UNIVERSITA' VITA-SALUTE SAN RAFFAELE

CORSO DI DOTTORATO DI RICERCA IN FILOSOFIA

Curriculum in Discipline Filosofiche

Trust, Science and Disinformation

Tutor: Professor Carlo Martini

Co-Tutor: Dottor Piero Ronzani

Tesi di DOTTORATO di RICERCA di Luca Ausili

matr. 020164

Ciclo di dottorato XXXVII

SSD M-FIL/02 LOGICA E FILOSOFIA DELLA SCIENZA

Anno Accademico 2024/2025

CONSULTAZIONE TESI DI DOTTORATO DI RICERCA

Il sottoscritto Luca Ausili

Matricola: 020164

Nato ad Ancona

Il 06/12/1993

Autore della tesi di dottorato dal titolo

Trust, Science and Disinformation

□ AUTORIZZA la consultazione della tesi

☑ NON AUTORIZZA la consultazione della tesi per 12 mesi a partire dalla data di conseguimento del titolo e precisamente dal 30/01/2025 al 30/01/2026.

Poiché:

□ l'intera ricerca o parti di essa sono potenzialmente soggette a brevettabilità/ *The whole project or part of it might be subject to patentability;*

☑ ci sono parti di tesi che sono già state sottoposte a un editore o sono in attesa di pubblicazione/ *Parts of the thesis have been or are being submitted to a publisher or are in press;*

□ la tesi è finanziata da enti esterni che vantano dei diritti su di esse e sulla loro pubblicazione/ *the thesis project is financed by external bodies that have rights over it and on its publication.*

È fatto divieto di riprodurre, in tutto o in parte, quanto in essa contenuto / / *Copyright the contents of the thesis in whole or in part is forbidden*

Data/Date: 15/01/2025

Firma/Signature

Luca Ausili

DECLARATION

This thesis has been:

- composed by myself and has not been used in any previous application for a degree. Throughout the text I use both 'I' and 'We' interchangeably.
- has been written according to the editing guidelines approved by the University.

In Chapter 1, section 1.3 and subsections 1.3.1, 1.3.2, 1.3.3, and section 1.4, and subsection 1.4.1 are based on: Ausili, L., The Inseparable Link between Disinformation and Attitudes, under review at *Inquiry.*

In Chapter 5, sections 5.3, 5.4, subsection 5.4.1, section 5.5, and subsections 5.5.1 and 5.5.2 are based on: Ausili, L. (2024) The Political and Socio-Epistemic Risks of Quantification, *Topoi*, https://doi.org/10.1007/s11245-024-10092-1.

All sources of information are acknowledged by means of reference.

Acknowledgments

In the past three years, I have had the chance to pursue a research project to which I have dedicated all the energies, competences, and passion that I have. Now that this project comes to an end, I cannot conclude it without mentioning the people who have been part of it, and without whom I could not have accomplished any result.

First, I must thank my supervisor, Professor Carlo Martini, for his endless support, enthusiasm, for having allowed me to constantly work with joy, calm and passion, and for having taught me that doing research is not having evocative or fancy ideas, but being able to divide an issue in all its components and study each of them with meticulousness and scientific precision. Then, I thank my co-supervisor, Doctor Piero Ronzani, for his support, time, comments, countless reading recommendations, for his pieces of advice on how to improve in academic writing, and for having given me the basis of empirical research.

During these years, I have had the great opportunity to spend an entire year at the University of Helsinki. Here, I have found a wonderful research environment. During the weekly TINT seminars and daily conversations, I have discovered new and useful research lines and related literatures that have majorly contributed to the present thesis. In particular, I must thank my supervisor at the University of Helsinki, Professor Caterina Marchionni, for her kindness, support, countless conversations, and for having helped me to improve as a researcher. Among the wonderful people that I have known in Helsinki, I would also like to thank Charlie Kurth, Michiru Nagatsu, Jaakko Kuorikoski, Säde Hormio and Teemu Lari who have given very useful comments to my work, and who have made my year in Helsinki special and joyful.

During these doctoral years, I have also had the wonderful opportunity to collaborate, for six months, with the association BergamoScienza, as part of my mandatory stage. During these months, I have been able to understand how impactful and important their activities are, and also how complex it is to design and structure activities that allow scientifically laypeople to discover science and scientific research in an efficient and entertaining way.

I must thank all the members of the association, and particularly Nicola Quadri, who has helped and guided me with great kindness during my months with BergamoScienza, and Professor Gianvito Martino, who with his endless enthusiasm has made me very passionate about the activities of the association.

I must also thank all the members of the Faculty of the Department of Philosophy at Vita-Salute San Raffaele University who have dedicated their time to me for conversations and discussions about common research topics. In particular, Doctor Bianca Cepollaro, Professor Francesca Boccuni, Professor Francesca Pongiglione, Professor Roberta Sala, Professor Claudia Bianchi, Professor Roberto Mordacci, and Professor Francesco Valagussa.

A special thank goes to my colleagues and friends, who have been part of this journey with me, and without whom these years would not have been as magical as they have been. In particular, I owe a debt of gratitude to the following (in Alphabet order): Alessandro Anzà, Alessandro Volpe, Alessandro Volpi, Davide Battisti, Eleonora Volta, Federico Bina, Gianluca Gemmani, Giorgia Adorno, Giulio Pennacchioni, Giuseppe De Ruvo, Francesca Cesarano, Mara Floris, Margherita Ghiara, Maurizio Mascitti, Nicolò D'Agruma, Paolo Colizzi, Paolo Polizzi, Simone Marsilio, Stefano Pinzan.

In conclusion, my eternal love and gratitude goes to my parents, who have always supported me throughout this journey, to Blaise, who has always watched over me, until the last day, and to Allegra, the strongest and bravest person I know.

Abstract

In this dissertation, my aim is to analyse the reasons and causes behind the spread of disinformation. In particular, my main interest is in what I will refer to as "scientific disinformation", namely the kind of disinformation that regards scientific issues and topics, with a particular focus on climate change. The thesis is divided into five chapters, each with a specific purpose. The first chapter is dedicated to the definition of disinformation. I review the different definitions of disinformation available in literature and argue that it is a necessary condition, for a content to be defined as disinformative, that its creator shares it to mislead the audience. In the second chapter, I analyse the role that the digital environment, and particularly the development of social media platforms, has in the proliferation of disinformation online. I argue that the current digital environment represents a fertile soil for the proliferation of disinformation, however, the spread of the latter is not directly and entirely imputable to it. The deeper reasons and causes for the creation and spread of disinformation are indeed antecedent to the recent technological developments, and inextricably dependent on the behaviour of the users. In the third chapter, I criticise the literature on disinformation for being too anchored in an idea of the digital environment as subordinate to the offline one. I argue that, while the literature on disinformation is rightfully dedicated to verifying the hypothesis that disinformation proliferates as a result of people's social identity protective attitudes, more research efforts should be invested on the possibility that sharing and receiving disinformation is related to people's social identity formation processes. In the fourth chapter, I consider the possibility that the spread of scientific disinformation is caused by a diffused distrust in science. I argue that this is largely related to non-epistemic reasons, but also to epistemic ones, and I identify the particular social groups that are more likely to distrust science and share scientific disinformation. In the fifth and last chapter, I consider the possibility that the current communication between science and society presents some shortcomings that could have indirectly facilitated the formation of scientific disinformation. In the conclusive section, I review the various passages of the thesis, the positions endorsed and indicate possible future developments and lines of research.

Abstract (Italiano)

L'obiettivo di questa tesi è fornire un'analisi delle ragioni e cause della proliferazione di disinformazione. Nello specifico, mi concentrerò su un tipo particolare di disinformazione che verte su temi e questioni di natura scientifica, e che per questo chiamerò "disinformazione scientifica", con particolare attenzione per la disinformazione sul cambiamento climatico e la scienza climatica. La tesi è composta da cinque capitoli. Il fine del primo capitolo è quello di definire in maniera chiara e precisa il concetto di disinformazione. Per farlo, presenterò le varie definizioni di disinformazione che sono state proposte in letteratura, e difenderò una posizione che ritiene sia una condizione necessaria, perché un contenuto possa dirsi disinformativo, che chi lo crea e diffonde abbia intenzione di ingannare la propria audience. Nel secondo capitolo analizzerò il ruolo dell'ambiente digitale, e in particolar modo dei social media, relativamente alla proliferazione di disinformazione. Sebbene alcune caratteristiche rendano l'ambiente digitale un contesto comunicativo all'interno del quale la proliferazione di disinformazione è facilitata, ritengo tuttavia che questo non abbia direttamente e totalmente causato questo fenomeno. Le ragioni profonde della creazione e diffusione di disinformazione sono infatti antecedenti allo sviluppo delle piattaforme digitali, e strettamente legate alla volontà, alle credenze, e al comportamento dei loro utenti. Nel terzo capitolo muovo una serie di critiche alla letteratura sulla disinformazione. In particolare, sostengo che la letteratura è ancora troppo ancorata a una visione secondo la quale l'ambiente digitale è subordinato a quello offline, e che pertanto si ritenga che online le persone si limitino a interagire sulla base delle credenze formate fuori dalla rete. Sebbene l'ipotesi che le persone condividono disinformazione a causa di una propensione a proteggere la propria identità sociale e di gruppo sia supportata dai dati, io ritengo che la ricezione e condivisione di disinformazione siano pratiche che hanno un grande impatto sulla formazione delle identità sociali, e che maggiore attenzione debba essere posta su questo aspetto. Nel quarto capitolo, considero la possibilità che la proliferazione di disinformazione scientifica sia causata da una diffusa sfiducia nei confronti della scienza e delle istituzioni scientifiche. Sostengo che questa sfiducia è soprattutto legata a questioni di natura non epistemica, ma in alcuni casi anche epistemica, e definisco i gruppi sociali che sono più inclini ad accettare e condividere disinformazione scientifica.

Nel quinto e ultimo capitolo, prendo in considerazione la possibilità che la proliferazione di disinformazione scientifica sia stata causata, indirettamente, da alcune inefficienze nella comunicazione tra scienza e società. Nelle conclusioni, mi limito a ripercorre i passaggi principali della tesi, a riassumere le posizioni difese, e a indicare possibili sviluppi di ricerca futuri.

Contents

Introduction **5**

1. Defining Disinformation: An Attitudinal Account **11**

1.1. Bad Information Can Be Dangerous **12**

1.1.1. False Information Can Cause Severe Health Consequences 12

1.1.2. Promoting Violence 13

1.1.3. Compromising Democratic Processes 15

1.1.4. Climate Change: A Peculiar Case 16

1.2. Distinguishing Different Forms of Bad Information: An Attitude-Based Taxonomy **19**

1.3. Refuting the Attitude-Based Taxonomy **25**

1.3.1. The Shortcomings of the DIGC Account of Disinformation 27

1.3.2. No Solutions From Contextualism 28

1.3.3. Further Categorical Mistakes Caused by the DIGC Account 30

1.4. The Inescapability of Attitudes **31**

1.4.1. The "Gateway Belief Model" and the Inescapability of Attitudes 33

1.5. A Defence of the Attitudinal Account of Disinformation **36**

Conclusion of Chapter 1 ..... **41**

2. Dynamics of Disinformation Spread: Technological Development, Microtargeting, Echo Chambers **44**

2.1. Different Approaches to Disinformation: US and EU **45**

2.2. How Does Disinformation Proliferate? **49**

2.2.1. False and Misleading Memories 54

2.2.2. Illusory Truth Effect 55

2.2.3. Sense Making 58

2.2.4. Lack of Accuracy 59

2.2.5. The Web Plays by Its Own Rules 60

2.2.6. Refuting Cognitive Determinism 61

2.3. Are Echo Chambers Real? **62**

2.3.1. Opinion Segregation Is Not Inevitable 66

2.3.2. Isolation and Polarisation Are Not Necessarily Irrational 70

2.4. Don't Multiply Entities Beyond Necessity **74**

2.5. Echo Chambers are Irrational Decisional Patterns **77**

Conclusion of Chapter 2 ..... **81**

3. Disinformation Influences Social Identity Formation Processes **84**

3.1. Motivated Reasoning and Disinformation **85**

3.2. Still, Not the Real World **90**

3.3. Social Identities Are Not Fixed **93**

3.3.1. Becoming a "Real British": The Use of Identitarian Rethoric to Spread Disinformation During the Brexit Campaign 99

3.4. Are We All on the Same Page? Sharing Narratives Through Disinformation **102**

3.4.1. Far-Right Disinformation on YouTube: The US Case 104

3.5. A New Look for Disinformation Sharing: Cooperation **105**

Conclusion of Chapter 3 ..... **110**

4. The Army of the Left Behind: Social Exclusion in a Digital World **113**

4.1. Distrust in Science and its Impact in the Spread of Scientific Disinformation **114**

4.1.1. A Perceived Lack of Benevolence Triggers Distrust in Science 114

4.1.2. Who Is More Likely to Develop a Social Status-Based Distrust in Science? Minorities, Reactionary and the Socially Excluded 119

4.2. Social Exclusion and Distrust in Science **123**

4.2.1. Where Future Expectations Are Low, Disinformation Is High 126

4.3. Social Exclusion Is Higher Order Evidence **130**

Conclusion of Chapter 4 ..... **134**

5. Unexpected Drivers to Disinformation: The Shortcomings of the Communication Between Science and Society **137**

5.1. The Conceptual Evolution of Objectivity **137**

5.2. Transparency **140**

5.3. Quantification **142**

5.4. Numbers and Indicators in the Political Discourse **143**

5.4.1. From Quantification to Disinformation 147

5.5. The Epistemic Benefits of Diversity **150**

5.5.1. Group Decision Making Benefits from Diversity 150

5.5.2. Diversity Needs Conflict 153

Conclusion of Chapter 5 ..... **155**

Conclusive Remarks **158**

References **165**

Introduction

In year 1921, French historian Marc Bloch published some considerations regarding his experience in the battlefield, during World War I. In particular, Bloch focused on the spread of false news that characterized the entire conflicting period. In a few pages, Bloch made some considerations that, a century later, are still of enormous relevance and usefulness for anyone who is interested in the same topic. First, Bloch noticed that before his times, it was unconceivable to consider errors, mistakes and false news as possible objects of study and analysis. It is a modern novelty of his time to consider the errors that people make as historically, culturally, and socially significant factors. However, Bloch argued that false stories and news have always moved the crowds, conditioned people's decision making, and caused major consequences in the history of humankind, and this has been greatly true during the 1914-1918 conflict as well. A few years later, in 1921, Bloch argued that World War I could have been interpreted as an "immense experiment of social psychology of unprecedented fruitfulness" (Bloch 2014, p. 107, my translation). From the experience within the peculiar research environment that is the wartime, Bloch derived two major conclusions. First, false news, their formation and spread, as a topic of study belongs to a peculiar domain of research, which is social psychology, a newly born field at the time in which Bloch was writing. This is due to the fact that false news cannot be studied considering the single individual. Their spread is a collective phenomenon and no experiment, independently of how well designed it is, could ever replicate the "magnificent wholeness that only a long period of time and uncountable mouths can give to it." (Bloch 2014, p. 105, my translation). Moreover, and this is the second point made by Bloch, for false information to spread and radicalize in people's minds, there is only one necessary and sufficient condition, namely "finding within the society where this spreads a supportive *breeding ground*. In it, people unconsciously manifest their prejudices, hate, fears, and all of their strong emotions" (Bloch 2014, p. 106, my emphasis and translation). These words by Bloch are coherent with many results and arguments currently discussed within the literature on the spread of false and misleading information, and I argue that they also indicate the most fruitful research approach to study and understand this phenomenon.

In this work, I investigate the reasons and causes behind the spread of a peculiar form of misleading information, namely disinformation, particularly about scientific

topics and claims, which I will refer to in the rest of this work as *scientific disinformation*. In particular, I will focus on studies related to scientific disinformation about climate change, which has specific features that are highly relevant. As I will argue, clearly many things have changed since Bloch raised the reported points in 1921, particularly in terms of communication technology, but the intuition underlying them remains correct, and therefore worth of being considered in our times as well.

Indeed, today as in 1921, the proliferation of false and misleading contents is possible only because there is a "breeding ground" (Bloch 2014, my translation) that is composed of people's feelings, fears, prejudices and beliefs. The main claim that I will defend in this work is that the current spread of disinformation is first and foremost a social phenomenon, and specifically the result of both suppressed and manifested frictions and social conflicts. To the prejudices, hate, fears and strong emotions mentioned by Bloch, I add people's distrusting attitudes, biases, anger, resentment and social isolation as the main causes of the spread of disinformation. Alongside this general claim, I argue that even if the recent development of information communication technologies has surely facilitated the proliferation of disinformation, still the latter is not directly imputable to content recommendation systems or other algorithmically regulated systems. The spread of disinformation online is not an unexpected negative externality, a suboptimal component of a progressive adaptation to a new communicative environment, namely, a phenomenon independent of people's behaviour and will. It is the result of the exploitation of the new communicative possibilities offered within the digital environment by interest groups who aim at conditioning people's decision making and the public debate. Moreover, it is also the result of a condition of social polarisation and resentment that has not originated online, but that online has found a fertile soil to manifest. Throughout this work, I aim to show that the spread of disinformation is the result of complex dynamics that involve epistemic and non-epistemic factors, as well as historical and socio-political elements. As I will argue again in the conclusive remarks, my aim is not only to describe why disinformation proliferates and highlight the extent to which the spread of disinformation is a very dangerous phenomenon. I also claim that the spread of disinformation, in particular the scientific one, is a very useful social indicator, capable of revealing hidden or underexplored social dynamics. To accomplish this task, I will divide this work in the following chapters and related topics.

In Chapter 1, in the first section, I consider recent historical and harmful events that have been characterised and influenced by the presence of disinformation. I dedicate the longest subsection to the debate on climate change to show what disinformative strategies have been adopted to contrast climate science. Then, I discuss the various definitions of disinformation currently considered in literature. The main contribution of this chapter, to the overall discussion, is to argue that disinformation is a phenomenon that inherently involves both epistemic and non-epistemic factors, and that both the intentionality of the creators of disinformation and of the audience that receives it are crucial to properly define and understand it.

In Chapter 2, I present a review of the literature regarding the possible reasons behind the apparently relentless spread of disinformation. I focus on the role played by the recently developed information communication technologies and social media platforms. I show that the digital environment facilitates the proliferation of disinformation by incentivizing the spread of poorly accurate contents, and by fostering users' automatic cognitive mechanisms, including biased reasoning processes. I also consider another risk associated to the current digital exchanges, namely the progressive formation of clusters of like-minded users that eventually radicalize and polarise. In this Chapter, I argue that social polarisation, which is often considered to be a result of the current digital environment, and also a consequence of the exposure to disinformation, is actually part of the cause for the dissemination of disinformation online and originated well before the development of the current digital environment. It is only since very recently, namely since disinformation has proliferated enough to pollute the digital environment and become an endemic presence, that it is possible that certain people, who are not currently polarised or radicalized, become so as a result of the disinformation they are exposed to, online and on social media platforms.

In Chapter 3, I elaborate on the latter point, starting from a critique to the literature on disinformation, which appears too anchored in the idea that people form their group, social, and individual identities outside of the digital environment, in the "real" world, and then interact online as fully formed and crystalized individuals. I claim that, in a world that daily becomes more digital, many of the most relevant experiences that condition social identity formation processes take place into the digital environment, and sharing disinformation with others is a peculiar kind of action that is highly important and

impactful for people's social identity formation as well. In this Chapter, I start from this general critique to argue that the spread of disinformation is also highly related to mechanisms of social identity formation, in a process that includes the consolidation of narratives and worldviews through the dissemination of disinformative contents. In particular, I argue that there are people who are incentivized to cooperate with disinformation creators, and the best way to cooperate with them is to share their contents and make their narratives broadly popular. Apart from people with specific material interests, the large majority of those who endorse this incentive are people who feel unfairly treated and not respected within their social environment.

Following the lines traced in the previous chapter, in Chapter 4, I consider the possibility that the spread of scientific disinformation is imputable to a diffused form of distrust in science. In particular, I focus on the possible non-epistemic reasons to distrust science, and particularly on the possibility that various social groups believe that scientific institutions lack in benevolence. I claim that alongside other social groups who distrust science for peculiar reason, there is a particular social group composed by those who feel left behind and socially excluded that is very likely to develop a socially motivated distrust in science. In the attempt to properly define this social category, I consider different variables and empirical studies. I argue that these indicate a promising research direction and motivate further exploration and analysis. In the conclusive part of this chapter, I argue that associated to the mentioned non-epistemic reasons, there could be particular epistemic dynamics through which those who feel socially excluded reject scientific information, and that make them vulnerable to scientific disinformation.

In Chapter 5, I consider the possibility that there are some possibly unexpected factors that could have contributed to the spread of disinformation, which are related to the current characteristics of scientific communication. I consider transparency and quantification. In particular, I focus on the latter one, and on the communicative values that numerically represented scientific results have from a socio-political and socio-epistemic level. In this chapter, I argue that, in recent years, the political discourse in democratic societies has been characterised by an instrumental and uncritical exploitation of the communicative benefits of quantified scientific knowledge. This has determined major socio-political and socio-epistemic shortcomings. First, the emergence of populist forces, particularly those who gained support leveraging on anti-expert and anti-scientific

rethoric, and second, the loss of the epistemic benefits that derive from the diversity of views and perspectives that characterize plural societies. The aim of this chapter is to show that, inter alia, the spread of disinformation could have been the indirect result of unfortunate and myopic rhetorical practices within the political discourse.

In conclusion, I resume the considerations made throughout this dissertation and describe possible future developments. In particular, I argue that to contrast the spread of disinformation about climate change, and to incentivize the public support for mitigation policies, it is necessary to consider the socio-epistemic causes of the spread of anti-scientific attitudes. In particular, it is vital to have an approach grounded on the reality of things and practices, and that enables to determine the causes of the formation of the "breeding ground", to use again Bloch (2014)'s term, that is prone to accept anti-scientific claims, and able to trigger the spread of scientific disinformation. I believe that these causes and reasons are largely socio-political and socio-epistemic in nature, and that researching and studying the phenomenon of disinformation is the way to uncover them.

1. Defining Disinformation: An Attitudinal Account

In the literature, various accounts argue that a necessary condition to define content as disinformative is that those who produce it have misleading intentions. I label the accounts that include this requirement as "attitudinal". In this chapter, I defend an attitudinal account, and I argue that a propositional content does not need to be false to be disinformative by definition. In other words, I defend a conceptualization of disinformative contents as alethically neutral ones that are disseminated with misleading intentions. The element of originality I add to this account is that the definition of disinformation should also consider the attitudinal state of the audience that receives a piece of disinformation. It is indeed crucial to point out that it is an essential component of disinformation that it is never spread randomly, but with a specific target in mind, namely the audience that is more likely not only to accept disinformation, but to share it with others. It is for this reason that the requirement of alethic neutrality is meaningful, because disinformation is often produced using factually true contents that are shared in a misleading or decontextualized way with an audience that is prone to grasp the false implicatures that it contains. The chapter is divided as follows.

