
 ‘prebunking’ 326–7  
 public opinion manipulation 162, 167, 172  
 resilience to correction 256, 258, 326  
 as threat to democracy 156  
*see also* public opinion manipulation

misleading  
 bullshit 50  
 Dentith’s fake news definition 27, 32–3, 37t, 38  
 fake news, misleading by commission and omission 290, 291  
 fake news, misleading effect of 4, 5, 46, 47, 58, 289–90  
 Gelfert: fake news as ‘misleading by design’/intention to mislead 27, 29–31, 37t, 38, 47–51, 56, 288, 311, 320–1, 324  
 literal truth can be misleading 46, 47, 58, 289–90  
 misleading defeater 296–8, 300  
 receiver-based accounts of fake news 288–9

moon landing (as fake) 82, 83, 124, 127

Mukerji, Nikil  
 definition of fake news 27, 33–5, 37, 37t, 38, 288, 289, 320  
 bullshit/lying distinction 50n.10

Murrow, Edward R. 77

Nazi Germany 150, 162, 165, 173

news 20, 52  
 accuracy and epistemic quality of 1

media channel and 1, 2  
as truth warranting utterances 20

news abstinence 11, 12–13, 286, 306  
acquiring false/irrelevant beliefs as condition for 13, 286, 287, 294, 299–300, 301–2, 306

bias-inducing information 296, 298, 299

cognitively difficult/time consuming to discriminate genuine from fake news as condition for 12–13, 286, 300–1

defeasibility framework of justification 286, 298, 302–3

epistemic consequentialism 286–7, 301–2

epistemic overconfidence as challenge to 305

fake news environment 287, 290–1

fake news environment as condition for 12, 286, 299–300, 302

justification of motivated news  
abstinence 12–13, 286, 287, 299–303, 306

as lesser harm defense 306

misleading defeater 296–8, 300

motivated ignorance 287, 292–6

motivated ignorance, epistemic value of 296–9

motivated ignorance/self-deception distinction 295–6

motivated ignorance/willful ignorance distinction 294–5

reliabilism about justification 286–7, 301–2

trivial truth 296, 298–9

undercutting defeater 302

willful ignorance 293–4

news abstinence: objections  
news abstinence as catalyst for fake news and misinformation 287, 303–4

news abstinence gives rise to echo chambers 287, 304–5

news consumers/consumption 8, 324  
distinctive features of contemporary news consumption 321–3

fake news, consumer-centered view of 6–7

news production and consumption under conditions of the ‘attention economy’ 321–5

personalized exposure to news 2, 7

news distributors/distribution 8, 20

news producers/production 8  
democratization of production 2  
distinctive features of contemporary news production 321–3

news production and consumption under conditions of the ‘attention economy’ 321–5

news sources 52–4  
accurate information 53, 54, 292  
apparent news sources as sources of false belief and objects of misguided trust 246–9

bad news sources 55

dubious sources 292

good news sources 53, 54–5, 60

liberal bias 182–4, 200

online world as source for print and broadcast media 323

original source 214, 251, 260, 265, 267, 268–70, 292

partisan news sources 200

vertical/horizontal evaluation of 61

*New York Times* 74, 76, 317

NPR (National Public Radio) 75, 76

NSA (National Security Agency) 146, 334, 335

Obama, Barack 322

objectivity 316  
journalistic objectivity and balance 39, 75–6, 77

Oreskes, Michael 75

*Oxford English Dictionary* 20, 134

partisanship 171, 190, 200, 304, 316–17  
broadcasters 322–3  
partisan/non-partisan journalism 76–7

PDR (Principle of Democratic Reason) 136, 137–8, 152

amendment in light of PV 149

conspiracy theories and 138, 146–9, 152

dangerous epistemic consequences for public opinion formation 138, 146–9

fault from an epistemic point of view 138–46

PV (Preemption View) as alternative to 138  
*see also* epistemic authorities

Pew Research Center 208, 217, 220

philosophy 41, 157, 310

pictures 24n.15, 168

Pizzagate 21, 22, 23t, 25

platforms 172  
algorithms 172  
misinformation 172  
self-regulations against fake news 8  
*see also* Internet

Plato 140, 159–60, 161

polarization  
experts: epistemic insensitivity and epistemic obstruction 184, 190, 200  
group polarization 7, 184  
ideological polarization 2  
prevention of 8

PolitiFact 77

Pomerantsev, Peter 164, 165, 171, 173

post-enquiry 10, 109  
*alethic telos* 117, 124, 125  
background assumptions 116–17

post-enquiry (*cont.*)