In the first part of this chapter, section 1.1 and related subsections, I report some cases of harmful events triggered by the influence of false and misleading information. I dedicate a specific subsection (1.1.4) to the spread of false information about climate science to point out some important peculiarities. In the second part of this chapter, starting from section 1.2, I consider different theories of information, and related systems of categorization between information, misinformation and disinformation. I will show that many authors, even with very different accounts, endorse the idea that a content is disinformative when intentionally spread to mislead. There are few accounts who reject the latter claim, including the one proposed by Simion (2023). In section 1.3, and related subsections, I describe the main characteristics of this position and present its shortcomings. I will argue that Simion (2023)'s account has problems that depend on the attempt to define disinformation only on epistemic grounds, thus independently of the attitudes of both the creators and the audience. In the conclusive section 1.5, I defend an account of disinformation based on the claim that, for a piece of information to be categorized as disinformation, its creator must have misleading intentions and therefore

target the groups who are more likely to accept and share the received contents on their own. This further requirement marks a distinction with the already common attitudinal position present in literature, and that considers misleading intentionality as a necessary and sufficient requirement for disinformation. The one that I argue for is a more specific form of intentionality, which is not only limited to misleading the audience, but misleading a specific audience which is targeted because it is the one who is more likely to endorse and share the disinformation received.

1.1. Bad Information Can Be Dangerous

In the latest edition (WEF 2024) of the Report on Global Risks, by the World Economic Forum, experts ranked misinformation and disinformation as the most severe global risks on the short-term period (2 years), above climate change, polarisation, cyber security, interstates conflicts and others. Even in a longer-term perspective (10 years), misinformation and disinformation are still present among the top five most severe global risks, arguably due to the development of artificial intelligence (AI), which raises further concerns among experts because it might be used to produce disinformation via deepfakes and other AI-generated contents (Bontridder and Poullet 2021). Experts are visibly concerned about the spread of false information, and this for the simple fact that, in recent years, it has been evident how this could cause negative effects on the social, economic, and political domain, but also on people's health and wellbeing. During the COVID-19 pandemic, the WHO (2020) argued that alongside the viral pandemic also an "infodemic" was taking place, namely a proliferation of false information and news about the virus. However, the realization that the spread of false information can be severely harmful and dangerous is antecedent to the COVID-19 pandemic. In this section, I report some recent historical episodes to show the variegate and severe effects that might be caused by the proliferation of false information.

1.1.1. False Information Can Cause Severe Health Consequences

During the recent COVID-19 pandemic, scholars have reported how the exposure to false information about the virus and related sanitary recommendations had negative consequences on those who received them. In the early weeks of the COVID-19

pandemic, exposure to false information on TV channels was positively correlated with higher numbers of infections and deaths, both related to non-compliance with health recommendations, including social distancing and mask wearing (Ash et al. 2024, Hmielowski et al. 2023, Simonov et al. 2020), which implies that harmful consequences like being exposed to the virus and getting infected had been paid not only by those influenced by false information, but also by anyone socially and physically close to them.

Another example of how the spread of false information has impacted people's health is the case of HIV denialism. Chigwedere et al. (2008) estimated that between 2000 and 2005 around 330.000 human lives were lost due to the lack of perfectly implementable treatment programs, made using antiretroviral drugs to prevent mother to child infection. During the time frame considered, the President of South Africa was Thabo Mbeki, well known for his anti-scientific positions about the correlation between AIDS development and HIV infection, who refused Global Fund's support, medicines, and to implement a system of health assistance and medicines' provision for AIDS affected patients. Grebe and Nattrass (2012) verified the existence of a correlation between trusting Tshabalala-Minang, the Ministry of Health in Mbeki's Government, and believing in anti-scientific and conspiratorial claims about AIDS and HIV, plus, they also measured the correlation between this latter variable and the probability to implement some useful health recommendation, like wearing a condom during sexual interchange. Results showed that trust in Tshabalala-Minang made it seven times more likely to believe in AIDS-related false information, and also that believing in the latter makes it two times more likely to not use a condom during sexual interchange and violate other health recommendations regarding HIV. These results show that the spread of false and misleading information can be majorly impactful on people's decision making, and cause extremely severe consequences when spread by institutional and highly trusted figures.

1.1.2. Promoting Violence

False information can also motivate and promote acts of physical violence against other people, as seen in Myanmar and Sudan. As resumed by Mozur (2018), given the impressive share of Facebook users in Myanmar, members of the military designed a communicative campaign, based on false information mostly spread on Facebook, to incentivize hate and encourage terrible actions against the Muslim community of

Rohingya, whose genocide has been going on for the last 50 years. This campaign managed to incite murders, and even after that Facebook shut some profiles down, still the campaign went on using fake profiles. Similarly, in 2019, after a violent and bloody repression of a manifestation in Khartoum, it turned out that a company called New Waves, run by former members of the Egyptian army, had designed a multi-platforms campaign to support and praise the Sudan military. New Waves paid 180 US dollars a month to anyone willing to share pre-crafted contents on their social media, including messages showing support to the brutal actions of the Sudan army (Walsh and Rashwan 2019). The latter case is important also because allows to point out two relevant elements.

First, the messages shared did not include false contents, since an endorsement has no truth value. The aim of this communicative strategy was to make it look like the vast majority of the population supported the endorsement included in the posts, which is another possible way to mislead through the dissemination of certain contents. Second, false information can be spread for reasons of profit. Monetary gain was the reason why a group of teenagers from Veles, Macedonia, created almost 100 websites where they published false information in support of the Republican candidate at the 2016 US Presidential elections, Donald Trump. It has been estimated that with this activity they earned approximately 8000 $ a month (Guess and Lyons 2020, p. 14), using clickbait as their main technique to attract attention and foster the virality and spread of their contents. This famous episode showed how achievable it is to design and spread an infodemic, and more importantly, that false information is often created by groups of interest with the intention to cause impactful socio-political consequences.

After an unfortunate episode of the recent North American history, the assault of Capitol Hill on January 2021, the spread of false information on social media was taken into account as one of the factors that have fuelled the rioters and contributed to the organization of the assault (Bizel and Singh 2023). In cases like these, were extreme actions are performed by individuals and groups, it is always complicated to determine the causal role that false information has ultimately had in people's decision making, but it is surely a fact that the spread of false information in far-right platforms of discussion, before the assault, represented a piece of evidence of the triggering ability that false contents have when they reach the proper audience.

1.1.3. Compromising Democratic Processes

False information can be used to alter democratic processes like elections and referendums. One of the most famous and recent cases is related to the 2016 US Presidential elections, which was characterized by two episodes that are worth of consideration. First, the Cambridge Analytica scandal. A group with a political intent, the election of Donald Trump[1], exploited a leak in Facebook's data protection to target specific users and send them tailor made false information to influence their vote. Second, the Russian interference during the same campaign[2], aimed at the election of the Republic candidate as well. The Russian Intelligence Research Agency (IRA), a "troll factory", as it has been called by Bastos and Farkas (2019), hired people to create false information and target specific groups, in the attempt to orientate them in the desired voting direction (Guess and Lyon 2020). There is considerable evidence of the Russian interference during these electoral campaigns, however, as shown by Benkler et al. (2018), the evidence of its actual impact on voters' decisions is more blurred and difficult to determine. Independently of the latter aspect, there are important aspects to point out of this episode. It is important to notice that the Russian-made false information primarily aimed at spreading scepticism across the audience. The idea was not to convince the North American voters of the truth of something in particular, but compromising their confidence in most of the things heard or said within the public debate. As I will show in the next subsection, this is a typical communicative strategy of the creators of false and misleading information, used also to deny climate science. To conclude this sub-section, it is worth highlighting that the episodes mentioned are interesting not only because they show how bad information can be crafted and used strategically to influence debates, platforms of public discussion, and democratic decisions, but also how the latter strategy has been professionalised. To influence the 2016 US elections and the Brexit referendum (in Chapter 3, I consider the Brexit case in greater detail), contents were manufactured and studied to achieve virality, elicit specific reactions, and their circulation was

[1] Cambridge Analytica also worked to support candidate Ted Cruz and to influence the UK electorate during the Brexit referendum in the direction of a pro-'leave' vote. However, the link with Donald Trump is arguably the strongest, since the former Vice-President of Cambridge Analytica, Steve Bannon, was also the Chief Executive Officer of Trump's 2016 presidential campaign.

[2] Reported and documented in: Report on the Investigation into Russian Interference In the 2016 Presidential Election, also known as the Mueller Report.

conceived to target specific social media users (Alcott and Gentzkow 2017). In other words, a designed disinformative campaign is a more complex mechanism than a simple dissemination of false contents online.

1.1.4. Climate Change: A Peculiar Case

One of the contexts where the potentially nefarious effects of the spread of bad information are better noticeable is the debate on climate change. Scholars argue that the spread of false and misleading information has played a prominent role in limiting the public discussion of the topic, and also the political consensus for the implementation of mitigation strategies (see Treen, Williams, O'Neill 2020 for a review). The history of the debate on climate change is also characterized by some characteristic aspects, that are worth pointing out.

Considered as a topic of political discussion and public debate, climate change presents some features that do not contribute to its popularity. In the US context for instance, climate change is very poorly discussed, to the point that it has been defined a "background noise" in the US society (Norgaard 2011). There are various factors contributing to this. Arguably, one of the most impactful is that climate change lacks in salience (Leiserowitz 2007) since it is easily perceived as the longest-term issue discussed in science and politics. However incorrect, since many of the effects of climate change are already visible, the latter perception has a great communicative force, and it makes very complicated the development of a prolonged and constant public discussion on this topic (Norgaard 2011, Chap. 6). In the vicinity of extreme climate events, people manifest a higher interest in the issue, with those directly invested showing a spike in both climate change belief and interest (Sisco, Bosetti, Weber 2017). After the disastrous effects of Hurricane Katrina, in 2005, the climate crisis became one of the major topics of discussion in the American society. However, this positive trend did not last or turn into a stable debate. Actually, during and after the passage of Hurricanes Helene and Milton in 2024, the US based debate on the topic was useful only to highlight the increasing levels of polarisation regarding climate issues and incentivized no major changes in the public debate. On the contrary, even during the conclusive days of the Presidential campaign, immediately after the passage of the Hurricanes, climate change was severely ignored by the public opinion (Jacobo 2024). Communicatively, a topic that is

characterized by an almost inherent lack of salience can be easily set aside as irrelevant, and the long-term positive effects of a mitigation strategy that is currently costly, can be also fall prey of arguments conceived to leverage on people's cognitive propensity to focus on short term utility (Weber and Johnson 2015). These features make climate change, and related science, vulnerable from a communicative point of view.

Climate change is also one of the most politicized scientific topics. From the 60s and the 70s, in the US, conservative organizations and societies developed as a form of cultural response to the progressivism and related movements that were spreading all around the country. Climate change became an emblematic topic, the symbol of the path that the country was taking in those years, and that these emerging conservative community aimed at contrasting (Dunlap and McCright 2010). This process resulted in what has been called the "climate change countermovement" (CCCM, Brulle 2014), namely the development of a set of think thanks, conservative media, and later Web and social media figures. Part of the communication strategy of these figures was to produce false and misleading news about climate change and related science that, with time, have reached not only American citizens and digital media users but have spread all around the world. With the financial support of interest groups, the mentioned figures, that compose the CCCM, have created an information network that produced and shared large numbers of anti-scientific contents.

One example is the Global Climate Coalition, established in 1989, financed by the American Petroleum Industry and the Exxon Foundation, and that became a producer of false information about climate science to contrast pro environmental initiatives including, inter alia, the American commitment to the Kyoto Protocol, in 1997 (Dunlap and Brulle 2020, Dunlap and McCright 2010, Pooley 2010). With time, this countermovement has been joined by politicians, and particularly by progressively larger shares of the Republican Party. Starting from the 1980s with President Reagan, until the epitome of Donald Trump's Presidency, climate change has become one of the most polarizing issues across the US political spectrum, with recent data reporting that only 12 % of Republicans believe that the mitigation of climate change should be a priority for the country (Kennedy and Tyson 2024). What is worth pointing out of this dynamic of events is that the one implemented by interest groups, with the various channels of information mentioned, is a strategically conceived system of false information sharing.

Not only various sources share the same misleading contents, but these are part of a communicative structure that daily produces contents in the attempt to influence the public debate and opinion formation. The principle behind this process is analogous to the cases previously mentioned. The use of false information to influence an election or referendum, independently of the efficiency of this attempt, is analogous to the use aimed at influencing the public opinion on a social issue. However, there is an important difference. With some exceptions (see for instance Harvey et al. 2018, who mention some cases of plain and blatantly false claims about climate science), the intention behind the spread of false information about climate change has often not been the one to convince the audience of an alternative truth to the one proposed by climate science. Instead, the funders and actors of the false information machine aimed at persuading the public opinion that anthropogenic climate change was an unsettled scientific issue, namely one that was largely debated within the scientific community (Oreskes and Conway 2010).

This is notoriously not the case. The scientific community is indeed very compact when claiming that anthropogenic climate change is taking place and that mitigating interventions are urgent (Cook et al. 2016). Moreover, within the very small component of scientists who do not believe in climate change, there is a very small share of experts in climate science (3%, Young and Fitz 2021). However, it is very difficult to make these levels of scientific consensus widely known. In recent years, there has been an increase in the share of US citizens who are reportedly aware of the scientific consensus on the occurrence of climate change (to see this in detail, consider, in the following order, the results reported by: Rosentiel 2009, Leiserowitz et al. 2014, Ballew et al. 2024), but this share remains low (56% of US citizens), and even in EU, where this share is higher on average than in the US, data showed that the awareness on scientific consensus about the anthropogenic nature of climate change struggles to settle (Lübke 2022).

Another communicative strategy used by the spreaders of false information about climate change, which is even more problematic in its effects, is questioning the methodologies used by climate science, and even more seriously, the motivations and interests of climate scientists. In the attempt to shed some doubt about the veracity of the reported data, what is questioned is not the competence but the integrity and the commitment to objectivity of the scientists, in particular of those who collaborate with international and governmental institutions, like the IPCC (McCright and Dunlap 2010,

Dunlap and McCright 2015, Hess 2014). The producers of false information aim not only at discrediting the integrity and deontology of experts, but more specifically, at convincing that those in positions of power, as the members of scientific institutions are claimed to be, are inevitably corrupt and hence unreliable. The latter is a very common communicative strategy, which has been used also in the political discourse, by populist and anti-intellectual forces, which have contributed to compromise the relation of trust between the scientific institutions and part of the civic society (I will elaborate more on these aspects in Chapters 4 and 5).

1.2. Distinguishing Different Forms of Bad Information: An Attitude-Based Taxonomy

In the previous section, and related subsections, I have used "false information" and "bad information" as interchangeable umbrella terms to mention instances of dissemination of false and misleading claims. In this section, I make some terminological order among the various forms of information, including disinformation, the one I will focus on in the rest of this work.

Currently, the most commonly used taxonomy of the different forms of information includes misinformation, disinformation and malinformation. To determine to which category a propositional content belongs, it is necessary to consider two variables: the truth value of the content, and the intentionality of the creator of the content. In this chapter, the figure of the creators of disinformation will always be accompanied to the one of the sharers of disinformation, but it is important to clarify that the two are not always overlapping or equivalent. The context of diffusion and exchange of information that I will consider the most is the digital one. The digital context is composed by social media platforms and Web pages where it is possible to share contents with other users. In this peculiar environment, as Floridi (2014) has noticed, it is no longer possible to sharply distinguish between producers and consumers of contents, as it is the case with respect to traditional media, where it is always possible to determine who is the producer of a newspaper article for instance, and who is the consumer of it, namely the reader. In the digital environment, the line between these two figures is often much more blurred and sometimes almost indeterminable. In this particular communicative context, it would

be inaccurate to consider only the intentionality of the individual who materially designs and produces an item of information, namely a social media post, reel, video, image, or news article. I claim that whoever shares a content while having the same amount of knowledge and awareness about the reliability of that content of its creator should be considered a creator herself, namely, her intentionality should be considered when determining to what kind of information the shared content belongs to. In other words, considering the previously mentioned case of the IRA, the group that created and spread false information during the 2016 US Presidential campaign (see 1.1.3) the intentionality of those who materially produced the contents to share, and that of those whose duty was limited to sharing those contents on social media, matter to the same extent, at least for purposes related to the definition and categorization of different forms of information. Similarly, another variable that requires a clarification is intentionality. In this chapter I will use "intentionality" and "attitude" interchangeably. With both terms I simply refer to the goals and reasons that a certain agent has for sharing a content.

Now that I have clarified the meanings of the most relevant terms involved, it is possible to partition the various forms of information. The taxonomy is divided as follows. If a content is false and shared without the intention to deceive the audience, this is an item of misinformation. If a content is false and shared with the intention to deceive the audience, this is an item of disinformation. If a content is true but intentionally shared in a way that could be deceitful and conducive to the endorsement of false beliefs, this is an item of malinformation (see Table 1).

Table 1	True Content	Intention to Mislead
Misinformation		
Disinformation		
Malinformation		

With this categorisation, it would already be possible to explore most of the literature. However, it is possible to enrich this picture by considering the possible varieties that have been proposed of this distinction, and in particular the different conceptualisations of the notion of disinformation that have been offered, and that depend on the different positions defended in the debate in philosophy of information.

A useful guide through some different conceptualisations of disinformation has been offered by Fallis (2014). In his analysis, it is considered the pragmatic approach, in the theory of information, offered by Fetzer (2004) who interpreted "information" as an alethically neutral term used to denote propositional contents conceived to represent a certain state of things. From this perspective, the only necessary requirement for a content to be definable as an item of information is the speaker's intentionality. If a speaker uses that content to communicate, to let others know something, this is an item of information. From such a radical pragmatic standpoint, misinformation and disinformation are peculiar forms of information, with the latter being misleading, and intentionally communicated to mislead, while the former is only misleading in its communicative effects, but with no intentionality involved.

Dretske (1983; 2008) defended a completely different theory of information and endorses a "semantic" approach. Dretske argued that misinformation and disinformation are not and cannot be peculiar kinds or sub-species of information, since the latter can

only be constituted by contingently true propositional contents, whose informative nature is independent from the intentions of the speakers. Misinformation, disinformation and false information in general, from Dretske (1983)'s perspective, are kinds of information in the same way "a decoy duck is a kind of duck" (p. 57). In his work, Fallis (2014) considered another position which is coherent with the semantic approach of Dretske. In particular, Fallis focused on the evolution that the ideas of misinformation and disinformation have had in Floridi's scientific production. At first, Floridi used the term "disinformation" only as a way to denote a form of defective information (1996), an idea that later changed as the more general theory of information that lies behind evolved into a strongly semantically one. In Floridi (2007; 2013)'s works, information is defined as contingently true, well-formed and meaningful data, which implies that an item of information cannot be false, and consequently misinformation and disinformation cannot be defined as peculiar forms of information but must be something else. More precisely, given a semantic content, information is the factually and contingently true component of the content, whereas misinformation is the false one, and when the speaker is aware of the misinformative nature of a semantic content, and communicates it nonetheless, as if it were a piece of information, this content turns into an instance of disinformation.

Fallis (2014)'s critique to this idea is twofold. First, Floridi's definition of disinformation leads to the categorization of implausible lies as items of disinformation, which are not, according to Fallis, because, since they are implausible, they are also very unlikely to mislead the audience. Second, Floridi (2011)'s formulation is said to be too narrow because it cannot include what Fallis called "true disinformation", even if it is not very clear what Fallis meant with this term. To clarify his point, Fallis made the example of politicians who, to convey a misleading message, highlight only certain aspects that considered in isolation are factually true. Therefore, from Fallis' perspective, "true disinformation" is a form of instrumental and misleading use of true contents to convey false implicatures, or something closer to cherry picking. In any case, it seems possible to argue that this blurred notion of "true disinformation", used by Fallis, is equivalent to that of malinformation. The way to solve the detected problems, particularly in Floridi's account of disinformation, Fallis argued, is to endorse a "functionalist" account of disinformation, which states that disinformation is misleading information *with the function* of misleading the audience (Fallis 2014, p. 625; see also Fallis 2016). As seen

from this brief analysis, the different approaches and definitions of disinformation depend on the intentionality of the speakers, whether this is constitutive and necessary for information, as in the case of a pragmatic approach, like the one defended by Fetzer, or whether this is an external factor that only clarifies the conditions of proliferation and use, which remain nonetheless useless to determine whether a propositional content is a piece of information or not, as it is the case from a semantic perspective. Moreover, from the perspective of the semantic theories of information, false information is a contradiction, thus if falsity is a condition for misinformation and disinformation, the latter are not peculiar kinds of information, whereas from the functionalist or pragmatic perspective, the concepts of false information and true disinformation are not problematic or inherently contradictory, and hence usable.

Given the evident variety of perspectives present in this debate, in recent years there have been scholars who have tried to make some order within this literature and develop a clearer definition of misinformation and disinformation that is independent of the endorsement of any peculiar theory of information. Baines and Elliott (2020), for instance, have used Hempel (1952)'s theory of concept formation to define a scientifically accurate taxonomy which included four different categories of "representational data", the primitive concept[3] of this categorization, which included: information, misinformation, disinformation and malinformation. To navigate through this taxonomy, it is necessary to know whether speaker's intentionality is deceitful, and whether truth equivalence holds between the shared message and the propositional content included within. More recently, Søe (2021) has defined a "unified" account of information, misinformation and disinformation which is grounded on the Gricean distinction between natural and non-natural meaning, and the related Scarantino and Piccinini (2010)'s idea of natural and non-natural information. An item of information is natural when its propositional content is related to the physical world, hence its truth value is independent of speakers attitudes and communicative intentions, and dependent only on the current state of the world. Conversely, a piece of information is non-natural when its truth value is speakers' dependent, meaning that it is related to social, linguistic and conversational conventions, in the broadest sense. Due to speakers' dependency, it is only this latter category that can be internally divided in different categories of information on the basis

[3] In Hempel (1952)'s sense.

of speakers' intentions and misleading efficacy. More specifically, Søe (2021) argued that a propositional content is informative when intentionally not misleading, misinformative when characterised by unintended misleadingness, and disinformative when intentionally misleading. Interestingly, this taxonomy is different from the ones previously considered because the categorization regards contents that are alethically neutral, which means that truth value is not a decisive factor for internal differentiation, something that allows to effortlessly label as instances of disinformation all those of malinformation, which is something that Fallis considered preferable as well, but that Søe (2021) managed to do without appealing to the problematic notion of "true disinformation".

Moreover, I argue that a reductionist approach that gets rid of the category of malinformation, and that lists its cases as instances of disinformation, is more coherent with the results currently available in the literature. In fact, there is evidence showing that the vast majority of misinformation and disinformation spread on social media, on the Web, and on other communication platforms is not constituted by plainly false contents but by misleading, partisan and manipulative ones (Roozenbeek and Van der Linden 2024, Chap. 4). Empirical evidence supporting the latter view comes, for instance, from Brennen et al. (2020)'s study on information during the COVID-19 pandemic. In this work, the authors considered a sample of detected items of misinformation regarding the virus, the pandemic, Governments' reactions and measures, and other related topics. Results show that 59% of their total sample consisted of misinformation and disinformation made using true contents and information which was twisted, distorted, partially reported, as well as de-contextualised or re-contextualised, to convey a misleading content. More specifically, 87% of the misinformation and disinformation shared within interactions on social media was of the latter kind, with a limited 12% of the total misinformation and disinformation on social media being completely false, fully designed and fabricated (Brennen et al. 2020). However, it should come with no surprise that the vast majority of misinformation and disinformation includes true contents and is the result of a distortion of factual truths and similar attempts. Plainly implausible or false statements are much less likely to be shared by mistake or negligence, hence without intentionality, as in the case of misinformation, because the speaker is in the epistemic position to recognize the implausible or problematic content of the reported information. It can surely happen sometimes, but it is sound to imagine that it occurs in the minority

of the cases. Similarly, in the case of disinformation, it is less likely to deceive an audience with a totally made-up content, while it is clever to distort truth, present facts in a de-contextualised way, and other strategies conceived to deceive the audience with a much more plausible and apparently reliable content. I argue that it is not necessary to extend the taxonomy with the inclusion of the category of malinformation, since within misinformation and disinformation it is possible to list also the cases of misleading uses of factually true contents, seen with respect to Brennen et al. (2020)'s work, and because these belong to the kind of phenomena that the ideas of misinformation and disinformation are conceived to grasp, and given the theorical effort made by Søe (2021), it is possible to safely do so.

To conclude this section, it is worth noticing that even in presence of important theoretical differences, the idea that disinformation is, at least partially, dependent on the speaker's intentionality has been widely accepted across the literature, even among authors who endorse very different theories of information. For this reason, I have labelled the taxonomy between the different forms of information "attitudinal", which simply indicates that to determine to what category an item of information belongs to it is necessary to systematically consider the intentions that the creator, speaker, and sharer has at the moment of proliferation. However, there are accounts who refuse the latter perspective, and in the upcoming section I discuss one of them. I believe that by considering the account defended by Simion (2023) in particular, it would be possible to make a further step in the conceptual analysis of disinformation. In particular, it will be possible to notice that the definition of a propositional content as an item of disinformation cannot be independent from considerations regarding the attitudes of both the creator and recipient of the item.

1.3. Refuting the Attitude-Based Taxonomy

According to Simion (2023), what determine whether a propositional content is an item of information, misinformation or disinformation are purely epistemic factors, totally independent from the attitudes or intentions of the creator. Simion offers a knowledge-based account of disinformation that is grounded on the refutation of the following premises, claimed to be common in literature. First, disinformation is a kind of information. Second, disinformation is a particular kind of misinformation. Third,

disinformation is always intentionally spread by someone. With these assumptions being refuted, the idea that grounds Simion's proposal is that what makes a propositional content a disinformative one is its epistemic effect on the context of proliferation, and particularly on the audience that received it. From Simion's perspective, if a content has the disposition to make the audience who receives it more ignorant than before, then this is rightly defined as disinformation. More specifically, Simion's formulation is the following:

"Disinformation as ignorance generating content (DIGC): X is disinformation in a context C iff X is a content unit communicated at C that has a disposition to generate or increase ignorance at C in normal conditions." (Simion 2023, p. 6).

To make this position clearer, Simion introduces a further requirement that makes more explicit the kind of epistemic effect that is crucial for her proposal. Simion's idea is that disinformation's peculiar effect is to induce an epistemic reaction that makes someone believing less in something true. In other words, a content *c* is a piece of disinformation if before that the audience had received *c*, its members were more inclined to believe in the truth of a true content, *p*, than they were after receiving it. This passage can be formalised using conditional probabilities, and that is how Simion represents it. A content *c* is disinformative for an audience A, with respect to a certain proposition *p*, if and only if A's subjective probability $Pr(p|c) < Pr(p)$, given that *p* is true. Behind this proposal there is a conception of evidence as "probability raisers that A is in a position to know", and an idea of what it means for an agent to be in a position to know which implies that individuals S are in a position to know a fact F if they have a "cognitive process with the function of generating knowledge that can easily uptake F in cognisers of S's type." (Simion 2023, p. 8)[4]. It is possible to notice how this proposal is coherent with the mentioned refusal of the idea that disinformation is inextricably linked to deceiving intentions. According to Simion it is possible for someone to communicate something that has a disinformative effect, without having any intention to deceive the audience. I believe that this account presents a series of shortcomings that should be pointed out. In the next sub-sections, I present all of them, considering examples from the literature on

[4] These two premises will not be discussed in the rest of this work.

scientific disinformation, and more specifically, I will focus on the results collected about the communication of climate science, and vaccines' safety.

1.3.1. The Shortcomings of the DIGC Account of Disinformation

As seen, according to Simion, for *c* to be a disinformative content it needs to be such that: Pr (*p*|*c*) < Pr (*p*), given that *p* is (objectively) true. Several problems arise from the empirical application of this requirement.

Arguably, the most problematic effect is that the DIGC account determines the definition of certain propositions as instances of disinformation even when those are exactly the opposite of that, namely well-supported scientific facts. In a study by Nyhan et al. (2014) it is reported that after receiving scientific information about the absence of a causal link between vaccination and autism, some parents empirically manifested a decline of an already low willingness to vaccinate their children. Even if the scientific information was successful in limiting the belief in the existence of a causal link between vaccination and autism, it had the described effect on the beliefs about the general safety of the vaccine, followed by the expressed intent to not vaccinate the children. In the same study, different strategies were also tested to see which one could influence parents' disposition to vaccinate their children. In particular, the scholars tested whether exposing the parents to the dramatic story of an infant suffering from the disease that the vaccine prevents, or to the images of a sick infant suffering from the same disease, could have played any effect. In both cases, the effect was negative, meaning that the parents were even more sceptical about the safety of the vaccine afterwards, and consequently even less prone to vaccinate their children than before. More recently, Pluviano et al. (2017) tested three different strategies as well, including the exposure to images of sick patients, and reported similar results, showing a double backfire effect, namely increased willingness to not vaccinate the children, and a weak effect on participants' sceptical priors on vaccination. Analogous results were detected by Betsch and Sachse (2013), who reported an increase in beliefs about vaccines' dangers, within their study population, following the exposure to messages denying the risks implied by the vaccination. An analysis of these results through Simion's account would offer the following picture. Define *p* as: "vaccines are safe", a proposition that Simion indicates as a possibly good example of *p* (Simion 2023, p. 9) and define *c* as an information about the absence of

major risks related to a specific vaccination. It seems that Pr($p|c$) < Pr (p) takes place here, and therefore c must be defined as a disinformative content, which is clearly problematic. The same would happen for propositions s and q, where the former instantiates the content of the reported dramatic story that Nyhan and colleagues used in their study, and the latter instantiates the content of the images of the sick infant, which were real and true, and nonetheless had the same effect on the audience's subjective probability over p. It seems at least counterintuitive to define an item of scientific information, whose spread is sometimes conceived to contrast disinformation, as being disinformation itself.

1.3.2. No Solutions From Contextualism

One possible reply to the latter critique could be based on the explicit contextualist nature of Simion's account, which allows to define a certain piece of communicative content as disinformative in a context but not in another, due to the fact that it makes only one of the two groups more ignorant after its reception. In particular, Simion (2023) claims that contexts can differ due to: "[…] the evidential backgrounds of the audience members, the shared presuppositions, extant social relations, and social norms." (p. 6). In the examples proposed, both the groups who intensified their disbelief towards vaccines' safety were already sceptical before the study even took place, meaning that, within the group of the sceptics, the shared presuppositions included that vaccines are not safe. Therefore, this specific context is one of a kind, an in particular one where the contents used in the mentioned study have a disinformative effect. I argue that the latter is not a good argumentative strategy for the following reason. Even allowing for the contextual argument to play a role in this case, still the DIGC account leads to a contradiction, namely that propositions 'c' and 'not c' (henceforth, '$\neg\ c$') are at same time both disinformative in a certain context, meaning that they can have the same effect over a true proposition p. Being c a claim of the kind: "there is no scientifically supported link between vaccines and autism", and p once again "vaccines are safe", it would be possible to have this combination of results Pr($p|c$) < Pr (p) and also Pr($p|\neg\ c$) < Pr (p), since there are no reasons to suppose that someone who is already sceptical about vaccines' safety will be inclined to decrease her belief in p once aware of $\neg\ c$. This result, that a communicative content and its negation both have the ignorance generative effect

required by Simion, and hence that both instantiate an item of disinformation, is unpalatable. Nonetheless, a supporter of the DIGC account could bite the bullet and claim that a propositional content, even if true and scientifically supported, could be disinformation for an audience due to the peculiarities of the context of proliferation, therefore stressing the contextual nature of the account. I believe that the latter is not a good strategy, and for the practical reason that a useful definition of disinformation needs to aim at objectivity, namely being independent from representational factors. This does not deny at all the possibility to consider the context, something that is more than meaningful to do, but it does require to not fall into relativism, which would make the distinction between information and disinformation pointless, and the DIGC account unfalsifiable. Any non-trivial content could turn out to be disinformative by considering the contextual variable. For any detectable backfire effect, the context could be represented as one which is such that a communicative content of a certain kind is disinformation in that context due to its epistemic effect, whereas in another the very opposite content is disinformative as well. I claim that the idea of disinformation as ignorance-generating is meaningful and worth exploring, but in this elaboration it might lead to too undesirable consequences, one very practical above all, namely providing an argument for the creators and spreaders of disinformation, who could argue that their contents are not disinformative for everybody, claiming that their target groups are only those for which this is not the case. Moreover, by considering certain specific contexts, the unfortunate results that follow from the contextualist nature of DIGC fully emerge. Consider groups that are bounded together by background credence, shared assumptions, and by the respect of some descriptive and injunctive social norms. Within the latter groups there are also peculiar ones, characterized by the endorsement of extreme ideas and worldviews, which can often acquire the form of echo chambers. In Chapter 2, I dedicate considerable space to an analysis of the structural and epistemic features of echo chambers. For the moment, consider echo chambers as groups of likeminded individuals, strongly committed to a set of common beliefs, and characterized by the propensity to not only systematically reject every information that comes from outgroups, but also to reinforce the beliefs endorsed within the group after this rejection (this epistemic mechanism will be largely discussed in sections 2.4 and 2.5). If we consider the case of an echo chamber within which anti-scientific ideas of some kind are spread and strongly

endorsed, an unfortunate conclusion for the DIGC's account follows straightforwardly. Offering counterevidence to groups of this kind will not only turn inefficient, but will make the members radicalizing their credence, and so reducing their belief in a true scientific statement, something like the previously mentioned '*p*'. Considering this latter point and the context of an antiscientific echo chamber, DIGC results respected when '*c*' is a scientific datum or claim that goes against one of the main claims endorsed within the antiscientific echo chamber. Once again, the result is that scientific and reliable information is labelled as disinformation by Simion's account, and in the case of anti-scientific echo chambers, this undesirable effect seems to be the result of DIGC's unbounded contextual relativism.

1.3.3. Further Categorical Mistakes Caused by the DIGC Account

There is another possible perspective from which to consider the DIGC account, and that determines inefficiencies of the described kind. Part of the literature on disinformation has started focusing on the phenomenon of the motivated rejection of science. This research approach studies the possible predictors for the rejection of specific scientific arguments or results, due to their incompatibility with people's personal positions, political ideas, and worldviews. From the results collected thus far, it is possible to claim that the rejection of science is content dependent, namely that people tend to reject those scientific contents, theories or discoveries that clash with their worldviews and deeper beliefs. It is true that there are predictors of generalized anti-scientific attitudes, like having a conspiratorial worldview for instance (for a precise analysis on this point see Lewandowsky et al. 2013), but in general, and coherently with the motivated reasoning literature, if a scientific claim is incompatible with someone's priors and worldviews it is very likely to be rejected. Results have shown that, for instance, supporters of the free market and conservatives are more likely to reject climate science and be sceptical about climate change. Arguably, this effect is due to the incompatibility between the consequences that its acceptance and attempted mitigation would have on the economy and on individual freedom (Hornsey et al. 2016, Van der Linden, et al. 2017, Lewandowsky and Oberauer 2016; 2021).

At a first sight, the mentioned empirical results might seem coherent with the contextualist nature of DIGC. Given different groups and contexts, different dynamics

occur, and so it is meaningful to think that different communicative contents could have different effects, with only some of them resulting disinformative, and only in specific contexts. However, with a closer look, it turns out that these results are of no actual support for the DIGC's perspective.

Once again, a proper thing to do is to see what happens when the DIGC's formula for disinformation detection is applied to cases like the mentioned ones. Consider the proposition: "anthropogenic climate change is occurring", another proper example of a true proposition of *p*-kind. In an environment where a strong nationalist ideology is spread and strongly rooted, people are reportedly inclined to be sceptical about a proposition like *p*. Consider *j* to be an agent who is a member of this environment and has a nationalist attitude, while *c* a communicative content that elicits *j*'s latent nationalist attitude to the point that this has an effect, indirect, but still real, on *j*'s beliefs about climate change. Even in this case, DIGC's requirement that $\Pr(p|c) < \Pr(p)$ for *c* to be defined as an instance of disinformation, is respected. However, this would be the case even if *c* was not related at all to climate change. Something like a cheering nationalist statement, which by itself seems to have nothing disinformative, or akin common examples of disinformation, turns out to be a piece of disinformation within the described context. This does not seem a good outcome for the trivial reason that it is perfectly possible to be a nationalist, and express this position with as many claims of the *c*-kind as one likes, without being forced, in any epistemic sense, to defend a denialist or sceptical position regarding climate change, or to automatically become more ignorant with respect to that topic every time that one's nationalist position is defended or strengthened. Otherwise, there would be an uncountable set of *c*-kind contents which all turn to be definable as disinformative according to Simion's account.

1.4. The Inescapability of Attitudes

One of the main premises of the DIGC account is that whether a content *c* is an item of disinformation is an independent matter of the attitudes of the speaker in the moment of the creation and sharing of *c*. This premise is coherent with the formalisation of the account, and the technical instrument used to represent the audience's changes of belief states, namely conditional probability. The problem is that having only conditional

probability in one's toolkit limits the descriptive potential of a model, since this can describe only the variations of an agent's state of belief that regard the propositional contents for which priors are known (Pr (*p*)). However, when trying to understand what lies beneath instances of polarisation, it is necessary to be able to take also people's attitudes into account, not only their priors, since the former play a crucial role in the elaboration of scientific contents. In other words, having only conditional probability at disposal, it is not possible to assess the reasons behind different beliefs' updates, within a group, following the exposure to the same piece of information ('*c'*), as in the mentioned examples (Nyhan et al. 2014; Pluviano et al. 2017; Betsch and Sachse 2013). As argued by Mahtani (2020), in cases in which it is necessary to understand why people manifest different reactions to the same piece of information, even given equal initial states of awareness, it is necessary to refer to both the conscious and unconscious attitudes that an agent holds in order to fully grasp what is going on, and something different from conditional probability is needed to take them into account. The descriptive limits of conditional probability do not imply that it is not possible to consider attitudes while preserving a Bayesian approach, as the one envisaged in Simion (2023)'s model of disinformation. An example of this possibility had been offered by Lewandowsky and Cook (2016).

In presence of polarising tendences, with two separate groups of agents reacting in opposite ways to the reception of the same piece of information, Lewandowsky and Cook (2016) managed to find a way to individuate and point out the factors that have determined the detected polarisation, namely by using Bayesian networks. In the context of a study of the determinants for the acceptance of claims about climate change, Lewandowsky and Cook (2016) found that people's worldviews were determinant for their differences in beliefs' update following the exposure to the same content. Specifically, to be a free market's supporter turned out to be negatively correlated with the acceptance of the occurrence of anthropogenic climate change, a result that has already been detected in literature (Lewandowsky and Oberauer 2016; 2021), as previously mentioned. In Lewandowsky and Cook (2016)'s study, the possibility to include the audience's attitudes and worldviews determined a descriptive improvement. Considering attitudes and worldviews is indeed crucial to understand people's different

reactions, following the exposure to the same scientific content, as also shown by the results reported in the following section.

1.4.1. The "Gateway Belief Model" and the Inescapability of Attitudes

In recent years, some scholars have argued that scientific disinformation has proliferated and has been accepted because of people's low scientific literacy and limited awareness about the current scientific debates. Considering climate science, as seen in 1.1.4, most of the efforts made to alter people's opinion formation about climate change are conceived to make the scientific debate look more internally divided than it actually is (Cook et al. 2016), which is an attempt that would be inefficient in presence of a well-informed audience. Therefore, some scholars have argued that communicating the existing levels of expert consensus on various scientific issues, including climate change, would have majorly beneficial effects on people's state of credence, including rendering them almost immune to scientific disinformation. The scholars who have proposed and tested this idea argued that the acquired awareness about experts' consensus would work as a "gateway belief" (henceforth I will denote this model as GBM). The "gateway" metaphor is used to indicate that a first order revision of beliefs regarding the current levels of scientific consensus on a topic would lead to a series of indirectly caused revisions of other beliefs about the same topic, including the related policy issues. With respect to climate change, the hypothesis endorsed by the GBM is that showing a simple message that reports the level of scientific consensus about the occurrence of anthropogenic climate change could have positive effects on people's belief regarding the occurrence of climate change, about its anthropogenic nature, and consequently incentivize the support for pro-environmental policies.

Van der Linden et al. (2019) reported results that support this hypothesis. More precisely, those who received the following message: "97 % of climate scientists have concluded that human-caused global warming is happening" reflected a very significant and positive effect[5] on all the following variables: perceived scientific consensus (a measure for one's personal knowledge of the current level of scientific consensus about the occurrence of climate change), belief in global warming, in its human causation,

[5] p value < 0.001.

apprehension for the occurrence of climate change, support for action against global warming. The strongest positive effect recorded was on people's perceived scientific consensus about climate change, which is probably unsurprising. This result was followed by the others in the reported order, which reflects the relative strength in terms of statistical effect. Very interestingly, conservative participants showed a stronger effect than liberals for perceived scientific consensus[6]. As anticipated, the authors of the study argued that the reportedly entrenched cognitive and credential changes were related to a mechanism of the following kind. At first, the message caused a major effect on perceived scientific consensus, which then caused a related second order effect on people's personal beliefs about climate change and related topics, like its anthropogenic nature, which then caused a third order effect on people's position about the urgency and aptness of policy interventions that closed the chain of beliefs' update.

Empirical results that support the idea behind the GBM have also been collected with respect to vaccines' hesitancy. Fisher et al. (2023) tested the efficiency of different messages in reducing US citizens' hesitancy to get vaccinated against COVID-19. In this study, the 44% of the total participants self-declared to be hesitant to get vaccinated, with the majority being sure of their desire to not get vaccinated ("no" group), and others being less certain ("not sure" group). Fisher and colleagues used different messages and discovered that among the "not sure" group, the most efficient message was one reporting a descriptive norm that stated: "millions already did that", whereas for the "no" group the most efficient message was a morally charged message saying: "protect others". The latter result is of particular interest since for a person who, at first, is sceptical about vaccines' safety, to end up believing that by getting vaccinated it is also possible to protect others from getting sick, it means that the belief about vaccines' efficiency has probably changed, otherwise the person would have an inconsistent state of mind. This suggests that something similar to the process of subordinate belief-changings seen with respect to climate consensus occurs for vaccines as well. Analogous results have been collected by Ronzani et al. (2022) who reported a positive effect of scientific messages on vaccination

[6] One possible explanation for this could be related to selective exposure, namely to the fact that conservatives are more likely to inhabit epistemic environments where this kind of information does not proliferate much, whereas liberals, on average, are more exposed to climate science.

intentions and perceived vaccines safety. Interestingly, this is the case independently of the different levels of trust in scientists endorsed among the participants.

Therefore, the GBM has pointed out an important effect, and Van der Linden et al. (2019) based the following claims on it. On the one hand, it is easier to change misperceptions than worldviews. It is neither economic nor efficient to contrast people's worldviews and deeper beliefs when these are in open contrast with science. The best strategy is to simply communicate, in the clearest way, the state of art and the level of scientific consensus on a certain topic. Once this latter content is accepted, then everything follows accordingly. On the other hand, communicating scientific facts does not necessarily cause polarisation. Moreover, even if a negative correlation between belief in climate change and conservativism is detectable, Van der Linden et al. (2017) reported that this was cancelled out by the exposure to scientific consensus, with no reported backfire effect. Lastly, and most importantly, the reported consensus effect is achievable only as long as the communicated messages remain politically neutral. It is crucial to keep politics out of scientific communication, and not let any scientific message or communication being represented as inherently partisan. The aptness of the latter recommendation results as evident in the work of Bolsen and Druckman (2018), where the positive effect of a communication about the scientific consensus on climate change was cancelled out by the politicization of the topic. In their study, Bolsen and Druckman (2018, p. 395) reported that, not only politicizing scientific issues determined a loss of the positive effect on perceived scientific consensus, but also that, once the positive effect of the information on scientific consensus was cancelled out by the effect of political views, nothing worked to restore it afterwards. Thus, the politicization of scientific issues seems irreversible in its epistemic effects. Even more challenging from the GBM's perspectives are the results offered by Aklin and Urpelainen (2014). In this study, information about considerable scientific consensus had very limited effects on people's support of environmental policies, whereas messages reporting even limited levels of disagreement among scientists were predictors of a reduction in both credence in climate change and demand for political interventions to contrast it. Interestingly, Aklin and Urpelainen (2014) reported that the different levels of trust in scientists, detectable among the participants of their study, predicted polarisation, namely opposite beliefs' update following the exposure to the same message. Among participants with high levels of

distrust in scientists, the higher the reported scientific agreement on an issue was, the more they became sceptical and manifested a decreased support for related policy regulation.

Strangely, the latter results support and contrast at the same time the GBM. On one hand, in Aklin and Urpelainen (2014)'s work, messages of scientific consensus had no analogous effects to the ones reported by Van der Linden et al. (2019), while on the other hand, information of scientific consensus had no generalized positive effects on people's support for environmental intervention, on the contrary, given the presence of distrusters in science, polarisation within the research group inevitably occurred. It is possible that, since in Aklin and Urpelainen (2014)'s work policy support is directly questioned, the political dimension is elicited by the question itself, with the different perspectives and attitudes endorsed within the population emerging and causing a division between different groups, as envisaged by Van der Linden et al. (2017). The latter are empirical questions that need to be directly tested to be answered, but the underlying point remains evident and solid. In presence of polarising tendences within a group, to understand what factors might have played a role in causing them, it is necessary to consider people's attitudes and perspectives, like their trust in science, political tendences, and worldviews, since the latter are also able to revert any possible positive effect caused by scientific information and messages about the scientific consensus on a topic, as those used by Van der Linden et al. (2019).

1.5. A Defence of the Attitudinal Account of Disinformation

Instances of polarisation, or simply of variegate epistemic reactions within a group, following the exposure to the same piece of scientific information, are common and natural. Not everyone updates her state of belief in the same way after receiving pieces of information regarding climate issues or vaccines for instance. This is not only due to the fact that people have different priors regarding these topics, but also because not everyone trusts science to the same extent, or because scientific messages can often appear politicized. To be able to consider these aspects would determine a descriptive improvement of an account like Simion (2023)'s that is limited by its own formalization to consider only people's priors. However, this descriptive improvement would not be accompanied by a normative one, since the shortcomings previously mentioned would

remain. Consider a true, scientifically supported proposition p, and given a group that, after the reception of a content c polarises with respect to p, namely, for some $\Pr(p|c) < \Pr(p)$ is the case, while for others $\Pr(p|c) > \Pr(p)$ is the case. To be able to state what reasons determined the reported epistemic effect in the former group would be a useful descriptive improvement, since it would allow to determine whether c is indeed an item of disinformation or not. However, at its present stage, the DIGC account described by Simion (2023) would state that c is disinformative as soon as $\Pr(p|c) < \Pr(p)$ occurs. As seen, c can generate the described effect while being a scientifically accurate claim, which is a suboptimal result. This impasse is due to the fact that the idea of disinformation is not related or conceived to be used to denote a peculiar epistemic process. For this reason, its determination or definition cannot be reduced exclusively to a limited set of epistemic dynamics. Disinformation is a concept that inherently combines epistemic and non-epistemic aspects, and whose real peculiarity is actually related to attitudinal factors. To consider the attitudes and perspectives of the receivers of a propositional content is crucial to better understand the predictors of the acceptance of a disinformative content and the rates of vulnerability that different individuals have to different items of disinformation. Therefore, I argue that to define a propositional content as an item of disinformation it is necessary to drop at least one of Simion (2023)'s premises. In particular, it is necessary to systematically determine the intentions of the creator and spreader of a content for the latter to be safely defined as disinformative. Therefore, in support of the attitudinal approach, it seems well motivated the claim that an item of disinformation is such when its creator(s) have misleading intentions. However, even this requirement is incomplete and needs to be expanded.