- conspiracy theories 124, 125
- definition 125, 131
- epistemic filters 124–6
- fake news and 131
- normatively aberrant enquiries 116, 117, 119, 125–6, 131
- science denialism 10, 109, 126, 128, 131
- see also* enquiry; science denialism

post-truth era/phenomenon 10, 156, 167

- bullshit 134
- definition 134
- detachment of public opinion from truth 134
- emotion and personal belief as truth-unrelated factors 134
- Enlightenment principles and 10, 136
- fake news 156
- ‘fake news’ in the post-truth regime 315–21
- ideologies 136
- irrational public opinion 134
- ‘post-truth politics’ as old phenomenon 173, 174
- sources of 10, 135–6
- value of truth and reality 34
- see also* expert/expertise

propaganda 4, 5, 11, 69, 70

- deceitful propaganda 34
- definition 24, 162
- fake news propaganda as bread-and-butter of totalitarian regimes 163
- fake news/propaganda distinction 24
- fake news/propaganda relation 25, 34, 157, 158
- as old phenomenon 156–7, 174, 310
- public opinion manipulation: Propaganda and Censorship model 11, 162–4, 173, 174
- as threat to democracy 156
- see also* public opinion manipulation

public opinion

- agents of misinformation 135
- individual and social biases 135
- irrational public opinion 134
- PDR (Principle of Democratic Reason): dangerous epistemic consequences for public opinion formation 138, 146–9
- post-truth and 134
- structural features of communication leading public opinion away from truth 135–6
- truth-unrelated factors leading public opinion away from truth 134–6

public opinion manipulation 158–9, 173–4

- epistemic autonomy 11, 159, 161–2, 164, 166–7
- epistemic dependence 159

fake news 159, 165–6, 173

Internet 156, 169, 170–3, 174

misinformation 162, 167, 172

Model A: Pleasing and Seducing the Audience 11, 159–62, 173, 174

Model B: Propaganda and Censorship 11, 162–4, 173, 174

Model C: Disinformation by Epistemic Pollution 11, 164–7, 173–4

social media 156, 165, 170–3, 174

*see also* fake news as novelty; fake news as old phenomenon

public sphere 156, 166, 170

PUCT (Principle of Unrestricted Critical Thinking) 10, 136, 137–8, 152

- amendment in light of PV 149
- conspiracy theories and 152
- see also* epistemic authorities

Putin, Vladimir 79, 165, 247

PV (Preemption View) 140

- as alternative to PDR (Principle of Democratic Reason) 138
- arguments in support of PV
- restrictions 149–52
- conspiracy theories and 148–9
- deference to authorities while developing reasoning skills 151–2

Higher-Order Undercutting Defeat Argument 143–6

Track Record Argument 141–3, 141*t*

*see also* epistemic authorities

Rawls, John 222, 299

reductionism 3

- testimonial reductionism 313, 325, 328
- referenda 1, 292
- 2016 Brexit referendum 1, 160

re-posting 245, 261–2

- benefits of 258–9, 260
- as bent testimony 250, 254–5
- credentialing 245–6
- echo chambers and 258–9
- epistemically trustworthy person and 12, 255, 259–62
- epistemic risk: ‘bent credentialing’ 249–51, 255
- epistemic risk: false belief 245, 251
- epistemic risk: misguided trust 246, 251–2, 260
- epistemic virtue, call for 251–2
- epistemic virtues relevant to re-posting 253–5
- filter bubbles and 258
- historical analogs of 250

psychological evidence about the risks of 255–8

trustworthiness and 3, 12, 253, 254–5, 259  
*see also* social media

right-wing politics 134

fake news and US right-wing politics 171

Fox News as right-biased 207–8, 217, 322

US right-wing media bubble/echo chamber 164, 166, 171  
*see also* expert/expertise

Rini, Regina

bullshit dimension of fake news 26

definition of fake news 27, 28–9, 37t, 38, 48n.6, 126, 248, 288

re-posting as ‘bent kind of testimony’ 250, 254–5

Roberts, Cokie 75–6

Russia

2016 US elections 316–17

fake news 165–6, 171

‘fake news’ as instrument of state censorship 70

fake news and the Russian threat 79

information warfare 171

Pizzagate and 21, 22, 25

public opinion manipulation 162, 173

*Russia Today* 167

Ukraine 166, 247

*Washington Post* 73

satire

*The Babylon Bee* 249

*Colbert Report* 316, 318

fake news/satiric news parodies relation 23–4

*The Onion* 4, 47, 51, 249

*Private Eye* 47

satiric news parodies 4, 23, 48, 51, 58, 249, 316, 318

science

scientific fraud 228, 229  
*see also* fake science; Google and retracted articles; science denialism

*Science* (journal) 228

science denialism 10, 109, 131

*alethic telos* 117–18, 125, 128, 129

background assumptions 117–19, 125, 126, 128

case study: anti-vaxxer 112, 124, 128–31

case study: flat earth 112, 124, 126–8, 131

climate change deniers 165, 216, 348–9

conspiracy theories and 123, 124–5, 126, 128–9, 131

defeater 111, 115, 116, 119, 122, 125

epistemic filters and 10, 119, 126, 128–9, 130, 131

fake news and 10, 109–10, 126–9, 130, 131  
as normative aberration 109, 112, 125, 127, 128, 131

post-enquiry 10, 109, 126, 128, 131

scientific enquiry and 109  
*see also* enquiry; post-enquiry

search engines 277

retraction in 229, 238–9

top link heuristics 236–7  
*see also* algorithms; Google; Google and retracted articles; Yahoo