In particular, I argue that it is crucial to point out that the attitude of the creator(s) of disinformation is not independent from the one of the receivers. What defines the attitude and intentionality of the spreader of disinformation is not only the goal of deceiving the audience, namely making them believe in something false, but maximizing the chances that the latter happens, and this could be guaranteed only by targeting a specific audience, the one that is more inclined to accept as true the received disinformative content and behave accordingly. As seen in section 1.1 and related subsections, the recent history of disinformation has reported various cases of interested spreaders who tried to reach specific groups of people and trigger specific reactions. The

latter process was not left to chance. Disinformation has been spread strategically, often with the idea to reach the widest amount of people, but also to micro target individuals and groups who are more likely to fall into the disinformation trap. In other words, the fact that following the reception of true and scientifically accurate contents, polarizing dynamics might occur, and that these are reportedly correlated with people's differences in worldviews and perspectives, is not a minor aspect from the standpoint of the creator of disinformation. To be aware of the fact that there is a group who is epistemically and attitudinally inclined to accept certain contents is a necessary requirement to target the members of that group, since they are supposedly[7] more likely to accept and share those contents once they receive them. This perspective is supported for instance by Bennett and Livingston (2018) who argued that: "*disinformation* invites looking at more systematic disruptions of authoritative information flows due to strategic deceptions that may appear very credible to those consuming them." (p. 124). I argue that what ultimately distinguishes disinformation from other forms of information is precisely this combination of attitudes and intentions between creator(s) and audience, combined with the misleadingness of the contents shared from the former to the latter. Taking stock of all the aspects raised in this Chapter, it is now possible to define disinformation as follows.

I have previously argued that Søe (2021)'s account of disinformation as alethically neutral and intentionally misleading by definition is the best available in literature for two reasons. On the one hand, alethic neutrality allows to dispense with the category of malinformation and categorize both true and false items of misleading information as disinformative. On the other hand, this definition implies that the attitudes of the creator(s) of misleading contents are considered an essential component of the concept of disinformation. This requirement is conceptually proper and practically desirable for two reasons, an epistemic and an ethical one.

Indeed, whether a misleading content has been intentionally created is important to determine the eventual blameworthiness of its creator(s), namely it is relevant from an ethical point of view. Spreading disinformation online and on social media is now recognised as a potentially harmful gesture, for which spreaders should take full responsibilities. In 2024, anti-immigration riots took place all over the UK, and these have

[7] In Chapter 2, I will report studies showing that sharing a content and believing in its truth are not necessarily correlated aspects.

been fuelled by the spread of disinformation and hateful contents on various social media. In particular, the false story of a Muslim immigrant having killed six children in Southport was the most popular one among rioters (BBC 2024). Soon, the principal responsible for the spread of this specific item of disinformation was identified and sentenced to 20 months of jail (Fung 2024). Therefore, distinguishing between an item of misinformation and one of disinformation is also distinguishing between a blameworthy and responsible individual and an innocent one, which is not a trivial difference.

Epistemically speaking, the attitude of the creator determines whether a content is disinformative because the fact that it is conceived to mislead a particular audience implies also that the content has been conceived to be appealing for that particular audience, which is supposed to be one who is more likely to find the received content convincing, acceptable, and shareable with others. The intention to mislead can clearly acquire various forms, and it could fail, but surely presumes an attitude that is very likely to be translated in an effort to reach not only as much people as possible, but the fittest ones, namely those who are more likely to be persuaded by the received contents. In other words, the successful spread of disinformation is often the result of a correct expectation regarding both people's attitudes and their beliefs' formation processes. Moreover, the systematic inclusion of the creators' misleading intentionality as a necessary requirement for disinformation allows to avoid categorical mistakes of the kind committed by the DIGC account.

Furthermore, I argue that an account of disinformation ought to systematically consider the audience's attitudes to be complete. In particular, I argue that to fall prey of disinformation, the audience needs to have an attitude to be misled, which is not expressible with a reference to a unique state of mind and hence should be figured out as a range of attitudinal states. I claim that having an attitude to be misled means to be very likely to form a false belief out of the reception of an intentionally disinformative content. I endorse the idea that disinformation is alethically neutral by definition, thus, an intentionally misleading propositional content, spread to disinform a targeted audience, can be factually true even if conceived to imply something false. For a false implicature to be grasped out of a propositional content *c*, the audience has to be aligned to a certain narrative and worldview, namely the one of the creator(s) of *c*. Consider the following example. In 2023, an article posted on Breitbart News reported that "San Francisco braces

the first snow since 1976 as more winter storms arrive". Even if the content reported in the title was factually true, the article was written and structured to convey anti-climate change implicatures. Breitbart News is arguably the most famous alt-right news network in the USA, and a platform of daily disinformation sharing, particularly about climate change. For an implicature of the described kind to be grasped, and a false belief formed or reinforced accordingly, it is a necessary condition that the audience is aligned with the narrative and perspective of the authors of the article. Conversely, in the case of factually false disinformative contents, the acceptance of disinformation can be determined by ignorance, as the supporters of the GBM account argued, but also by motivated reasoning, or distrust in science, as it was the case in some of the reported examples, but also, and probably for the vast majority of cases, by inaccuracy and cognitive laziness (Pennycook and Rand 2019). Considering instances of disinformation related to climate change, the EDMO (EU's interdisciplinary network on disinformation) has reported various items of disinformation spread all over the EU area in 2022. Among the most diffused false stories there were many that included very easily debunkable stories, and which nonetheless managed to spread widely and being largely reported. The latter effect could be due to any of the mentioned aspects. Someone might be completely unaware of the state of art in climate science, hence having no reason not to believe that, for instance, the Arctic ice had reached its peak since decades in 2022, which was one of the most diffused items of disinformation that year (EDMO 2022). Others could have been distrusters of the scientific community who were prone to find intellectually attractive any content that contrasted scientific institutions and related recommendations. The point is that for a disinformative content to be accepted, and even shared, it is necessary that the misleading intentions of the creators find a fertile soil in their audience.

Therefore, the important point to highlight and stress is that the creation, sharing and spread of intentionally misleading contents in the digital environment is a socio-epistemic phenomenon that cannot be studied without taking into systematic account its finalized and interested nature. The intentionality behind disinformation's misleadingness is not purposeless. Recent history shows that the spread of disinformation is instrumental to interest groups with specific goals and ideological purposes. There would be no disinformation without the latter aspects. There could be fake news and misinformation, but not disinformation, which is inherently dependent on the intention to deceive an

audience by exploiting people's tendences, fears, biases, and other aspects that are influential during decision making. For all these reasons, any socio-epistemic analysis or definition of disinformation that is independent of both the creators' and audience's attitudes is incomplete because it does not reflect an essential component, namely that the success of disinformation in conditioning people's decision making, and polluting the public debate on relevant issues, is the result of the combination of two compatible attitudes, the one of the creators and the one of the audience. In more precise terms, I argue that the best definition of disinformation is the following.

An item of disinformation is an alethically neutral, propositional content, which is created and spread with an intentionally misleading attitude, and which is accepted as true and shared with others by an audience which has an attitude to be misled. This attitudinal state can consist of mere ignorance about the received content, propensity to biased and motivated reasoning processes that are conducive to the acceptance of the received content, ideological alignment to the narratives and worldviews endorsed by the creators of the received pieces of disinformation. Therefore, the attitudinal state that makes an audience prone to accept disinformation consists of the combination of epistemic and non-epistemic factors.

Conclusion of Chapter 1

In this chapter, I first listed some of the possible harmful effects that follow from the spread of false and misleading contents. The latter can persuade people to engage in behaviours that are dangerous for themselves and for the others, or that can trigger violent actions. The pervasive presence of misleading contents, within a social environment, can also pollute the public debate which might fail to properly discuss urgent topics, as it is still happening with respect to climate change, particularly in the US context. Apart from making explicit the dangers and risks related with the spread of false information, the main goal of this chapter was to properly define disinformation. I have considered the different possible definitions that have been offered of the latter kind of information, showing that many scholars, even with important theoretical differences, agree that what distinguishes disinformation from other forms of information is the misleading intentionality of its creator. Nonetheless, there are divergent voices, and I have considered in particular the account offered by Simion (2023) who argued that disinformation could

be defined exclusively on epistemic grounds, namely as a propositional content that makes more ignorant those who receive it. This account suffers of various shortcomings, and I have argued that those are related to its very theorical premise, namely that the definition of disinformation should be independent from considerations about the intentionality of the creators. The main claim that I have defended in this chapter is that disinformation is inherently related and dependent on a combination of epistemic and non-epistemic factors. In particular, disinformation is different from misinformation and information because what characterizes it is the combination of the misleading intentionality of the agent who creates it, and the propensity of the audience that receives it to be misled, namely, to grasp the disinformative content, accept it as true, and possibly share it with others. In this chapter, I have stressed the importance of recognizing that the latter process does not occur by chance, but it is the result of the deliberate attempt of the creators of disinformation to target the audience who is more likely to behave as described, namely, to accept, spread, and act upon the received pieces of disinformation.

2. Dynamics of Disinformation Spread: Technological Development, Microtargeting, Echo Chambers

In the first part of this chapter, I consider the role that the recent technological revolution, and in particular the development of new information-communication technologies, has had on the proliferation of disinformation, and the specific factors that are enabling the spread of false and misleading contents, online and on social media platforms. In particular, I will point out that the digital environment is inherently designed to maximize users' engagement, and their time spent on the platforms, which is a mechanism that inevitably compromises contents' accuracy, and that facilitates biased reasoning processes. In the second part of this chapter, starting in section 2.3, I focus on the topic of echo chambers for two reasons. First, there is an urgent need for conceptual clarification. In literature there are various and very different uses of the term 'echo chambers', and it is necessary to determine in a clearer way what this is meant to denote. Starting from Nguyen (2020)'s definition, I argue that these represent groups of individuals characterized by a common epistemic vice, namely an irrational mechanism of update of their state of belief. Second, the literature on echo chambers includes important results that allow me to argue that the current spread of disinformation and misinformation is not entirely imputable to the recent technological revolution. More precisely, I argue that the digital environment represents a fertile soil for the proliferation of disinformation, and for the problematic consequences that derive from it, with socio-political polarisation being one of the most relevant ones. However, it would be wrong to claim that the spread of disinformation, the formation of echo chambers, and the polarisation of the society are direct results of the development of the current digital environment, and in particular of content recommendation systems, that force people to cluster in groups of like-minded users. On the contrary, polarised groups and disinformation disseminators have found in the digital environment the perfect space to interact, share contents, and implement communication strategies that have direct and real effects on the society.

The chapter is structured as follows. In section 2.1, I make an historical analysis of the different national approaches to information regulation and to the contrast of disinformation dissemination. In section 2.2, and related subsections, I discuss some of

the dynamics that facilitate the proliferation of disinformation online and on social media. Sections 2.3, 2.4, 2.5, and related subsections, are dedicated to echo chambers.

2.1. Different Approaches to Disinformation: US and EU

When considering the problem of disinformation from a historical perspective, some doubts may legitimately arise. It might seem that national and international governments, alongside digital platforms, started facing the problem of how to regulate the flows of information online a little too late. It was to be expected that in an open and free environment like the digital one, someone could have tried to use it inappropriately and to spread misleading and false information to pursue personal or group interests. A similar critique could be made regarding the responsibilities of social media companies, which should have been clarified in advance, particularly regarding their role in the control and regulation of the information that circulates on their platforms. This is a very difficult issue to settle now, since social media companies occupy a position of great power and economic influence. To have a clearer perspective on these aspects, a brief historical analysis might be useful.

Internet's early days were characterised by considerable optimism. The diffused perception was that a free circulation of information, from multiple sources, would have meant free access for many to the largest number of pieces of information ever collected, and that the benefits for every democratic system and relative public debate would have been massive. The possibility of the proliferation of false information and fake news has been envisioned as a possible risk, but more as a "top-down" phenomenon, meaning one related to the possibility that governments, international institutions and business companies could have used it to make propagandistic messages and exercise their control over the people (Ballardini 1995). The idea that disinformation and propagandistic claims could have been produced and shared also intra-internet users was not treated as a serious risk. Alongside the optimism of the early days of the digital era, there are other technical and socio-political factors to consider. In the 90s there was a tendency to consider the dynamics of exchange and interaction within the Web as unchangeable, or at least stable for a long while. In other words, the rhythm of technological innovation witnessed in the following 30 years, regarding in particular the information communication technologies (ICTs), had not been predicted by many commentators. This factor has contributed to the

present situation, where governments and international institutions struggle to keep the pace of the information technology industry. It is necessary to predict the upcoming developments of the ICTs' technologies and regulate them to prevent possibly dangerous future applications (similar discussions are now daily made regarding the regulation of AI). However, in order to do that, it is first necessary to recover the delay accumulated in the early days of internet enthusiasm. This is not an easy task to accomplish since most of the expertise needed is employed in industries and companies that actively produce information technology. Moreover, as seen in the previous paragraph, many political forces have made use of disinformation not only to gain consensus but also, as I will show later, to characterize their political identity and target specific groups of potential voters. Therefore, it is very unlikely that this kind of politicians would do their part in facing the issues related to the spread of disinformation online. Nonetheless, in recent years, both at a national and international level, regulations of social media platforms have been implemented.

In 2016, the European Union has adopted the General Data Protection Regulation (GDPR) that defines the legal boundaries, within the European Community, about data regulation and privacy protection, and in 2018 the European Commission ratified the Code of Practice on Disinformation, updated in 2022 with even stronger demands, which binds the biggest and most important social media companies in the World, who all signed the Code, to commit to actively fight disinformation and monitoring users' activities in their platforms. This event clearly represented a concrete intervention to regulate social media activity and control the influence that this has on European citizens. More implicitly, it also represented the concrete endorsement of the idea that disinformation can be harmful and cause really detrimental effects on people's lives and on socio-democratic stability. Much more complicated is the situation in the US, and for fascinating historical and cultural reasons.

As Fukuyama and Grotto (2020) showed, in Europe, since the end of World War II, each country has taken a different approach to media regulation, first with newspapers, then with television, and now with social media and digital platforms. Fukuyama and Grotto (2020) take the example of France and Germany to show how different two approaches to media regulation are. According to Hallin and Mancini (2004), France and Germany reflect two different ways to manage the relation between politics and media.

France belongs to the "polarised pluralist model", where public broadcast is openly politicized, whereas Germany belongs to the "democratic corporatist model", where politically neutral media forces are constantly on the rise, with a Länder-level public broadcasting financially supported to prize balance and neutrality. These differences are the result of historical and cultural mechanisms. In France, since the end of World War II, the attempt has always been the one to favour a plural French-speaking media environment, protected by market pressures, through national financial support. In Germany, a plural media environment has been considered one of the most important things to preserve in order to prevent the insurgence of new totalitarian forces. Thus, the Länder-based system is conceived as a way to avoid the centralization of media power in a single institution.

These different national approaches to traditional media management, and the reasons behind them, have perfectly been translated in national regulations regarding the activity of social media companies. In Europe, the various models of media regulation have an element in common, the awareness of the role and influence that media can have on the democratic processes. This historical awareness has had an impact on the recent history of EU's media regulation, both in terms of data protection and of disinformation prevention, and it possibly explains the attention and the need for solid regulation expressed in recent years. The US system of traditional media regulation is very different, as it is the recent story of social media regulation as well. First of all, the US came out from World War II in a completely different situation than the one of the European countries, with no comparable urgency to regulate the media in a way that would never allow the centralization of information production in totalitarian hands. Therefore, the plurality of information sources has been accomplished and regulated as in any other sector, by market rules and competition, with no need for a central media regulation. People can freely choose the sources that they prefer, and plurality of sources will come as a natural and positive consequence of market rules. This general neoliberal tendency famously reached its peak in the 80s, but less famously it reached another peak during the early days of the internet era. The Web has been interpreted from a neoliberal perspective as a major chance, being it able to offer free information to consumers with minimal government intervention involved, just by leaving competition regulate information provision and sharing (Fukuyama and Grotto 2020, p. 211), and this optimism brought to

a loose approach regarding the possible risks of an unregulated spread of information, including the possibility that users share false news. Consequently, the ruling neoliberal perspective, that equalized information to any other good, made it possible for US based Web sources and proto-social media to implement business models and, more importantly, algorithmic designs conceived to reach the widest possible audience, with no journalistic deontological duty to preserve information's quality or a proper and plural public debate, since the main duty of any media source was simply to maximise the number of their readers. This system has contributed to the media degradation that has invested the US in many ways, with visible results that today contribute to the difficulties in media regulation. Many US journals, particularly local ones, used to heavily depend on revenues from advertising, but now companies promote their products on social media for the vast majority. In absence of revenues from advertising, and in a concurrency-based environment like the US one, where national aids are very limited, many local newspapers did not survive the outbreak of Web sources and social media. The numbers of this fall have been merciless, with a reduction of 26% of newsroom employment in the interval 2008-2021 (Walker 2021). The disappearance of so many newspapers reshaped the distribution of media coverage in the US territory, with the consequent occurrence of what have been called "news deserts" (see the report by Abenathy 2018), namely areas where there is not even one local newspaper left. In these areas, mostly localized in the heartland of the US territory, people can now only rely on the internet and on social media for information gathering. According to Gottfried and Shearer (2016), the 49 % of US citizens reported getting news from social media, with Facebook on top of the list. The resulting media degradation has been severe, with results showing high levels of distrust in traditional media, young people being disincentivized and disinterested to get informed via whatever channel they might prefer, and increasing information polarisation as well (Starr 2020). One of the most important consequences of the settlement of this media environment has been that supply and demand for extreme and radical contents has reached a point of equilibrium on a high quantity level. This has come at the expenses of good journalistic practices. In particular, it has negatively impacted the quality of the information provided and has led to the presence of an increasingly more sensationalist language and rethoric, functional to elicit readers' emotional spectrum more than the analytic one. In recent years, the progressive isolation of far-right media channels in

closed clusters, and scandals like Cambridge Analytica, and the Russian interference during the 2016 Presidential Campaign, have determined a turn in the North American history of media regulation. After these extremely serious events, the hypothesis of a regulation acquired public and political support, but legal issues complicated media regulation. Social media companies have been able to convincingly represent themselves as mere providers of communications tools, irresponsible for the contents that people share within their platforms. Two factors made that possible. Regarding content control, the First Amendment still represents an important protection for social media corporations. Conceived to preserve and protect freedom of speech, the First Amendment represents an important aspect to consider when discussing how to limit or control a privately owned and managed communication tool. Alongside, and maybe more impactfully, the Section 230 of the Communication Decency Act, adopted in 1996, makes platforms almost immune from liability for the activities of their users and the contents that these create and share. This has contributed to the protection of platforms and indirectly justified the picture that these have offered of themselves as mere providers of spaces of communication and interaction. Therefore, it is legally and politically complicated to regulate the activities of social media platforms in the US context, and even more, to force them to practice some form of self-control and to actively engage in the fight against disinformation. It is important to notice that this is not a relevant issue only for the social media users located in the US. In a globalized and connected world, the spread of disinformation is a universal matter of interest, since the use that social media companies make of users' data, the availability of the latter for third parties who might use them to influence people's decision making, and similar issues are of a kind that have an international relevance. To better see the aptness of the latter point, it is necessary to consider how disinformation proliferates on social media and why this environment is the most fertile for the spread of disinformation.

2.2. How Does Disinformation Proliferate?

In Chapter 1, I have mentioned Floridi (2014)'s claim that in the current digital environment, and particularly on social media, anyone is at the same time a consumer and a producer of pieces of information, even by simply sharing contents created by others. As Tucker et al. (2018) have rightfully argued, in an informational and communicative

environment of this kind, the role of the gatekeepers, who should supervise the debate that occurs within, is compromised, and the risk of proliferation of unreliable sources and inaccurate pieces of information is inevitably high. Moreover, false and misleading contents circulate not only because it is difficult to monitor communication online, but also because the current digital environment is designed in a way that recommends contents not on the basis of their accuracy, but for their ability to entertain social media users, in the attempt to maximize the time spent within the platform.

Online, and more specifically on social media platforms, companies and contents' creators compete to gain the attention of the users and turn them into consumers of products and services. This mechanism, which has been defined an "attention economy" (Williams 2018), has caused an overabundance of contents, given the limited attention of the users. For this reason, the algorithmic logic that regulates contents' recommendation is designed to constantly improve the users' profiling in order to make their digital experience as tailor made as possible.

Therefore, social media platforms are environments designed exclusively to be engaging, and thus inherently modelled to prize contents that are more likely to be interesting for large numbers of users, independently of the accuracy of the information included. Unfortunately, people do not spend time on social media only to be entertained, or to discover new products they might be interested in, but also to gather pieces of information and knowledge. As a consequence of this, the described mechanisms and logic of content recommendation have deeply influenced the work of information providers. To be competitive on social media platforms, media and news networks must adapt their proposal, in terms of published contents, to an environment that prizes virality more than everything else, accuracy included. Seeking virality might lead to suboptimal practices, like the production of partisan contents, in the attempt to target specific groups and be more often recommended to them, or to address topics that are more likely to trigger users' reactions, independently of their actual relevance in the public debate. As the accuracy and quality of the information shared is not a priority, the soil for the spread of misleading and false information becomes fertile. As provocatively claimed by Diaz Ruiz (2023), in an environment regulated by a business logic, as social media platforms currently are, where the common goal of contents' creators is reaching virality, the spread

of misinformation[8] is an "expected outcome" (Diaz Ruiz 2023, p. 15) not just a shortcoming, meaning that it is inevitable for structural reasons of platforms' design.

In particular, the communicative dynamics that regulate interactions within social media platforms foster the sharing of highly emotionally and morally charged contents which are very entertaining for the users, and hence able to trigger various reactions, including likes and comments, but also to incentivize their sharing, and in general to increase their virality. More precisely, as Brady et al. (2020) have highlighted, social media incentivize the proliferation of morally and emotionally charged contents because interactions occur within large audiences and in an impersonal way, thus for users who want to communicate something with as many people as possible, it is necessary to share contents that elicit an immediate reaction in the audience, and such that the intended receivers can immediately grasp the underlying message.

Therefore, for ecological reasons, contents' creators and social media users are incentivized to communicate using impactful but simple contents. Social media represent a great platform for people to let the others know who they are, and therefore, those who aim to express their group membership, identity, and political views are incentivized to share morally and emotionally charged contents as well, so that the other users can easily recognize the latter as signals of group membership, or messages of solidarity and complicity. Unfortunately, social media users tend to signal their belonging by sharing contents against out groups, more than contents in support of their group. Rathje et al. (2021) showed that contents that include messages against out-groups, and that express out-group animosity, are more likely to be shared and to become viral. It is important to notice that, in an environment where social identity is made salient by structural and communicative features (Brady et al. 2020), the proliferation of emotionally charged contents, shared against out groups, could be conducive not only to polarisation, that I will consider in greater detail in section 2.3, but also to the spread of disinformation.

Rathje et al. (2021) reported that political disinformation often includes negative contents about out-groups, and since these are the ones more shared by social media users, it is very likely that pieces of disinformation proliferate because certain users share them to manifest their group membership. Moreover, the latest results worsen this picture showing that there is a tight connection between contents that express moral outrage and

[8] In Diaz Ruiz (2023)'s work misinformation is intended as generically false and misleading information.

false and misleading ones. McLoughlin et al. (2024) have reported that social media users tend to share, on average, more contents that elicit moral outrage independently of their reliability and trustworthiness. Therefore, it very likely that users share pieces of disinformation and misinformation, even by mistake, because those foster moral outrage and the related emotions, with anger among all. Coherently with this possibility, McLoughlin et al. (2024) reported results showing that false and misleading contents are better able to elicit moral outrage in users than contents from reliable sources. Even more explicitly, Carrasco-Farré (2022) has analysed 92.112 different news article with the intent to spot the most peculiar features that characterise the misinformation that proliferates online and showed that the use of highly emotive language is one of the principal ones.

In Carrasco-Farré's work the umbrella term 'misinformation' is used to cover a wide variety of false news including 'junk science', which is a label that indicates something like pseudo-science and scientific disinformation. This variety of terms is motivated by the fact that false contents are not always of the same kind, and for that reason, it is necessary to design a scalable definition for false news, to be able to face it independently of its contingent form. Carrasco-Farré's most important result is to have shown that it is possible to study all the different forms of false news by considering only two variables, the level of emotivity that a piece of news appeals to, and the cognitive effort required to elaborate it. Carrasco-Farré (2022) showed that, in general, items of misinformation demand less cognitive effort than reliable news, and try to evoke emotions, with words' choices and appeals to morality, considerably more often than accurate news. Interestingly, among all the different forms of false news considered in this study, junk science is the closest one to factual and reliable information. In other words, pieces of scientific disinformation and misinformation turned out to be the ones that resemble the scientifically accurate ones the most, which might indicate that for the users it is complicated to distinguish the two. Leaving this aspect aside, in a nutshell, the main result offered by Carrasco-Farré (2022) is that appeal to emotions and low cognitive effort are the two most peculiar aspects that characterize misinformation and disinformation. The best explanation available is that these contents are conceived, structured, and presented to be easy to remember, easy to understand, and able to stimulate the more immediate, less critical and analytical reasoning process in the

receiver, and also that their creators, in the case of disinformation, try to exploit the current ecological features of the digital environment and make these contents going as viral as possible.

In another seminal study, Vosoughi et al. (2018) have tested whether appealing to emotions and designing very easily understandable false contents is efficient to attract and persuade the receivers. Their results show that this strategy is more than efficient. The comparison between false and true stories revealed that false ones spread in considerably less time, and reach more people than accurate ones, with political news being those with the largest audience, even more than scientifically related ones. Pennycook and Rand (2021) reported evidence of the efficacy of appealing to emotions, particularly fear and anger, showing that the latter emotion is a predictor of higher rates of sharing on social media (more on this will be said in Chapter 4). What the mentioned results seem to reflect is that the low cognitive effort required to elaborate false contents fosters an immediate sense of clarity and intelligibility, which paves the way for the acceptance of the message included. Moreover, the ability to stimulate the proper emotions fosters reactions that include likes, comments, and sharing that contribute to their virality. Further empirical evidence of this comes from the work of Stieglitz and Dang-Xuan (2014), who reported that, with a data set of 165000 tweets, the emotionally charged ones were much more likely to be retweeted. Considering a specific category of creators of scientific disinformation, those who create anti-vaccine messages strongly rely on the emotive reactions and status of their audience (Mitra et al. 2016, Hughes et al. 2021), as it might be the case of parents who are deciding whether to vaccinate their children or not (Shelby and Ernst 2013). Therefore, with respect to disinformation, the fact that false contents spread faster and further than accurate ones seems to be related to the ability of their creators to exploit social media's mechanisms of content recommendation and design contents that are more likely to reach a broader audience than accurate ones, and in a considerably limited amount of time, as reported by Vosoughi et al. (2018).

From a cognitive perspective, the fact that emotionally charged and low cognitively demanding contents are more likely to be accepted or shared can be descriptively explained considering the work of Kahneman and Tversky (1979), and previously of Simon (1955), who have shown that to have a complete view of the factors

that influence people's decisional dynamics, it is necessary to consider the context of decision making. As seen, in the digital environment, information proliferates in high quantities and its elaboration occurs in a very limited amount of time, particularly on social media platforms. In a context like this, people are ecologically incentivized to elaborate contents using their more intuitive, emotionally sensitive, less analytic cognitive mechanisms and heuristics available. The latter is not necessarily a negative dynamic, since people benefit from being able to decide quickly in many situations, but this could also yield to biased reasoning processes. Therefore, the spread of disinformation, online and on social media, might be due to biased and emotionally driven responses. Empirical studies show that cognitive biases and automatic reasoning processes play a significant role in the selection and acceptance of the contents found online (Lau and Coiera 2007). In the next sub-sections, I will consider a series of cognitive processes that are conducive to the acceptance of false and misleading contents. More specifically, I will focus on the reasoning processes that are more likely to play a role within the digital environment, namely during interactions between social media users, and during information gathering online and on social media.

2.2.1. False and Misleading Memories

Misinformation and disinformation have significant effects on people's memory. Studies in psychology and cognitive sciences have shown that being exposed to false and misleading contents alters people's recollection of past events, thus compromising their factual accuracy (Greenspan and Loftus 2020, Loftus and Klemfuss 2023), which means that new pieces of disinformation and misinformation could be accepted as a result of the effect that previously accepted disinformation and misinformation has on memory. This mnemonic effect is complicated by two interrelated factors.

First, it is more likely to be committed to false or inaccurate memories whose content is coherent and aligned with one's current beliefs. This is due to the general effect that confirmation bias plays on states of credence. Confirmation bias is a cognitive mechanism that makes a content looking more accurate due to its coherence with previously endorsed beliefs (Nickerson 1998, for a more recent discussion on this bias see Peters 2022). Reportedly, this has an important effect on the acceptance of items of disinformation (Del Vicario et al. 2016), with a resulting reinforcement effect on the same

priors that determine its emergence, and it influences also the recollection of past events, or the formation of biased memories (Lewandowsky et al. 2012). Second, many studies have now confirmed that disinformation and misinformation generate a "continued influence effect" (CIE), which means that a piece of misinformation or disinformation can have a very long-term effect on the recollection of past events, even after the exposure to counterevidence or retractions about the accuracy of the misinformative (or disinformative) content considered. The latter effect clearly complicates the elaboration of an efficient, and long lasting, debunking strategy (Lewandowsky et al. 2012). Psychological studies suggest that the CIE of misinformation is possible because a misleading content has a more pervasive cognitive effect than its correction and debunking (Ecker et al. 2022), and this for a simple reason. In order to correct a misleading content, it is almost inevitable to recall and mention it. The debunking of a fake or misleading news article for instance, often consists of another article which reports the content of the misleading news and an explanation of what was wrong about it. The latter apparently harmless and inevitable process determines the continuous recollection, and related misleading effect on memory, of the same content that the correction was meant to solve. Moreover, the numbers of encounters with a determinate content influence its perceived accuracy, hence corrective attempts might even end up reinforcing false beliefs. In fact, a sense of familiarity that results from repeated encounters can be followed by the so-called "illusory truth effect", a cognitive bias that could determine the reinforcement of the same credence that the debunking content was meant to correct. Due to its proven effects on the spread and acceptance of misinformation and disinformation, it is worth dedicating a separate sub-section to the illusory truth effect.

2.2.2. Illusory Truth Effect

There are research outcomes showing that repeated exposure to the same content makes it looking more accurate, reliable, and even true (Fazio et al. 2015, Wang et al. 2016, Mattavelli et al. 2024). This cognitive effect is called "illusory truth effect" and is believed to be particularly dangerous on social media and the Web, since people have only partial control on their rates of interaction with different contents, being it algorithmically determined and related to one's activities online, and also by the contents shared by friends and frequently visited pages. This is concerning with respect to fake

news and items of misinformation and disinformation. Particularly on social media platforms, that are designed to prize virality, creators of disinformation can exploit the illusory truth effect by making a piece of disinformation shared by as many sources as possible and thus make it more likely to appear multiple times in many users' feeds. For all these reasons, scholars who are interested in disinformation, fake news, and related issues, have considered the illusory truth effect a relevant cognitive bias to study. Importantly, as Swire et al. (2017) have noticed, it is important to be able to empirically distinguish between the effect of familiarity bias and of recollection. Familiarity is a fast, automatic cognitive process, that binds recognition with accuracy, whereas recollection is slower and inherently related to the contextual features of the previous encounter with the same informative content, and so arguably not involved with biased reasoning, or knowledge formation processes.

Focusing on the pervasiveness of illusory truth effect and its impact on the sharing of false contents, Vellani et al. (2023) showed that, interestingly, even after a single repetition, people considered the news that they have already encountered to be more accurate than those seen only once. This is the case also for fake news and items from non-credible sources, and even in presence of a research population who reported a general aim to share information they considered true. In this study, news related with politically charged issues were deliberately excluded, but there are scholars who have investigated the possible combination between people's political orientation and the illusory truth effect. Pennycook et al. (2018) have collected important results showing that repetition plays no effect only for plainly absurd fake news. However, as soon as the news become more believable, being them true or false, repetition has a significant effect in the attribution of accuracy. In this study as well, participants reported a clear intention not to share contents they did not consider true, and even in this case, only one reencounter was sufficient for the illusory truth effect to occur. With respect to political convergence, results showed that participants judged the pieces of information that were aligned with their political ideas as more accurate than the discordant ones, but there was no combined effect between repetition and political valence, meaning that an illusory truth effect was detected for both politically discordant and concordant news, and independently of their truth or falsity. Moreover, Pennycook et al. (2018) reported the detection of a direct proportion between the number of exposures to a content and the

level of accuracy attributed to that same content, with this effect being detected also for the false ones. Thus, the illusory truth effect represents a plausible explanation for instances of immediate and sloppy acceptance of false contents, and also a major obstacle in the fight against disinformation. At this point, it is fruitful to discuss an important, and perhaps counterintuitive aspect.

Working on a data set of shared tweets, Vosoughi et al. (2018) found out that false news that appeared to media users more novel than true news were more likely to be shared. Therefore, novelty turned out to be a predictor of the sharing of false contents. However, in this subsection I have reported results showing that repeated exposure is a plausible cause of the spread of misinformation and disinformation online, so these results may appear to be in contrast, but they are not. The cognitive effects caused by repeated exposure concern the perceived accuracy of a content, and even if a subjective attribution of accuracy could predict the sharing of a content, it is important to notice that sharing is a peculiar action that follows its own dynamics, and that it is not necessarily related with perceived accuracy or belief. It has been empirically shown that people share contents they are not doxastically committed to or that do not consider to be epistemically accurate (Pennycook and Rand 2020, Pennycook et al. 2021). The latter is due to the fact that people often have non epistemic reasons to share a piece of information in a public platform, or with a specific audience. Someone might be motivated to share a content to signal one's group belonging and social identity, or to manifest their support to a certain position or idea (Piazza and Croce 2022, Bergamaschi Ganapini 2021, more on this on Chapter 3). None of the latter motivations is related to the perceived accuracy of the information shared, and in fact the epistemic and non-epistemic dimensions might also clash sometimes. Therefore, it is important to be aware that the illusory truth effect could lead someone astray, and that it can make an item of disinformation looking accurate simply due to the repeated encounters with it. However, it is important not to attribute to an effect of this kind the entire explanation of the dynamics involved in contents' sharing. Nonetheless, the reported empirical results are very important, and this also because, as Vellani et al. (2023) have rightfully stressed, they point out the cruciality of timing in the detection and debunking of items of disinformation and misinformation.

2.2.3. Sense Making

Another possible reason for the spread of false contents is the need for sense making. Particularly in conditions of difficulty, some people might find psychological relief in false and conspiratorial beliefs, at least as long as these apparently explain the current experienced struggles and bad feelings. Goertzel (1994) showed a correlation between living in conditions of social insecurity and the tendency to accept conspiracy theories. When there is no immediate explanation for a certain situation, or when the reasons are complex and variegate, there are people who are prone to accept, or behave as if they accepted, even blatantly false and implausible explanations. This phenomenon has been studied by Sunstein and Vermeule (2009), who stressed that the kind of sense making associated to the acceptance of false and conspiratorial claims is mainly related to the detection of individuals or groups who are imputable for the occurred event or situation. The importance of determining who is blameworthy for a certain condition, even when there is no agent who is directly responsible, has been discussed by Popper (1972, pp. 123-125) as well. Popper recognised the existence of a "conspiracy theory of society" which consisted of a propensity to see social institutions as groups characterised by collective intentionality that always act consciously and following a precise and deliberate plan, as if they are able to determine every consequence of their decisions, and hence as always inherently imputable of something. This perspective paves the way for conspiratorial attitudes that are grounded on the exclusion of any kind of casual element from reality. The occurrence and psychological strength of mechanisms of this kind has been widely reported in literature. For instance, Van Proijen and Douglas (2017) detected a non-trivial link between the occurrence of impactful historical events and the spread of conspiracy theories related to them, and notably within the groups affected the most by the bad consequences that followed from the same impactful events. People feel lost and powerless and so "[...] engage in increased sense-making activities to imbue a situation with meaning and purpose. If belief in conspiracy theories is a way of making sense of a situation, it follows that such beliefs are increasingly likely to the extent that people experience uncertainty or a lack of control." (Van Proijen and Douglas 2017, p. 327). This necessity to restore a sense of control on one's condition might be an instance of the psychological need for cognitive closure that some people manifest, and which is indeed empirically correlated with conspiratorial thinking (Marchlewska et al. 2017).

Conspiratorial thinking is a peculiar phenomenon, and conspiracy theories are not instances of disinformation. On the contrary, the two should be kept sharply distinguished. However, there is a correlation between conspiratorial thinking and the acceptance of disinformation (Lewandowsky et al. 2013) which justifies the idea that in some cases, factors that predict the former could also predict the latter. Therefore, it is proper to consider the possibility that the spread of disinformation is caused by the need for sense making, particularly after the occurrence of impactful events, and notably among those who have suffered the most because of them.

2.2.4. Lack of Accuracy

There are positions in literature that criticize the idea that cognitive biases are the principal reason for the spread of false and misleading contents online. In a study that has now become pillar, Pennycook and Rand (2018) argued against the idea that the principal cause of the acceptance of inaccurate and unreliable contents is their coherence with previously endorsed beliefs, and for this reason, they focus on confirmation bias. The idea of Pennycook and Rand (2018) was that the low propensity of users to engage in analytical thinking is the real responsible factor for the spread of misinformation and disinformation online. Coherently, their results showed that people with higher propensities to analytical thinking were better able to distinguish between reliable and unreliable news, with political partisanship having no significant role in influencing analytical thinking, positively or negatively, a result that was interpreted as evidence of a nil influence of confirmation bias during information elaboration. From this study, it seems that the propensity to think analytically, and not in a loose or biased way, is the real predictor of people's considerations of accuracy.

There is room for a mediating point here, since it is possible that people are not simply and sharply separable between those who can or cannot think analytically, but that most of the times Web and social media users are not incentivized to do so due to environmental factors. Panizza et al. (2022) reported interesting results while working on a similar hypothesis. The authors showed that monetary incentives are efficient to increase people's accuracy online, both in isolation and in synergy with other forms of interventions, including messages and pieces of advice on how to detect accurate news. Therefore, it is possible that the role played by cognitive biases is overestimated, and that

these simply reflect low levels of attention, or incentives to be accurate, that people have when they elaborate determinate contents.

Before that the discussion moves forward, it is important to consider other mechanisms that could not be rightfully listed as cognitive biases, but that could play an important role during information evaluation.

2.2.5. The Web Plays by Its Own Rules

Alongside the efforts made to understand how disinformation, misinformation, and fake news spread, it is also necessary to understand how to make people better equipped to evaluate the pieces of information they run into while online and on social media. Some scholars have started working on this issue and offered useful results and promising strategies of interventions. An underexplored but significant risk is that social media users are led astray by the application of heuristics that are apt for environments that are different from the digital one, and in particular that are appropriate for certain sources of information (e.g. printed newspapers, educational or academic books and articles), while not for the Web and social media. This is the idea behind part of the work conducted by the Digital Inquiry Group (the former Stanford History Education Group). The aim of the studies conducted by this research team was to define a series of strategies that people could have adopted to orientate during information collection online. The research framework designed has been called "Civic Online Reasoning" and had a specific target group, high school students. The necessity to focus on this specific demographic group is twofold. First, and maybe counterintuitively, results show that the younger generations are no less exposed to risks of the mentioned kind. Those born and raised in an already digital world fare no better than the older in the evaluation of the accuracy of the contents found online (Bakke A. 2020, Gasser et al. 2012, Hargittai et al. 2010, List et al. 2016, Martzoukou et al. 2020). Moreover, teenagers spend a considerable amount of time on social media, and in general on their smartphones. The 78% of Italian teenagers reportedly spend more than four hours a day on the internet and social media (Di Cristofaro 2020), and the 62% of US teenagers more than 6 hours a day (Rideout and Robb 2019). These data indicate a severe exposure to online contents, which means a higher chance of running into fake and misleading ones. These results motivate the necessity to investigate the best possible strategies to improve young adults' critical

abilities to evaluate the accuracy of the information found online (Breakstone et al. 2021a; 2021b, McGrew et al. 2018). The idea is to make high school students aware of the fact that, within the digital space, there are different rules to follow and that it would be inefficient to apply there the same heuristics used when reading a textbook at school, or an article on a newspaper.

2.2.6. Refuting Cognitive Determinism

To efficiently study the spread of misinformation and disinformation, online and on social media, it is important to understand how automatic and biased reasoning processes influence people's evaluations of the information found online. However, as it is important to stress the distinction between sharing a piece of information and accepting its content, it is also crucial to keep sharply separate the ascription of accuracy to a content, that is determined by biases and automatic cognitive mechanisms, with the doxastic commitment to the same content. In a context of explanation, one example of a problematic trespassing of this kind, namely one that argues for the formation of a belief out of a biased attribution of accuracy, is included in the following quote from the mentioned work by Pennycook et al. (2018), who argued that: "Finally, our findings have implications beyond just fake news on social media. They suggest that politicians who continuously repeat false statements will be successful, at least to some extent, in convincing people that those statements are in fact true" (p. 1880).

There are at least two major risks that could follow from such an arbitrary extension of the explanatory force of illusory truth effect, or of any other cognitive bias. The first is to attribute some form of gullibility to people. To claim that political rethoric and persuasion can be reduced to simple repetition, and that the more a speaker repeats the same content the more convincing their words are for a considerable amount of people, is too simplistic and obliterates many crucial aspects that characterize politics as (also) the fine art of gaining consensus. Moreover, there is considerable evidence proving the limits of this approach. Mercier (2020, Chap. 9) reported the results of studies on the efficiency of the totalitarian propaganda in Germany, during the 1930s, showing that the repeated exposure to the same contents and ideas had no effect on the general population. The constant repetition of antisemitic claims, for instance, resulted convincing only among the already antisemitic groups, whereas negative results were recorded among

groups and communities where antisemitism was not diffused. Therefore, historical episodes and related evidence does not support Pennycook et al. (2018)'s claim. Moreover, the second risk related with the quoted claim is to confuse amusement with belief and action guidance. There could be many reasons behind a liked post that includes disinformation, or a positive comment below a false claim, and it is not a settled or simple issue to determine what kind of epistemic or discursive action is the one performed through sharing a content on a social media page or by re-posting someone else's contents (Marsili 2020). In general, the problem behind a claim like the mentioned one is that it is grounded on a deterministic picture of people's decision making, which is presented as something predictable on the exclusive basis of the contents spread within the inhabited environment, an approach that limits the explanatory possibilities of this literature and that does not allow to see all the dynamics, epistemic and not epistemic, that are hidden behind the spread of misinformation and disinformation, and which go beyond biased reasoning processes, as argued also by Altay et al. (2023). In the next section, I will consider a specific debate where this risk is run as well. I will elaborate on the discussion introduced in section 2.2, regarding the role that algorithmically regulated systems of recommendation have in the spread of false information on social media, and the possibly related polarisation of users' positions. I will focus in particular on the topic of echo chambers, and related literature. My claim is that the use and meaning of the metaphor of the 'echo chamber' is largely unclear and confusing, and that this should be used to denote groups whose members are characterised by a common, irrational decisional pattern.

2.3. Are Echo Chambers Real?

The development of the current digital environment has been indicated as one of the causes of the proliferation of false and misleading information (Starr 2020). The latter is due to some of the ecological features of the current digital environment that I have already mentioned, but that is worth recollecting. The current mechanism of content recommendation online and on social media is designed to maximize the time spent on the platform. For this reason, contents that are entertaining and that trigger users' reactions are more likely to become viral and reach a vast audience. Therefore, creators and spreaders of disinformation can exploit these features and design contents that are likely to incentivize users' reactions. Indeed, in a digital environment that prizes virality,

the creators of disinformation have strategically leveraged on two factors, the use of highly emotive language and the poor cognitive demand required to elaborate their contents. This picture is already worrisome, but a further element, already hinted in section 2.2, must be added to enrich it.

Some scholars have argued that the development of social media platforms, and the consequent shift of the public debate into the digital environment, not only has determined the proliferation of disinformation, but has also contributed to the current political polarisation, witnessed in many democratic societies. One of the discussed possibilities is that interactions on social media pollute and compromise the possibility of a fruitful public debate because many users share determinate contents with the exclusive intent of discrediting their political opponents (Crockett 2017, Van Bavel et al. 2021, Van der Linden et al. 2020, Rathje et al. 2021), and radicalizing the language used within the public debate. Other scholars have argued that the current system of content recommendation has further problematic political effects. As seen, on social media and on the Web in general, algorithmic recommendation systems select contents that are aligned with users' ideas and profiles. This includes also political contents. Therefore, many scholars have argued that a form of political microtargeting causes an over exposition to contents from like-minded sources, causing a radicalization of people's ideas and consequently polarisation. Recently, this worry has been increased by the development of AI-based systems, whose capacity to match people's profiles and contents can even refine the current recommendation systems in the near future, and hence possibly increase polarisation (Simchon, Edwards and Lewandowsky 2024). The risks related with political microtargeting, including the spread of political disinformation and its suboptimal effects on the public debate, have been object of an important discussion. In 2016, after the Russian interference during the US Presidential campaign, and the Cambridge Analytica scandal, the possibility that groups of interest could make use of social media platforms and exploit the mechanisms of content recommendation to target people with disinformation became reality. As reported by Zarouali et al. (2020), later in 2016 these concerns were furtherly supported by the fact that companies like IBM put on the market systems for personality-congruent targeting, open to everyone who can afford them, and that represent a precious tool for interest groups willing to be more precise in targeting their intended audience. The latter hypothesis is perfectly plausible,

since results have shown that microtargeted political contents are very persuasive when designed on the personality of the audience (Zarouali et al. 2020). In an environment where people are exposed to news that are conceived to be interesting for them and coherent with their profiles, and where extreme and highly emotive contents are very likely to become viral, it is rational to consider the risk that every social media user turns into a Narcissus, using Gunn (2019)'s metaphor, who mirrors themselves into their own social media feed. On the long term, the most worrisome perspective is that social media users will be clustered in ideologically homogeneous groups, where only determinate ideas and perspectives proliferate, and where the extreme contents shared foster the radicalization of the endorsed beliefs. Therefore, it is urgent to determine to what extent, in the digital environment, users are free to select their sources of information. Indeed, it is possible that interactions and information exchanges are already a priori filtered in a way that fully prevents the exposure to contents that are not aligned with users' priors, a condition that would represent a serious limitation of people's freedom and autonomy of choice.