Shepard, Alicia C. 76

Silverman, Craig 157

skepticism 63–4, 269, 315, 326

conspiratorial scepticism 335, 336, 337, 340, 350, 351

Snopes 77

Snowden, Edward 334, 340

Snyder, Timothy 165–6, 171

social epistemology 102, 222–4, 253, 255, 274, 330

non-ideal social epistemology 11, 161, 207, 224, 225

social media

distinctive features of contemporary news production and consumption 321–2, 323

fake news as novelty 156, 170–3, 174

fake news as political tools and 2

filter bubbles 2, 7

plurality, diversity, democracy and 171

public opinion manipulation 156, 165, 170–3, 174

social media bots 174, 207, 219–22, 225

spread of fake news 2, 26, 28, 136, 326  
*see also* algorithms; Facebook; Instagram; re-posting; Twitter

Sputnik 35

*The Sun* 248

Sunstein, Cass 206, 214, 216, 353–4

Talisse, Robert 57, 310

terrorism 69, 70, 74, 293

terrorist attack on 9/11 (as inside job) 82, 92–3, 124, 147, 334

Thomas of Monmouth 247, 250

tobacco industry 124, 158, 165

totalitarianism 162–4

troll factories/farms 158, 171, 174

Trump, Donald 75–6, 79, 166, 188

2016 US elections 6, 169, 208, 247, 248, 316–17

bullshit fake news 22, 23t

Trump, Donald (*cont.*)  
 echo chamber: Trump/Fox News relation 206, 207–8, 216, 217, 225  
 ‘fake news’, the term 4, 68–9, 157, 166, 317  
 lack of truthfulness 29  
 Trump supporters 189, 196–7, 217–18  
 truth of statements made by 219  
 Twitter 219  
 trust-network 336, 351, 354, 355  
 conspiracy theories and 13, 345–6, 347, 351, 353–4, 355  
 division of cognitive labour and role of trust-networks 341–4  
 trustworthiness  
   as epistemic virtue 253, 254–5  
   re-posting and 3, 12, 253, 254–5, 259  
   trustworthiness of an informant 341–2, 343, 345  
   *see also* epistemic trustworthiness  
 truth  
   clickbait and 249  
   diversity of opinions and 215  
   fake news: truth dimension of 25, 28, 29, 31, 32, 33–4, 37, 41  
   fake news as news lacking truth 6, 9, 20, 21–2, 23, 23t, 25, 27, 38, 287–8  
   intellectual virtue: love and desire for the truth 62  
   journalism and 34, 41, 39  
   scientific enquiry and 110–11  
   trivial truth 296, 298–9  
   truth dimension in fake news definition 25, 28, 29, 31, 32, 33–4, 37, 41  
 truthfulness  
   bullshit: lack of truthfulness 22, 29, 49–50  
   fake news distributors, lack of truthfulness 31  
   fake news as news lacking truthfulness (truth indifference) 6, 9, 20, 22–3, 23t, 26, 27, 38, 287–8, 289  
   journalism and 39, 41  
   persuasiveness rather than truthfulness as guiding principle in democratic systems 161  
   ‘truthiness’ 316  
 Twitter 169, 208, 269, 316  
   censorship by 65  
   fake news 218–19  
   falsehood diffuses farther, faster, deeper, and more broadly than the truth 218  
   retweeting 172, 219  
   rumor cascades 218  
 Trump, Donald 219  
   *see also* social media  
 UK (United Kingdom): Brexit 1, 38, 160  
 US (United States) *see* expert/expertise; right-wing politics; Trump, Donald  
 US 2016 elections 181  
   Facebook 2, 316, 326  
   fake news and 1, 2, 4, 22, 23t, 68, 166, 169, 247, 316–17, 326  
   Fox News and 208, 217  
   Macedonian teenagers 6, 22, 23t, 248, 320  
   Russia 316–17  
 veritism 245, 298, 311  
 videos 82, 172–3, 301  
 virtues *see* epistemic virtues; intellectual virtues  
*War of the Worlds* radio broadcast 249  
*Washington Post* 72–4, 76, 317  
 website 21, 220  
   fake news websites 57, 78, 157, 171, 324, 325  
   *see also* clickbait; Google and retracted articles; Macedonian teenagers; troll factories/farms  
 WikiLeaks 21  
 Yahoo 7  
 yellow journalism 157n.1, 168, 172  
 YouTube 82, 170, 172–3  
 Zimmermann, Fabian: definition of fake news 27, 35–7, 37t