A perspective that does not support the latter reading is the one of Sunstein (2017), who argued that the current digital environment is indeed able to facilitate polarisation, but not because users are a priori prevented, by algorithmically regulated systems, to interact and engage with diverse perspectives, but because they are not incentivized by the environment to do so. Sunstein (2017) argued that the current digital environment works in a way that fosters motivated reasoning, homophily, and confirmation bias which are the actual causes of people's clustering in homogeneous groups of like-minded users, and eventually of polarisation as well. From Sunstein (2017)'s standpoint, online, users are able and free to determine their interactions, but their decision making is strongly and negatively influenced by environmental factors that foster biased reasoning processes which could be conducive to the radicalization of the positions endorsed. For this reason, Sunstein (2017) is very concerned regarding the potential anti-democratic risks caused by the digitalization of human interactions and public discourse.

In the previous section, I presented a picture coherent with the one described by Sunstein (2017). In fact, I have argued that the spread of disinformation and misinformation might be dependent on the ecological features of the current digital

environment, that makes people systematically incentivized to reason fast[9] and to rely on their priors and automatic cognitive responses. Coherently with Sunstein (2017)'s argument, I argue that individual and group decision making still represents the decisive factor in terms of information selection, sharing activities, and consequently polarizing tendences. It is true that studies show that online dynamics are largely regulated by homophily, even if there is a considerable variability among social media platforms (Cinelli et al. 2021), and that groups of like-minded users get formed very easily as soon as an issue becomes a popular topic of public discussion (Cota et al. 2019), however, it would be inappropriate to claim that results like these prove that polarisation is automatic and inevitable in the digital environment, and particularly that this is caused by algorithmically regulated systems of content recommendation.

First of all, it is important to notice that, considering a variety of topics and different political systems, there is no uniformity of effects (Gunn 2019, Cardenal et al. 2019), namely it is possible to notice that, within the digital environment, people and groups behave very differently even while discussing the same topic. In the mentioned study by Cota et al. (2019), which is focused on the political debate on a specific issue, there was no uniformity of behaviour, or sharing activities, among groups with different political positions. In particular, those with a moderately oriented positions on the topic were the ones able to reach the broadest audience, simply because they were willing to consider a more heterogeneous set of sources of information than the more radical users. In this case, moderately partisan groups did not polarise as the debate proceeded, on the contrary, for the most part maintained a moderate position that allowed them to reach a broad audience with their contents, disconfirming the idea that progressive exposure to politically aligned contents inevitably leads to polarisation.

Considering the US context, Benkler et al. (2018) have argued that the Web and social media platforms cannot be exclusively responsible for the present polarizing tendences since the North American society has reflected largely nonhomogeneous polarizing dynamics. Among the US users, those who are currently more likely to be isolated in closed clusters are far-right supporters, and as Benkler et al. (2018) showed, this was not a direct effect of the emergence of the Web but the result of four decades of progressive radicalization of part of the US society, politics, and media channels. In other

[9] To use Kahneman (2011)'s terminology.

words, if the polarizing dynamics witnessed in recent years were to be attributed to algorithmically regulated contents' recommendation systems, the polarizing pattern should have been uniform and homogenous, and this has not been the case. This variability that emerges from the mentioned studies could be related with Sunstein (2017)'s idea. Clearly, individual and group decision making varies on the basis of the exposure to different sources of information, but this influence is not deterministically biased in a specific direction. Individuals and groups are free to react differently on the basis of their inclinations and attitudes, including their propensity to fall into cognitive biases and motivated reasoning, and these differences could explain the nonuniform dynamics reported in the mentioned studies. Therefore, Sunstein (2017)'s claim appears sound and well supported.

The idea that social media and the Web inevitably determine the polarisation of people's opinion is connected to the metaphor of the 'echo chamber', now very commonly used to represent closed groups of users who are exclusively exposed to contents they would agree with, to the point that they have no interactions with users endorsing opposing views. The main risk related to echo chambers is the one envisaged by Sunstein, namely that, with time, everyone will become the member of an echo chamber. At that stage, social media users will receive contents only from like-minded sources to the point that the interaction with diverse and opposite perspectives will be fully compromised. A digital environment that is partitioned into echo chambers is very likely to lead to polarisation and jeopardise democracy. In the next subsection, I review some results that do not deny the possibility that closed groups of this kind currently exist, or that the emergence of polarizing attitudes might be facilitated within the digital environment. However, from these results that I consider, it turns out that falling into an echo chamber is not the inevitable fate of every social media user, and that polarizing attitudes are very unlikely to be the direct product of the current digital environment.

2.3.1. Opinion Segregation Is Not Inevitable

Bakshy et al. (2015) showed that Facebook's algorithm filters politically divergent contents out from users' feeds, but only for a limited share, that nonetheless might be considered not trivial (Sunstein 2017, Chap. 4), and that varies on the basis of the users' worldviews. Among US users, liberals reflect a 5% ratio between the chances

of seeing variegate contents (including divergent ones) and ideologically aligned ones, whereas for conservatives this ratio is 8%, hence reflecting a dis-homogeneous exposure. Thus, there is evidence that contents are filtered, but to what extent this determines users' segregation is a different issue. In the impactful research that has been offered by Gentzkow and Shapiro (2011), the authors showed that in terms of news "segregation", namely users' isolation within groups and networks that share only contents that are aligned with their priors, there is no evidence that the Web is a place where most of the people are ideologically segregated, on the contrary, it turned out that the online environment is less ideologically segregated than the one offline, composed of real life interactions with family members, friends and peers. More recently, Nyhan et al. (2023) showed that it is true that, in general, the vast majority of interactions on social media occur among like-minded people. Again, on Facebook, items from like-minded users are three times more frequently seen than the ones shared by opposite opinions' holders, but only a very small fraction of these interactions regard politics or consist of the exchange of actual news. If we define extreme opinion isolation as a rate of exposure to aligned contents that is higher than the 75% of the total contents present on someone's Facebook feed, then only the 20% of Facebook users are in extreme opinion isolation. As for the study by Bakshy et al. (2015), it is possible to argue that 20% of Facebook users is not a small number at all, but still these results show that the vast majority of Facebook users are not extremely isolated, hence it would be inappropriate to claim that ideological isolation is the standard way of inhabiting social media spaces. Even more explicitly, Nelson and Webster (2017) reported that people belonging to different political sides have very similar news consumption patterns on social media, and that ideologically mixed audiences, which constitute the vast majority of social media users, spend most of their time on the same, mainstream, political websites, with a resulting news consumption among conservatives and liberals that turns out to be "remarkably similar" (Nelson and Webster 2017, p. 7). In general, data show that people have centrist information diets, and that most of the people who navigate online do not care about politics at all, and while it is true that people prefer congenial contents, it is not empirically proven that they actively avoid contents that are not in line with their positions, quite the opposite (Garrett 2017, Guess et al. 2018, Nyhan et al. 2023).

There is a standstill at this point, since it appears somehow true at the same time that social media are contributing to the current socio-political polarisation and that on social media people are not, by default, politically segregated into echo chambers. In fact, the contents offered in one's social media feed are ideologically oriented and aligned with user's profiles, and hence with their priors as well, but it is also true that the use of social media and online sources makes it more likely to be exposed to counter-attitudinal information. In other words, social media facilitate those who decide to inhabit an echo chamber, but these represent a minority of users, whereas the vast majority continue inhabiting a quite heterogeneous media environment. Moreover, even in presence of a desire and decision to be part of an ideologically homogeneous community, this does not rule out at all the possibility of receiving politically divergent news or contents when online (Barbera 2020, p. 38). In support of this claim, Flaxman et al. (2016) reported important results showing that while the news articles suggested on social media are "associated with higher ideological segregation than those an individual can read by directly visiting news websites" (p. 318), it is also true that social media and online search engines are the news sources where it is more likely to discover divergent perspectives and contrary positions, independently of one's activities, and which offline could be easily avoided through an accurate selection of one's sources of information. In general, their conclusion was that "the vast majority of online consumption mimicked traditional offline reading habits, with individuals directly visiting the pages of their favorite news outlets" (Flaxman et al. 2016). This conclusion downsizes the supposed effect of social media on people's ideological segregation and suggests that polarizing tendences do not originate online.

To recap, even if it is true that digital spaces facilitate the formation of politically homogeneous, and often radical, groups, it is also true that social media and search pages are environments where it is more likely than offline to run into opposite perspectives and contents that are divergent with users' opinions. This conclusion seems to deny the idea that, on social media, people are dragged into echo chambers with like-minded users, and with no chance of getting to know people with different ideas or worldviews. Even more importantly, what emerges is that those who inhabit echo chambers end up in one of them because they seek to be part of a homogenous environment of a specific kind, or because so profoundly conditioned by cognitive dynamics, like motivated and biased reasoning,

to eventually join extremely closed groups, a possible conclusion that would support Sunstein (2017)'s perspective. Therefore, to better understand the dynamics of group formation online, and consequent polarisation, it is necessary to focus on users' behaviour as much as on algorithmic recommendation system. As Nelson and Webster (2017) very rightfully pointed out while commenting their research outcomes:

"That we found no evidence of partisan selective exposure does not mean that we believe the fears of political polarisation to be unfounded. Instead, we believe that more work should be focused on how media audiences *engage* with political news, rather than how frequently they are exposed to it." (p. 10)

This conclusion marks an important passage in this discussion. The term 'echo chamber' has been introduced to denote closed groups, but more specifically to point out a shortcoming of the current digital environment, namely the inevitable polarisation of users due to their selected exposure to specific contents. However, results show that users are very unlikely to end up in an echo chamber simply because of them interacting online or gathering their information on social media, which implies that algorithmic mechanisms of content recommendation alone cannot be imputed of a direct polarizing effect. If the problem with the formation of polarised groups online is not fully attributable to mechanisms of content recommendation, then what should the term 'echo chamber' be used for?

The empirical evidence available shows that the current mechanisms at play surely facilitate the formation of clusters of like-minded users, but against a deterministic reading, people's attitudes and decision-making processes still result decisive to determine their segregation and eventual polarisation. Therefore, it is possible that the best use of the metaphor of the 'echo chamber' is not to denote groups formed by content recommendation dynamics, but groups characterised by specific decisional patterns or cognitive mechanisms, and that are suboptimal and irrational because conducive to radicalization and polarisation. In the upcoming subsections, I will consider two different possibilities.

2.3.2. Isolation and Polarisation Are Not Necessarily Irrational

I believe that the emergence of doubts regarding the proper meaning and use of the term 'echo chamber' are primarily due to the fact that the relative literature has been very confusing from a conceptual point of view. It has been said that echo chambers represent: "join groups formed around a shared narrative" (Cinelli et al. 2021, p. 1), or digital environments where the exposure to diverse perspectives is limited, while the exposure to similar narratives is reinforced (Nyhan et al. 2023), and also a "closed system where other voices are excluded by omission" (Avin et al. 2024, p. 1). The principal source of confusion regards whether echo chambers refer to digital spaces resulting from algorithmic mechanisms that prevent the exposure to certain contents, or to particular groups of ideologically homogenous users who unite on social media and digital platforms to share specific contents. In other words, it is not clear whether echo chambers should be used to represent certain ecological features of the digital environment, or decision-making dynamics. Given the picture offered in the previous section, the latter possibility seems more promising and empirically supported. Therefore, it is worth analysing the possibility to define 'echo chambers' the groups of social media users who systematically reject the pieces of evidence and information received by sources that endorse views that are opposite or in contrast with theirs.

This focus on users' behaviour, and not on the mechanisms of content recommendation, is emblematic in Lackey (2021)'s discussion on echo chambers, where the thesis endorsed is that the lack of diversity and of independent opinion sources, that is common of most of the characterizations of echo chambers, is not problematic in itself, at least from an epistemic point of view. On the contrary, it is perfectly rational in general, Lackey argues, not to consider certain sources, like motivated or notably unreliable ones, since no benefit or accurate pieces of knowledge could be gained by them. In other words, if what ultimately characterizes echo chambers, conceived as spaces of group interaction, is the selective attitude of their members regarding the trusted sources of information, then these are not definable as inherently irrational or epistemically vicious environments. What is decisive to evaluate whether a closed group stands on epistemically problematic grounds is the resulting selection of the trusted sources of information, the kind of contents spread within, and the beliefs formed accordingly. In other words, all that matters are the decisions taken within the group, not its selective attitude. The socially and

epistemically dangerous groups are those that are based on a systematic sharing of false, misleading, radicalizing, polarizing and hateful contents, that derive from bad sources of information. Therefore, from this perspective, if within a group what circulates the most are items of disinformation, fake news and false contents, and sharing such contents is what keeps the members of the group unite, then this is an instance of a poor and problematic digital environment, which could be labelled as an 'echo chamber'. Groups of this kind reportedly exist online. Vosoughi et al. (2018) showed that the spread of false news is indeed not mainly due to algorithmic or automatic mechanisms of sharing, but it is significantly more likely to be determined by user-to-user sharing. Most of the circulating false and misleading items of information are thus not directly proposed to social media users on their feeds, but received by other users, who could eventually form an echo chamber as interactions proceed. Lackey (2021)'s argument is coherent with the reported empirical evidence. If the digital environment does not fully prevent the access to opposing views, on the contrary, people can be sure that, at some point, an ideologically divergent content will be proposed to them, it is meaningful to argue that the current polarizing attitudes might be related to users' bad choices regarding their sources of information, and the contents shared with each other. However, if Lackey's position is endorsed, it is evident that the metaphor of the 'echo chamber' turns out to be useless and redundant, because it would be used to indicate nothing more than a bad selection of sources of information.

At this point, the only option left is to bite the bullet and consider the possibility that the best use of the term 'echo chamber' is to denote groups of social media users, connected by common views and ideas, who manifest polarizing tendences. This possible use of the term 'echo chamber', and related definition, is grounded on the idea that polarisation is inherently problematic and irrational. However, there are positions that argue against this assumption. Singer et al. (2019) claimed that polarisation within a group is not necessarily irrational, since it depends in what way it occurred. The authors argued that polarisation can sometimes derive from a "coherence-based" strategy of memory management, which means that people constantly receive new pieces of evidence and select which of them to memorize on the basis of their coherence with the beliefs already endorsed. The latter process of selection and storage is not irrational per se, on the contrary is quite simple and ordinary from an epistemic point of view, even if

it can be conducive to polarisation. In particular, it would be more epistemically problematic to immediately elaborate newly received items of information in function of the reinforcement of the already endorsed beliefs, namely to reason in a motivated way. However, if the latter does not occur, a polarised state of belief, that derives from a "coherence-based" strategy (Singer et al. 2019), is not epistemically vicious.

Begby (2024) argued for a position similar to the one of Singer et al. (2019), while defending a non-ideal argument on belief's polarisation. Begby (2024) argued that the epistemic environments within which people interact are imperfect and characterized by the constant influence of value judgments and partisan worldviews. Moreover, due to their cognitive and epistemic limitations, people are dependent on others' judgment to gather information and form beliefs. In evaluating the information received, people may be conditioned by peer agreement in forming their beliefs, to the point that they eventually polarise. This process, that Begby (2024) identified with the fall into an echo chamber, is said to not be necessarily irrational (pp. 534-535). An individual might be unfortunate in terms of the epistemic environment in which she happened to live, for instance one deeply conditioned by extreme views and disinformation. Moreover, even falling into an echo chamber does not necessarily represent an epistemically vicious fate per se, since in complex and imperfect societies, where everyone's judgment is conditioned by value judgments, it is rational to privilege, to a certain extent, the opinion of those with which someone already shared ideas and beliefs. Begby (2024)'s argument on this topic is more detailed than the brief description I have reported, but there is no space to elaborate it furtherly. However, the combination of the reported arguments is sufficient to state that it is not pacific to claim that beliefs' polarisation is always and inherently irrational. Moreover, Begby (2024)'s analysis of echo chambers constitutes a convincing argument against the idea that these should be used to refer to epistemically vicious environments, as long as what characterizes them is the polarising attitudes of their members, which was the starting hypothesis of the second part of this sub-section.

To sum up, in section 2.3, I have started considering the possibility that the development of social media platforms has not only contributed to the spread of disinformation, but also to the polarisation of the political and public debate, with these two elements being deeply connected. Thus, a better understanding of the potential causal role played by the digital environment in triggering polarising dynamics could be

beneficial for an analysis of the spread of disinformation as well. Alongside the discussion on polarisation and social media, part of the literature has started using the metaphor of the 'echo chambers' to represent groups of like-minded users who are exposed almost exclusively to contents that are aligned with their priors, and who interact only with each other. However evocative, the use of this metaphor has been confusing, and it has not always been clear what problematic mechanism it is supposed to describe. In some cases, echo chambers are used to indicate clusters of like-minded users that result from algorithmically determined recommendation systems. In this case, the problematic dynamic that they are supposed to represent is that it is currently inevitable, for every social media user, to become a member of a closed group of likeminded individuals who have only access to determinate contents that they are supposed to like. This perspective does not imply or require that, within echo chambers, only polarising or harmful contents are shared. From this view, it would be possible to be part of an echo chamber of supporters of a soccer team, or vegetarian food lovers, who as such will never receive, or share, contents that include meat-based recipes for instance. The problem from this perspective is not related to the contents shared, but to the fact itself that people, basing on their priors, are prevented from received contents not aligned with the latter. Authors like Sunstein (2017) who are particularly committed to the beneficial role of diversity of opinions and open discussion in democracy, are concerned with this possibility and related scenario. However, as seen, the available evidence does not support it. Social media users are currently not completely prevented from running into contents that are not coherent with their profiles. In other words, it is very likely that those who cluster in closed groups manage to do so as a consequence of their deliberate selection of the trusted sources of information, or as the result of biased reasoning processes. For this reason, I considered the possibility that the metaphor of the 'echo chambers' should not be used to represent closed groups that originate from suboptimal algorithmic dynamics, but the combination of the latter with specific group decisional patterns, that being irrational and epistemically vicious, are conducive to polarisation and radicalization. Therefore, I considered two apparently problematic characteristics that could be inextricably associated to echo chambers, and that could help defining them in a clearer way. At first, I considered the possibility that the most peculiar, vicious practice, that is characteristic of groups that can be defined as 'echo chambers', is their selective attitude, meaning the

a priori exclusion of every content deriving from specific sources of information. However, as Lackey has argued, to be selective with one's sources is not an irrational practice per se, since it all depends on the results of the selection, namely, whether this was perfectible or not. It could be said that echo chambers should denote the groups who are bad selectors of items and sources of information, but it would be redundant to have a specific term for such a contingent and common practice. Therefore, I have considered a different possibility, namely that the vicious characteristic of echo chambers is simply their propensity to polarisation, and thus their use should simply be the one of denoting radically polarised groups that have been formed online. Clearly, this possibility holds only as long as it is possible to argue that polarisation is, in itself, inherently vicious. However, I have reported the arguments from Singer et al. (2019) and Begby (2024) who claimed that even polarising tendences are not necessarily problematic from an epistemic point of view, since in some cases it is perfectly rational to polarise. Therefore, someone with a nominalist tendency might be tempted to drop the term 'echo chamber' and recommend stopping using it within this literature. Contrary to this conclusion, I think that it is possible to offer a conceptual clarification of the term which is still grounded on the idea of polarisation. The reported arguments against the idea that polarisation is inherently irrational did not rule out the possibility that some particular forms of polarisation are indeed irrational. On the contrary, they both defined precise boundaries to determine when a polarizing attitude is the result of an irrational and epistemically problematic practice. Begby (2021; 2024) in particular has defended an important position that is relevant in this discussion, and that I will recall in the next section. I believe that an efficient definition of echo chambers is already available in the literature, and once this is adopted, with the proper clarifications and auxiliary assumptions, it would be possible to define echo chambers as groups characterised by a peculiarly vicious epistemic dynamic, namely an irrational beliefs' update mechanism.

2.4. Don't Multiply Entities Beyond Necessity

Some kind of conceptual clarification is needed with respect to echo chambers. Fortunately, I believe that the solution for this conceptual impasse is already present in the literature, and it has been offered by Nguyen (2020) who has distinguished between

'echo chambers' and 'epistemic bubbles' and has defined which peculiar epistemic dynamics the terms should be used to describe.

Before engaging in the analysis of this distinction, it is useful to notice that, in the previously mentioned work, Lackey (2021, p. 207 footnote) defined Nguyen (2020)'s definition of echo chambers as "unorthodox". I think that it is useful to start this section by claiming that there is no orthodoxy to consider or respect in this debate, since there is no ultimate and universally agreed upon conception of echo chambers, as the previous section has shown. The term "echo chamber" has been introduced to denote a peculiar, and arguably harmful, typicality of the digital environment, which due to the mentioned empirical results, needs to be reconsidered and better conceptually engineered. In the previous section, I have reported various scholars arguing for the necessity to consider users' behaviour, instead of algorithmic recommendation logic, as the decisive factor for the constitution of homogenous clusters of like-minded users online. However, for the latter to be the case, it is necessary that users have free and open access to any content that is present online and on social media. In case of digital environments where access to certain contents is prevented, users' decision making, in terms of the selection of sources of information, is a priori constrained and conditioned. This can happen independently of users' behaviour and decisions regarding information gathering, as it is the case of those countries where the access to specific sources of information is banned.

Following Nguyen (2020)'s distinction, the latter scenario is the one of an 'epistemic bubble'. Nguyen (2020, p. 143) defines an "epistemic bubble" as a "social epistemic structure which has inadequate coverage through a process of exclusion by omission." Being into a bubble means being stuck in a condition of ignorance that lasts as long as the bubble itself holds, meaning, until someone within has access to sources of information that were previously inaccessible from the bubble. On the contrary, an 'echo chamber' is defined as "an epistemic community which creates a significant disparity in trust between members and non-members." (p. 142) with the latter being convicted of *"*some epistemic demerit, such as unreliability, epistemic maliciousness, or dishonesty.*"* (p. 143). What unites the members of an echo chamber is thus not only a set of shared ideas and worldviews, but also the respect of a norm of systematic rejection of every piece of information that comes from opposing perspectives. Repeated exposure to divergent contents requires a strong commitment and the most peculiar characteristic of echo

chambers can be attributed to their mechanisms of reinforcement and radicalization of the beliefs endorsed within. According to Nguyen (2020), in an echo chamber, beliefs' reinforcement occurs via a peculiar epistemic mechanism called "evidential preemption" (Begby 2021). The idea is that, within groups where homogeneous contents are constantly shared, it is possible that certain members warn the others of the upcoming reception of counterevidence by antagonistic groups. Then, once counterevidence is actually received, the apparent predictive ability is interpreted by the members as a sign of theorical superiority, as if any possible counterevidence or counterargument had already been taken into account within the group. The effect of the latter mechanism is twofold. On one side, preemption works by "neutralization", meaning that it neutralizes a priori the epistemic significance of any possible counterevidence, present or future, and contributes to the endorsed discredit of the sources from which the counterevidence comes (Begby 2021). Therefore, following Nguyen's distinction, epistemic bubbles are digital spaces that are caused by environmental reasons, and not by their members' behaviour. Here, people are not exposed to certain contents because of an external limitation the is independent from their behaviour. The resulting epistemic condition is one of mere ignorance that could be fixed by simple exposition to the unknown contents that are left outside the bubble. This case seems to be the one previously envisaged by authors like Sunstein (2017), whose worry is that social media, on the long run, can determine the constitution of groups where access to determinate contents is totally prevented, a risk that is worth considering but that for the moment appears under control, and that it is better represented by the idea of epistemic bubble than from that of echo chambers. Indeed, on the contrary, from Nguyen (2020)'s perspective, echo chambers are digital media-based groups, the formation of which is minimally related with environmental factors and ultimately determined by members' behaviour. As Lackey (2021) has rightfully argued, there are various ways to be part of a homogeneous group and to select the trusted sources of information, and in the case of echo chambers, Nguyen (2020) argued, the latter is the result of an epistemic control based on discredit and distrust of any source outside the members of the group.

There is a very interesting aspect that is worth highlighting at this point. Committing to Nguyen (2020)'s distinction implies the acceptance of the idea that the connection between the members of an echo chamber is maintained and reinforced through the mechanism of evidential preemption. As seen, for the latter to occur, exposure

to divergent views is a sufficient and necessary condition to trigger certain epistemic dynamics, that I describe in the next section. Therefore, from this analysis it turns out that, following Nguyen's distinction and definition, epistemic bubbles and echo chambers result to be not only distinct phenomena, but completely antithetic ones. The former are the result of a lack of exposure to divergent contents, which is independent of the behaviour and will of those stuck into the bubble, while the latter are dependent on the exposure to opposing contents to both get formed and constantly reinforced. Echo chambers are thus not peculiar digital spaces, but peculiar group decisional patterns, instantiated in normal conditions of exposure to variegate contents. In the next section, I argue that the term echo chambers, in Nguyen's conceptualization, should be used to denote groups characterized by specific epistemic vices and irrational decisional patterns, that I describe.

2.5. Echo Chambers are Irrational Decisional Patterns

From Nguyen's perspective, echo chambers get formed when the members of a group start to systematically discredit certain sources of information and reject their contents as a priori untrustworthy or fake. However, the necessary and most peculiar epistemic dynamic that characterizes certain groups as echo chambers is another one, and from Nguyen (2020)'s standpoint, it is related with their internal mechanisms of belief reinforcement, namely evidential preemption. If this idea of echo chambers is accepted, and it is stated that what ultimately distinguishes them from other forms of group formation is their mechanism of beliefs' reinforcement, through evidential preemption, then it is possible to argue that inhabiting an echo chamber implies the insurgence of an irrational state of belief, or in other words, the systematic violation of a requirement of rationality. To see why this is the case, it is necessary to consider what makes evidential preemption irrational, or as Begby (2021) clarified, under what circumstances evidential preemption is epistemically vicious. Begby (2021) argued that, when a group receives counterevidence of an endorsed belief, the previously described mechanism of preemption by neutralization (see 2.4) can only influence the assignments of credibility attributed to the source of the received counterevidence, and not to the belief that it was meant to deny (call this belief *p*). In other words, the reception of counterevidence by a distrusted sources can never constitute, in itself, evidence in support of *p*, namely

determine an increase in the belief in *p*. Consider *p* as a proposition that is endorsed within a group. Following Nguyen (2020)'s idea of evidential preemption, it is possible to state that what is necessary for this group to be defined as an echo chamber is precisely that the reception of counterevidence against *p* is interpreted as evidence for *p,* namely, that it determines an increase in the belief in *p*. The latter is not an epistemically acceptable dynamic, since it represents an irrational form of evidential preemption according to Begby (2021). Therefore, we have now a definition of echo chambers that clarifies the epistemically vicious practice that ultimately characterises them, and this is related to the fact that when counterevidence or opposing views are encountered, this event itself, independently of the actual content received, constitutes evidence of the robustness of the theory endorsed within the chamber. The resulting non-falsifiability of the views shared within an echo chamber is then interpreted as a sort of epistemic strength, that makes immune to any possible counterevidence coming from the outside. While intuitively promising, there is still limited empirical evidence that an epistemic mechanism of this kind actually takes place. For instance, Quattrociocchi et al. (2016), who analysed the reception of debunking attempts within conspiratorial groups, offered results in support of evidential preemption. Moreover, from this work, it seemed that evidential preemption follows a linear increasing trend, meaning that the degree of belief-reinforcement is proportional to the number of pieces of counterevidence and opposing views encountered. However, further evidence is required to argue for the occurrence of the latter "disagreement-reinforcement mechanism", as Nguyen (2020) called it. Nonetheless, from a purely conceptual point of view, it is possible to state that this particular form of irrational evidential preemption could represent the most characterizing feature of echo chambers, and thus the ground of their definition and related use.

At this point, it is apt to clarify of what epistemic vice are culpable the members of echo chambers. Begby (2021) argued that evidential preemption is epistemically vicious when, within a group, after the reception of counterevidence against a common belief *p,* the fact of this reception alone determines an increase in the belief in *p.* In the rest of this section, I will elaborate further on this and determine what specific epistemic violation is instantiated in the latter dynamic. In particular, I consider the requirements that rational agents ought to respect when they become aware of something new, including counterevidence coming from sources endorsing opposing views and

perspectives. In what follows, I consider in particular Bradley (2017)'s reverse Bayesian account of rational beliefs' update.

A way to represent people's states of awareness is as divided in "prospects", which possible states of the world over which people form attitudes, namely subjective assignment of desirability and probability, which are progressively updated, expanded or reduced, as new information becomes available. Following Bradley (2017)'s account, it is possible to see that a domain of awareness can be expanded: *"*as a result of a refinement of the possibilities that she entertains*"* (p. 407). This refinement could depend on the discovery of a new level of relational complexity, including new possible features of the already considered prospects for instance. Otherwise, an individual can become aware of *"*one or more possibilities disjoint from those she entertained initially*"* (Bradley 2017, p. 407), meaning a completely new prospect, whose occurrence in real life would be incompatible with any of the other ones. Consider the following example to clarify this distinction. A person, who has been invited to a dinner party, is waiting to know in which of two restaurants the dinner is going to take place. This person could refine her state of awareness by, for instance, having access to the menu of the two restaurants. Otherwise, becoming aware of the fact that the restaurants considered by the host are actually three, and becoming aware of the name of the third restaurant, would represent an expansion of her state of awareness, since the dinner party can only take place in one restaurant, hence the three possible prospects, each associated to one restaurant, are incompatible. Consider the latter case. Bradley's claim is that: "[…] in cases where we become aware of prospects that are inconsistent with those that we previously took into consideration is that we should extend our relational attitudes to the new set in such a way as to conserve all prior relational beliefs and preferences" (Bradley 2017, p. 408). In other words, the claim is that, when a rational agent becomes aware of a more precise description of the prospects that compose her state of awareness, this should not imply any change of attitudes regarding the prospects, as they were known and conceived in the agent's pre-refinement domain of awareness. In the example offered, becoming aware of the fact that a third restaurant can be selected for the evening should not determine an attitudinal change regarding the other two restaurants, meaning assigning new and different subjective values of desirability and probability. Clearly, when a state of awareness is expanded and a new assignment of probability is made, since probability assignment must always add

up to 1, the probabilities assigned to the previously considered prospects will numerically change, but not the relative proportions. Therefore, when an individual becomes aware of an item of information that is incompatible with some of the prospects previously endorsed, it would be irrational to change the probability and desirability assigned to any of the latter.

Interestingly, this requirement of rationality is compatible and coherent with Begby (2021)'s take on the rationality of evidential preemption. Recall that Begby (2021) argued that it would be irrational, being committed to a proposition *p,* to consider the reception of counterevidence against *p*, by a discredited source, as evidence in support of *p,* since it could only constitute evidence for the testimonial reliability of the source. This claim is implied in Bradley's definition of a rational refinement of an agent's domain of awareness. Becoming aware of a proposition *q* that refines one's state of awareness, independently of the assigned credibility and desirability, cannot modify, in a rational way, the proportions, in terms of assigned credibility and desirability, among the other known propositions, and so it cannot rationally determine an increase in belief with respect to a previously endorsed proposition.

Consider a proposition *q* whose content is opposite to the one that exemplifies the worldview of an echo chamber, which is denoted with *p*. Those who reject *q* due to their doxastic commitment to *p*, come to know both *q* and the fact that they received it from a distrusted source. If, following the latter fact, namely the reception of a content opposite to *p*, the members of the echo chamber reinforce their doxastic commitment to *p*, namely assign higher probabilities to the truth of *p*, they violate both Begby (2021)'s requirement for a rational evidential preemption and Bradley's requirement for a rationally refined state of awareness.

Therefore, assuming Nguyen's conceptualization of echo chambers, it is possible to claim that the epistemic dynamic that defines echo chambers regards the process of beliefs' reinforcement that follows from the exposure to counterevidence. More precisely, echo chambers could be defined as groups of like-minded users who have access to all the sources of information available in their environment, and whose exposure to opposite views and counterevidence determines the progressive reinforcement of their commitment to a set of shared beliefs, in a way that is epistemically vicious and irrational.

Therefore, if this configuration of Nguyen's conceptualization of echo chambers is adopted as the standard in this literature, the resulting benefit would be the clarification of what this metaphor should be used for, and also the definition of the epistemic dynamic that characterizes echo chambers, that are nothing more than collective decisional patterns that epistemically problematic due to their dependence on an irrational configuration of the mechanism of evidential preemption, that consists in a violation of the requirements of rationality for the refinement[10] of one's state of belief. In other words, for a cluster of like-minded users, the price to pay for them to turn into an echo chamber is to systematically endorse an irrational state of awareness, at both individual and group level. Even in this preliminary configuration, this definition allows to clarify what is the kind of stubbornness whose formation must be prevented online and on social media. In a context that facilitates the formation of groups of like-minded users, it is crucial to avoid the formation of echo chambers, since it is clear that, according to the offered definition, given the epistemic dynamic that characterizes echo chambers, it is not only impossible to debunk the beliefs endorsed within, but it is counterproductive. This is clearly a problematic scenario considering the case of an echo chamber that is founded on scientific disinformation. In fact, every attempt to debunk the beliefs that characterize it would end up reinforcing the commitment of their members, with the risk of radicalizing a cluster of spreaders of disinformation regarding delicate topics. The severity of this risk motivates the concern expressed by Sunstein (2017), which nonetheless was founded on an unclear conceptualization of the notion of echo chambers. Given a clearer definition of the latter, it is possible to picture in a clearer way the possible risks related to users' segregation online, and to stress the importance that social media platforms are actively committed to guarantee a heterogeneous exposure to diverse ideas and opinions on their platforms, and to contrast the proliferation of disinformation, as explicitly required by the European Code of Practice on Disinformation.

Conclusion of Chapter 2

In the first part of this chapter, my aim was to describe some of the reasons that have determined the current spread of disinformation, online and on social media. In

[10] Following Bradley (2017)'s terminology.

section 2.1, I have considered some historical factors that have contributed to the current normative scenario regarding information management, and social media companies' responsibilities to contrast the proliferation of disinformation. Whereas, in section 2.2, and related subsections, I considered some cognitive mechanisms, including biased reasoning processes, that facilitate the proliferation of inaccurate contents. Then, from section 2.3 to the end of the chapter, I have discussed the topic of online segregation, namely the formation of groups of ideologically aligned users, and echo chambers. I have defended two main points in this chapter. The first is that the spread of disinformation is not fully imputable to the current mechanisms of content recommendation. Users' decision making, in terms of which sources to trust and what contents to share, is still largely free and dependent on people's goals and states of belief, hence, on both epistemic and non-epistemic factors. However, it is true that the current digital environment, and particularly social media platforms, are both conducive to biased reasoning processes, and to a form of cognitive laziness that disincentivizes the users from analytic thinking. Moreover, in a context that prizes virality, contents that include emotionally charged expressions, and that are designed to require low cognitive effort to be elaborated, are more likely to become viral and be widely spread, and disinformative contents often have both these features. Regarding information segregation, I have argued that it would be very dangerous, for every democratic system, if social media users isolate in clusters of like-minded users, thus preventing themselves from running into diverse perspectives and ideas. However, currently this does not seem to be the case, since the vast majority of social media users have variegated information diets, and the polarisation witnessed online regarding certain topics, particularly climate change in the US context, is very likely to be the reflection of a long-lasting polarising tendency that started offline. Nonetheless, online segregation is an issue to consider, and for this reason, I have claimed that it is urgent to make some conceptual order within this debate. Particularly, I have argued that the current use of the term 'echo chambers' is confusing. Starting from Nguyen (2020)'s conceptualization, I have argued that the metaphor of the echo chambers should be used to denote online-based groups that are characterized by a particular irrational mechanism of beliefs' reinforcement, that I have described in section 2.5.

3. Disinformation Influences Social Identity Formation Processes

In this chapter, I defend two main arguments. The first is that, even if motivated reasoning is one of the best explanations for the spread of disinformation online, a research approach that only considers the compatibility between certain disinformative contents and determinate group identities as the driver of the spread of disinformation would be incomplete. The main risk would be to ignore the fact that sharing disinformation is not only related to social identity maintenance and reinforcement, but also to social identity formation. I argue that being exposed to disinformation, and prone to share it with others, has consequences on people's state of belief and on the development and formation of their social identities, namely on their beliefs about themselves as socially situated agents, and about the positioning of their groups in the broader social hierarchy. Following this claim, in the second part of the chapter, I argue that to better understand the reasons behind people's decision to share disinformation, it would be helpful to consider the action of sharing disinformation as a form of intersubjective cooperation. More precisely, the chapter is structured as follows.

In the first three sections, I elaborate a critique to the literature on disinformation. In particular, I argue that this has largely focused on the possibility that the spread of disinformation is related to processes of motivated reasoning and identity protection, while ignoring the possible role that disinformation has in processes of identity formation. This is due to two interrelated factors. The first, which I consider in section 3.2, is that in this literature it is popular the assumption that people form their beliefs and identities outside of the digital environment, and then, online, they only defend and reinforce their commitment, an idea that I consider no longer tenable in this historical moment. The second, that I consider in section 3.3, is that a too static notion of social identity, i.e. one that presents social identities as sets of beliefs that rarely change, is diffused in the literature and limits the possibility of recognizing the processes of social conflict and identity formation that are involved in the phenomenon of disinformation. To show a concrete example of how disinformation can influence social identity formation processes, in subsection 3.3.1, I consider the analysis offered by Oysterman and Dawson (2020) of the disinformation techniques used during the Brexit campaign, in UK. In

section 3.4, I argue in favour of the idea that disinformation is inherently conceived to convey not only specific contents, but narratives and worldviews. Then, as I have done in the previous section, to offer a concrete example in support of the claim defended in section 3.4, in sub-section 3.4.1, I consider a concrete example, the one of some far-right oriented, disinformative YouTube channels, to show how a network of disinformation disseminators can convey a specific narrative through their contents. In the conclusive section, 3.5, I argue that it is possible to interpret the action of sharing disinformation as a form of cooperation. In particular, for certain individuals and groups, to share disinformation is primarily a way to contribute to the proliferation of the narrative that lies beneath a disinformative content, more than to the proliferation of the content itself, thus is a way to cooperate with its creators. In support of this claim, I report the results of studies in the social psychology of cooperation which show that the best predictors of cooperation are social motivations. In particular, people cooperate when they feel to be valuable members of a respected community.

My claim is that the rethoric and narrative endorsed by groups of disinformation disseminators, such as the YouTube-based network considered in sub-section 3.4.1, are designed to provide a social identity to people who feel lost, unfairly treated, and uncertain about their social status. Disinformation creators aim at creating a sense of belonging in the members of their audience, to make them willing to cooperate, and to cooperate, in this particular case, consists in sharing their contents. Then, progressively, as sharing disinformative contents becomes a routine, people's individual and social identities are progressively influenced, and eventually modified, by this activity.

3.1. Motivated Reasoning and Disinformation

In the first part of Chapter 2, I have reported results showing that the current spread of disinformation is due, in part, to the effect of automatic cognitive mechanisms and biased reasoning processes. However, considering scientific disinformation, it could be argued that other contingent factors play a role in the acceptance of false and misleading contents. Education, scientific literacy and quantitative reasoning could be mentioned as the most determinant factors for people's exposure to disinformation, and for the simple reason that the more a social media user is educated and competent in

quantitative and scientific reasoning, the less likely she should be to accept scientific disinformation as true and accurate. However, this is not always the case.

Actually, with respect to determinate scientific issues, education and scientific literacy can even be correlated with higher rates of rejection of scientific claims. The latter effect has been registered in studies concerning climate change. Conservatives are reportedly more likely to reject anthropogenic climate change (Hornsey et al. 2018), particularly in the US context (McCright and Dunlap 2011), and not support mitigation policies (Ziegler 2017). Empirical results showed that, among conservatives, the higher educated ones manifest the more radical distrusting attitudes toward climate science, proving that not only education and scientific literacy do not mitigate the effects of partisanship, but could even reinforce it (Kahan et al. 2012, Kahan 2015, Uscinki et al. 2017). Results of this kind allow to make a general and useful point. Define polarisation like in Chapter 1, namely as the occurrence of divergent updates of states of beliefs, within the members of a group, following the common exposure to the same informative content. Instances of polarisation, within a group, are not necessarily caused by misperceptions or wrongful uptakes, which could result from cognitive shortcomings for instance. Different reactions, to the reception of the same piece of information, could also be due to issues of trust regarding the source of information or to interested and motivated interpretations of the received contents (Nyhan 2021). In the Chapter 4, I will consider the possibility that the spread of scientific disinformation is fostered, and partially caused, by the pervasive presence of a specific form of distrust towards the scientific institutions. Instead, in the first part of this chapter, my aim is to report that many research efforts have pointed out the role that motivated interpretations of information have had, and still have, in the proliferation of disinformation.

If the digital environment is a particularly fertile soil for the emergence of certain cognitive biases, it is also possible that online it is easier for people to reason and elaborate informative contents in a motivated way, namely, in the now classic formulation by Kunda (1990, p. 480): “motivation may affect reasoning through reliance on a biased set of cognitive processes: strategies for accessing, constructing, and evaluating beliefs.” Motivated reasoning, in fact, can render an individual, or group, more exposed to certain cognitive biases, particularly to confirmation bias. Empirical evidence shows that the proliferation of misinformation and disinformation is facilitated by the fact that people

tend to share and accept contents that confirm what they already believe, even false and misleading ones (Del Vicario et al. 2016). This is due to the fact that contents that confirm people's priors are elaborated with poor cognitive effort and attention. The latter effect has been detected with respect to misinformation about climate change, whose proliferation resulted to be facilitated by confirmation bias (Zhou and Shen 2021). Moreover, confirmation bias can also influence the evaluation of sources of information and determine the rejection of contents coming from sources that are not aligned with one's priors, including fact checkers and debunkers (Hameleers and van der Meer 2020).

However, the effect and significance of motivated reasoning is more subtle and pervasive than the cognitive attractiveness of a content that confirms what is already believed. It is also distinct from the mere influence of previously endorsed ideas, worldviews and ideologies in the evaluation of newly acquired contents. To reason in a motivated way means to accept or reject contents for reasons that are independent from knowledge acquisition or belief coherence. Instead, motivated reasoning is related to dynamics of identity and group membership maintenance, reinforcement, and support, hence it is related to non-epistemic factors and motivations.

The literature on the effects of motivated reasoning on disinformation has focused mostly on politically determined motivated reasoning, and rightfully so. Empirical evidence reported that people manifest divergent responses to the reception of the same informative contents, in a way that is coherent with their political and group identities (Schaffner and Roche 2017, Hameleers 2023). Moreover, Osmundsen et al. (2021) showed that partisanship, unlike ignorance or low intuitive abilities and digital literacy, is also a great predictor of sharing tendences, since, in their study, the contents that were politically aligned with the ideas of the users resulted more likely to be shared. Similarly, Pretus et al. (2023) showed that far-right voters from Spain and USA were very likely to share false and misleading contents when these were associated with core identitarian values. Likewise, Rathje et al. (2023) reported that, even if incentives to be accurate can efficiently improve truth discernment, the motivation to share contents that are liked by members of one's social group outmatches any consideration of accuracy and trustworthiness with respect to the shared contents. Coherently with this picture, Pereira et al. (2021) reported results showing that people express a higher propensity to believe, and also a higher willingness to share, contents and news that are coherent with their

political affiliation, or that discredit their opponents. Moreover, after dividing the research population between conservatives and democrats, the former manifested a higher propensity to share more contents in general, and particularly, to share contents within homophilic groups. This result has been confirmed also by Boutyline and Willer (2016) who in their analysis of 260,000 Twitter users showed that conservatives tend to engage in more homogeneous online networks than liberals. Possibly, it is partially due to this propensity to share more contents in general, and to this tendency to communicate within homophilic groups, that conservatives resulted more likely to share inaccurate news, and also more prone than democrats to share contents that contrast the ideas of the opponent groups (Pereira et al. 2021).

It is possible to extrapolate interesting and general considerations from all these results. In a rich overview on the role of motivated reasoning in the spread of false and misleading contents, Galeotti (2020) claimed that, given the pervasive influence of the motivational component in the elaboration of the contents found online, large numbers of social media users are not driven by the attempt to be accurate in the selection of the contents to share. What ultimately seems to condition sharing activities are users' political ideas and group identities that might work as "epistemic selectors" (Galeotti 2020, p. 72). From this perspective, it is possible to interpret the propensity to share certain contents that discredit opposing groups as a way to express, and to let the other users know, one's group membership. There is evidence of this instrumental and communicative use also with respect to false and misleading contents, which are shared not only to discredit out-groups, but through this, to express one's group belonging (Van der Linden et al. 2020, McLoughlin et al. 2024, Brady et al. 2020, Crockett 2017, Rathje et al. 2021). I have already mentioned the fact that, particularly among conservatives, there is a higher propensity to share false contents (Jost et al. 2018, Roozenbeek et al. 2020) and an interesting aspect pointed out by Pereira et al. (2021) is that it is partisanship, more than ideology, that triggers motivated reasoning processes within conservative groups, as reflected also by Pretus et al. (2023), and by Swire-Thompson et al. (2019)'s work on the effect of partisanship over analytic thinking among the supporters of the Republican Party (US). In the attempt to explain these results, Pereira et al. (2021) argued that the strong effect played by partisanship is due to the fact that political parties resemble social groups and, for this reason, trigger the social identity protective mechanisms that are conducive

to motivated reasoning. If the latter is the case, the reported results might constitute evidence of the fact that people are strongly influenced by group membership in the evaluation of the information encountered. Thus, in the digital environment, the systematic acceptance and sharing of certain contents might be related with an attempt to be consistent with the ideas shared within one's social group. To return to the topic of scientific disinformation, it is plausible that the current spread of false contents regarding scientific topics, and the apparent reluctance in accepting scientific recommendations, might be related with the fact that certain scientific claims are incompatible with the ideas and worldviews endorsed within determinate social groups, and that their explicit rejection is made public to signal and reinforce one's group membership and commitment. There are results that support this hypothesis.

Lewandowsky and Oberauer (2021) reported results showing that nationalism, religiosity and free market are negatively correlated with the acceptance of climate and vaccines' science[11]. To test whether the collected results were indicators of motivated reasoning, the authors measured the correlation between the mentioned variables (free-market, nationalism, and religiosity) with the acceptance of another scientific topic, the theory of the heritability of intelligence. The latter had been selected in the attempt to find a scientific claim that represented no threat for religious, nationalist and free-market identities, neither in its essential significance, nor in its potential socio-political implications. Coherently with the authors' expectations, results showed no significant correlation with the mentioned variables, in support of the hypothesis that, even if certain political groups are more likely to manifest generic anti-scientific attitudes, it is the scientific topic under dispute that really matters for attitudes formation, which is conditioned by ideology and worldviews, coherently with the motivated reasoning hypothesis. Therefore, the results reported by Lewandowsky and Oberauer (2021) showed that the conservatives who reject climate science, or vaccines' safety, do not necessarily endorse a generic anti-scientific worldview, but reject specific scientific claims, either due to their incompatibility with the endorsed worldviews, or as a way to manifest their group membership and loyalty.

[11] These correlations were all very significant ($p < 0.001$).

3.2. Still, Not the Real World

At this point, the fact that the proliferation of disinformation is caused by a combination of epistemic and non-epistemic factors results evident. Focusing on the latter, in Chapter 2, I have highlighted that the communicative features of disinformation are part of the cause of its broad and pervasive presence online. Highly emotional, salient, and easy to elaborate contents are believed to elicit automatic and biased cognitive responses that might determine their acritical sharing, due to their coherence with users' states of belief, as in the case of confirmation bias, or due to ecological factors, like being disproportionately exposed to determinate contents, as in the case of illusory truth effect. At the same time, in the previous section, I have a reported various results that support the hypothesis that motivated reasoning is a promising explanation for the spread of disinformation. I have already stressed multiple times the importance of keeping the decision to share a certain disinformative content separate from the assumption that the sharer is doxastically committed to that content. To share a content, particularly on social media, is an action of its own, which could be determined by someone believing in the truth of the shared content, but that could also be totally detached by considerations of this kind, hence independent of epistemic reasons. Sharing a content, even when blatantly false, can be a way to manifest one's support for a political group or a way to signal someone's endorsement of a political position, worldview, or ideology, and again independently of whether the specific, shared content is actually believed or not. Motivated reasoning and the efficacy of emotive language can also work in synergy to foster the spread of false and misleading contents. In fact, the ability of certain contents to elicit anger and moral outrage incentivizes the non-epistemic reasons for disinformation and misinformation sharing, instead of the epistemic ones, including the desire to signal one's position and group membership through sharing activities on a digital platform (McLoughlin et al. 2024). As it seems at this point, the literature has been able to offer a relevant set of possible reasons behind the current spread of disinformation. However, there is an aspect that it is urgent to point out.

Even in their clear differences, the research approaches and explanations considered so far all seem to endorse two main assumptions. On the one hand, scholars assume that the most severe consequences of the spread of disinformation are related to the proliferation of risky behaviours, like not following sanitary recommendations during

a pandemic, or the pollution of the public debate online, which could exacerbate the polarisation of the society. These short-term issues are broadly considered to be the most severe risks of the spread of disinformation. On the other hand, in close connection with the previous point, the idea that seems to ground the current literature on disinformation is that the behaviour and attitudes that people manifest online are just the reflection, in the digital world, of the ones endorsed in the life offline, which remains the "real" one in a certain sense. The idea is that people form their personal and social identities outside of the Web and then interact online in a way that is mediated and determined by the pre-existing features that fully characterize them as individuals. The spread of false and misleading information too is then studied as an issue that regards fully formed individuals, whose different attitudes towards different pieces of disinformation are mediated by their beliefs and worldviews. Therefore, it is true that people are identity protective, but the implicit answer to the question of what identity they are protecting by engaging in motivated reasoning, what group membership they are manifesting, and what positions they are supporting, seems to be that they are protecting, claiming and supporting identities and positions formed during the interactions in the real, physical world. In the following sections, I argue that this perspective is incomplete and blamelessly myopic. Moreover, I claim that alongside the risk of polarisation and proliferation of dangerous conducts, further risks should be considered as research priorities in this literature.

In this chapter, the main argument that I defend is that, in a world that is becoming more digital by the day, there is an urgent need to consider the possibility that the presence of disinformation is becoming an endemic feature of the online environment, and an influential factor also, and perhaps above all, with respect to users' social identity *formation* processes, more so than their maintenance and reinforcement. Common sense suggests that the digital environment is destined to consolidate its position as the space where many of the activities that are relevant in the formation and definition of one's identity will take place in the near future. In the digital world, we study, work, make friends, know romantic partners, discover heritage and make sense of our cultural history, and this is relevant for our identity formation processes. The systems that made this revolution possible are related to communication technologies, and being aware of the effect that these have on people's social and individual identity formation is vital.

Following a similar argumentative line, Floridi (2011) explicitly argued that the communication technologies that have been recently implemented are not only able to influence the development of the users' identities and selves, an effect that qualifies them as "technologies of the self" but are actually the "most powerful technologies of the selves to which we have ever been exposed" (p. 561). If the potential that an environment structured and based on these new communication technologies, like the digital one, has in influencing identities' formation processes is massive, as Floridi (2011) seems to suggest (see also Floridi 2013, Durante 2017), then I argue that it is more than relevant to consider the problem of disinformation also from this perspective, since it is through these technological systems that it proliferates for the most part. In particular, I argue that the current global dynamics offer a scenario that depicts the spread and presence of disinformation as very likely to be an endemic component of the digital environment also on the long term. One reason is that, in this historical moment, the technological development is such that legislators and designers of regulatory interventions struggle in keeping the pace with technological advancements, which the creators of disinformation are notoriously able to exploit. Another factor to consider is that some political forces that have made a consistent use of disinformation in their communication and to pursue political goals, are currently in a position of power, with the re-election of President Trump, in November 2024, representing the emblematic symbol of this. It is therefore reasonable to expect that political figures of this kind will not prioritize efforts to limit the spread of disinformation online[12]. I argue that this is a worrisome scenario for all the reasons mentioned in section 1.1, and also for its possible negative effects on the processes of social and personal identities formation that take place in the digital environment, where disinformation proliferates the most. Consider, for instance, the discussed negative influence that disinformation has on people's memory. It is possible

[12] In support of this political reading, on January 7th 2025, a few days before the official nomination of Donald Trump as the new President of the United States, the CEO of Meta, Mark Zuckenberg, has announced that his company will get rid of the service of independent fact checkers in all their social media platforms (Fabebook, Instagram and Threads) claiming that this decision is motivated by the goal of preserving freedom of speech and limiting the censorship that third party fact checkers could be responsible for. Fact-checking on Meta will be appointed to the users, as it already happens on X, the social media owned by the principal supporter of the newly elected President Trump, namely Elon Musk. For this reason, this decision by Zuckenberg has been interpreted as mainly political, and of a kind that could cause an increase the spread of disinformation in its platforms due to the deterioration of the quality of fact checking, as Pennycook (2025) has argued.

that this negative influence will not be limited to the accuracy of the recalled past events (Greenspan and Loftus 2020, Loftus and Klemfuss 2023), but also to the recollection and elaboration of important historical events, which might be relevant for the definition of oneself as member of a specific community or social group, as already envisaged by Van Proijen and Douglas (2017) in their study of the possible effects that the proliferation of conspiracy theories could have on the formation of people's historical memory. In other words, one of the most pressing issues to address is the possibility that those who are highly exposed to items of disinformation, and who heavily share them on social media, for whatever reason, can be deeply influenced by these daily activities, with consequences on their conceptualization of themselves as persons, citizens, and members of a collective. With time, this process might also permanently compromise people's attitudes towards concepts like truth, knowledge and accuracy. In the rest of this chapter, I claim that the deliberate sharing of disinformation online is a kind of social action that majorly impacts in the definition of someone's social identity, and I argue that to see this in a clear way it is necessary to accomplish two tasks. First, reconsider the approach to social identity theory that the literature on disinformation has had so far. I will argue for a dynamic conception of social identity formation, against the static one which is currently popular in the literature[13]. Second, I will argue that it is fruitful to consider the action of sharing disinformation from an unusual perspective, namely as an instance of a cooperative action, to unveil further possible reasons behind its sharing.

3.3. Social Identities Are Not Fixed

It is a widely shared assumption that people are identity protective, and for this reason prone to motivated reasoning. To support this claim, it is common to refer to social identity theory, since this has offered various results that confirmed the efficacy of group belonging, and social interactions, in conditioning people's decision making and behaviour. However, I believe that the conceptualization of the idea of social identity that has become popular in the literature on disinformation is inappropriately static. More precisely, social identity has been conceived as a set of beliefs that derive from contingent factors like demographics, political and religious affiliation, education, and other aspects

[13] Galeotti (2020, p. 73) raised a similar concern, but without elaborating it furtherly, as I try in this chapter.

that matter in the process of self-determination as a socially situated individual, and that have often been depicted as almost immutable. The most emblematic example of this conceptualization of social identity is implicit in the proposal made by Hornsey and Fielding (2017) to contrast the spread of false and misleading contents. The authors' idea is that, to contrast the negative effects of motivated reasoning, including the acceptance of false claims, it is necessary to individuate the underlying beliefs that trigger the motivated reasoning and design a communicative strategy that "aligns with that motivation" (p. 469). The idea is, for instance, to communicate with those who are prone to reject scientific claims, for identity protective reasons, in a way that limits their perceived contrast and distance with the scientific institutions, and in general with the providers of the rejected information. I argue that this approach is grounded on two very unstable premises.

First, in highly polarised contexts, or regarding highly polarised issues, it is relatively easy to individuate the beliefs that are actually responsible for motivated reasoning. However, this is not always the case, since certain topics become polarising with time and not always in a predictable way. Moreover, during the discussion on echo chambers, I have reported results showing that on average social media users reflect moderate behaviours and are exposed to variegate contents (see 2.3.2), and thus it is very difficult to determine what contents and conversely what beliefs or attitudes could have determined a particular response or reaction.

Second, in line with the general claim made before, this kind of proposal is meaningful only given the assumption that people possess clearly defined social identities, that they use to orientate and catalogue the information they come across, and that these are stable, meaning very unlikely to vary. I claim that social identity theory does not justify this static view of social identity, since this has been primarily conceived as a dynamic concept. In their founding research on this topic, Turner and Tajfel (1986) emblematically defined social categorization as a "system of *orientation* and self-reference" (my emphasis, p. 283). The discovery and definition of the self is a continuous process that is functional to a progressive understanding of one's position in the society. Self-categorization, in other words, is foremost a comparative process. People make sense of themselves first by defining their own position within their social group, and secondly by understanding the status of their social group within the broader society. Moreover,

one of the main assumptions of social identity theory is that people strive both for authenticity and external recognition, which implies that, once aware of the low status ascribed to their social group, people might decide to leave it and aim for a better social positioning. The latter attempt inherently requires the possibility to engage in a process of group disidentification before striving for a new social identity, a process that requires time, effort, and that is not possible for everyone. This is why it is meaningful to think of social identity formation also as something that comes in degrees, as Huddy (2002) has argued.

Following Turner and Tajfel (1986)'s work, social identity theory shows that people do constantly make comparative considerations in terms of their own social positioning and regarding the external evaluation of their own social group, and that in presence of perceived low social status, people develop a related feeling of low self-esteem, which either causes the attempt to disidentify and change social group, or when the latter is not possible, change the parameters of confrontation with the higher ranked groups in order to not feel socially deprived any longer. All these evaluations and changes in terms of group belonging are possible only if social identity is conceived as something that is always questioned, challenged, and ready to be changed. Evidence of the aptness of a dynamic conception of social identity comes from the work of Egan (2020). In this work, Egan (2020) explicitly argued that it is crucial not to consider social identities as sets of never-changing beliefs, and reported results showing that, in the US society, there are citizens who, over a four-years period, have changed their self-identification as members of certain demographic categories and social groups in ways that are consistent with their political affiliation. In other words, for a non-trivial number of US citizens, the relevance of partisan affiliation has been such that it has determined a change in their identification as members of different groups and categories, namely it has progressively caused a change in their social identities, and also in a relatively short period of time. Therefore, social identity formation and maintenance appear to be a very fluid and complex process, whose complexity becomes even more evident when considering the emergence and the role played by social conflict.

I believe that considering the latter in this context is meaningful because, in various cases, the spread of disinformation is related to a dynamic of inter-group conflict. I have reported results showing that very often people share pieces of disinformation

because these include false and misleading contents that discredit a specific group or category[14], and consequently allow to signal one's group belonging (see 2.2). To consider these results in combination with the literature on social conflicts enables to enlarge the set of the possible explanations for disinformation sharing.

It is true that social identity theory offers evidence in support of the idea that group belonging, as soon as it is made salient, fosters inter-group favouritism and identity-based group and social conflicts (Tajfel 1970, Tajfel and Turner 1979) but there are various factors that might determine the emergence of social conflicts, and even more, that could condition the outcomes following from those. One crucial factor to consider when dealing with social conflict is the attitude of the groups involved with respect to the status quo of the society. When the legitimacy of the current social hierarchies is questioned, due to the lower ranked social groups feeling negatively privileged and deprived, and the higher ranked ones feeling in decline, social conflict spreads because the former aim at a higher social position, and the latter try to hold their status tight. This picture had been depicted and discussed, in the early days of the 20th Century, by Georg Simmel (1908), and later reconsidered and re-elaborated in particular by Lewis Coser (1956) who highlighted how Simmel recognised that for a low socially ranked groups to aim for a higher social status, and act accordingly, its members must become aware, at some point, not only of the reality of the current social structure, but more importantly of its illegitimacy and reversibility: "Before a social conflict between negatively and positively privileged groups can take place, before hostile attitudes are turned into social action, the negatively privileged group must first develop the awareness that it is, indeed, negatively privileged." (Lewis Coser 1956, p. 37).

From Coser's analysis of Simmel's work, it emerged that social conflict is not only a core component of historical and social transformations, but also an epistemically transformative experience for both parts of the conflict. Engaging in social conflict, with the aim of modifying the current hierarchical composition of the society, is a process that allows social groups, and their members, to better understand their separateness with others, the boundaries within which the group is inscribed, and other defining features of theirs as socially situated individuals and groups. Moreover, during conflict, the entire

[14] See also Van der Linden et al. (2020) for interesting results about the use of the term 'fake news' to discredit opposing views and out-groups.

social space becomes aware of values and norms that were "dormant" (Simmel's term) and that needed the trigger of the inter-group conflict to emerge. These values and group norms could then be the ground for new social rules and regulations, once the conflict is over and a new social order is established. These mechanisms of social transition and modification are, already in Simmel's view, all possible because of the triggering force of conflict, and grounded on a dynamic conception of social identity formation and modification mechanisms. Social conflict is then not only limited to the accomplishment of social changes, but it is crucial to understand one's personal and group's social positioning. The recognition of this defining and self-determining efficacy of social conflict allowed Simmel, already in 1908, to notice and explain why conciliatory attitudes by higher ranked social groups could possibly be rejected by lower ranked ones, which is an intuition that has later been confirmed amid social identity theory's empirical work (Tajfel and Turner 1986, p. 23). The reason for this, perhaps counterintuitive, result is that it is only through conflict that lower groups' opposition against the status quo could possibly be stated and maintained (Simmel 1908, p. 411), and hence it is an irreversible process due to its epistemic, not only social, relevance for those who feel socially disadvantaged. The fact that when social hierarchies are believed to be unstable and unfair, the attempts of inter-group conciliation can be rejected, is in contrast with Hornsey and Fielding (2017)'s proposal, whose limitations derive from its dependence on a static interpretation of social identity, namely as a fixed set of features that characterise different social groups, and that can always be determined simply by considering determinate properties of their members, like political ideas, religious affiliation, education and others. The fact that groups who believe to be of low social status are more likely to reject conciliatory attempts will be particularly relevant also in the next chapter, where I will try to show that the spreaders of disinformation are very often people who endorse this belief regarding their social status, which makes it even less desirable to pursue the proposal made by Hornsey and Fielding (2017).

Getting back to the main topic of this chapter, I have already pointed out that there are cases where particular social identities are characterized by defining features, including common beliefs, that could be antithetic with certain scientific results or claims, and hence are meaningfully considered to be possible predictors of disinformation sharing. However, this does mean that every instance of disinformation sharing should be

explained only in this way, namely as the result of a clash between a scientific claim and a belief or attitude that is typical of a specific social group. On the contrary, I believe that disinformation sharing could be rightfully interpreted also as a different form of social conflict. Many individuals might indeed engage in this peculiar form of conflict as part of a process of formation, understanding, or redefinition of one's social identity, but also as part of a process of disidentification, namely as a way to react against their social condition and the balance of forces currently present within the society. The sharing of disinformation about scientific issues, considered within a frame of social conflict, could acquire all these different meanings. Individuals can surely share disinformation to express group solidarity, for instance by sharing a content that denies a scientific claim that is incompatible with the beliefs endorsed within the group. This is very likely the case considering the compact rejection of climate science by US conservatives. However, it is also possible that sharing scientific disinformation is a way to point out and question the social norms that regulate the relation between science and society, and that for some groups might be dormant, in Simmel's sense. Disinformation sharing would then be a way to trigger a social conflict which is necessary, from the perspective of some groups, to question the legitimacy and fairness of the current role of science in the society. Sharing disinformation about climate science could then be at the same time a way to question the influence of scientific institutions, like the IPCC for instance, or to reject the idea that the mitigation of climate change is a priority, or simply to signal one's belonging as a member of a specific political group.

Social identity theory is a very useful theorical tool. It does not only claim that people are group and identity protective, but that there are various forms of social conflicts and possible reasons behind them. Thus, when dealing with a specific form of social conflict, to understand what motivates the conflict it is necessary to consider the social context and all the possible dynamics at stake. Otherwise, the risk is to interpret certain actions as manifestations of identity protective attitudes while the case is actually the opposite, namely that conflicting attitudes arise from people's desire to disidentify from their current social status. Therefore, social identity theory is crucial for the study of disinformation, in particular, to understand the dynamics implied in disinformation sharing, but only as long as the endorsed idea of social identity is dynamic and reflexive, in the discussed sense.

In the upcoming sections, I argue that some attitudes and behaviours on social media, including the sharing of disinformation, were part of social identity transformative processes, and not of confirmatory and protective ones. In the next subsection, I support this claim with an example that comes from the recent political history, in particular from the UK Brexit referendum.

3.3.1. Becoming a "Real British": The Use of Identitarian Rethoric to Spread Disinformation During the Brexit Campaign

In this section, I consider Oysterman and Dawson (2020)'s brilliant analysis of the pro-Leave movement's communicative rethoric during the campaign for the UK Brexit referendum. I believe that the originality of their conclusions depends on the dynamic conception of social identity formation processes that they have adopted. Oysterman and Dawson (2020) argued that the pro-Leave campaign managed to reframe the decision of whether to leave or not the European Union as an identity-based choice, namely one between a more and a less "us" (as British) thing to do, with a pro-Leave vote being represented as the more coherent one with the British identity. This identity-based reframe of the referendum has been accomplished with the usage of culturally fluent terminologies and slangs, aimed at generating what authors have called a "social tuning", namely a common interpretation of the decision in identitarian terms. Once the entire issue was reframed as an identity based one, the complexity of the topic was extremely reduced for those who effectively attuned to this identitarian narrative, and with lower cognitive effort invested in the elaboration of the issue, people became more exposed to disinformation, and in fact, some pro-Leave false and misleading contents managed to spread widely by leveraging on the identitarian reframe of the referendum[15]. In short, social tuning, accomplished by culturally fluent pro-Leave messages, not only caused people to rethink the Brexit issue in identitarian terms, but it also allowed pro-Leave disinformation to go unnoticed, due to the lower cognitive effort invested in the elaboration of the referendum. Having only a static conception of social identity in one's toolkit, the described communicative strategy could only be considered efficient among

[15] These false and misleading claims were not only political but also scientific, and some of them were related to climate change.

those with specific features, namely a strong nationalist identity and an a priori propensity to elaborate the possibility of the Brexit as an identity-related one. Differently, Oysterman and Dawson (2020) approached this issue while endorsing a more dynamic conception of social identity and as a result of it, they were in the theorical position to claim that the described information strategy has not only been able to redefine a decision in identity-based terms, but also to offer an answer to those in doubt of what was the most identity-congruent decision to make. From their analysis, it resulted that the pro-Leave rethoric was not only composed of a descriptive intent, namely amusing people with the idea that their decision was actually between what a real British would do or not do, but it also had a normative component, namely, to argue that the most "us" thing to do, namely what a "real British" should have done was voting Leave. In other words, the described communicative strategy was not only functional to attune many people with the idea that the Brexit should have been interpreted as a social identity related issue, but it offered also a narrative of what was the social identity congruent choice to make. Behind the explanation offered by Oysterman and Dawson (2020) there is the idea that "identities are dynamically constructed in the moment" (p. 182), meaning that it is not only one's social identity that explains the decision to vote for Leave, but also the other way around, namely that some of those who ended up voting for Leave, as a result of their own decision, formed or modified their social identity accordingly. A dynamic notion of identity formation is needed in this case to see how, following the reframe of a decision in identitarian terms, when accompanied with a normative prescription of what is the most identity congruent decision to make, it is possible that is the decision made that defines the social identity, and not only the opposite. It is important to stress these points to show the speculative benefits that a dynamic conception of social identity formation could have in the debate on the proliferation of disinformation. In the case of the Brexit campaign, the pro-Leave movement made use of disinformation to support their positions. With the conception of identity protection that is currently diffused in the literature, a static one, only two possible explanations of the pro-Leave disinformation spread would have been available. The first is that reframing the referendum in identitarian terms worked in reducing the level of perceived complexity of the entire issue, determining also a reduction of the cognitive effort invested in the elaboration of the matter, and making people more exposed to disinformation as a consequence of automatic cognitive

responses and biased reasoning processes, coherently with the dynamics described in Chapter 2. The second possible explanation has already been mentioned, namely the idea that the pro-Leave false claims were aligned with people's ideas, priors, worldviews and group identities, and hence their acceptance resulted from their convergence with people's beliefs or by motivated reasoning processes. A dynamic conception of identity formation allows to make a further claim, namely that, because of the way in which it was reframed, the decision to support the Leave movement constituted, in itself, a piece of information about one's social identity, giving an answer to someone who felt deprived of it, or confused about their social condition and status.

Following the latter discussion, it appears meaningful to claim that social identities are also formed, and not only defended or reinforced, as it happens within echo chambers for instance, through disinformation sharing, which supports my argument that the literature should adopt a more dynamic notion of social identity. The latter would allow to consider a long-term severe risk related with the spread of disinformation, namely that it systematically influences the processes of social identity formation, and it would allow to better understand disinformative strategies. Diaz Ruiz and Nilsson (2023) have argued that part of the rhetorical strategies of disinformation creators is to spread false and misleading contents by representing them in the form of identitarian claims and beliefs which, as such, are more difficult to debunk. This considerable use of an identity-based rethoric made by disinformation creators could enlarge its effects on people's identity formation, a scenario that must be considered and studied more extensively.

In this sub-section, I have only pointed out the potentially reflexive power that a certain action could have in terms of social identity formation, considering the real case of the Brexit vote, and it is important to remark that what I argue is not that motivated reasoning and identity protective attitudes are not proper explanations of disinformation acceptance or sharing. On the contrary, I have reported various empirical results supporting the idea that it is very likely that people accept and share items of disinformation because of identity and group protection reasons. However, I argue that focusing exclusively on the latter mechanism would represent a major mistake, since it would impede to notice that the spread of disinformation is also determined by processes of social disidentification and by reasons of identity formation. In the upcoming sections, I furtherly elaborate this idea to the specific kind of social action that is disinformation

sharing. I will argue that sharing disinformation has a potentially powerful reflexive effect on the social identity formation process of those who perform it. To accomplish this task, in a way that is theoretically efficient, I will consider the sharing of disinformation under a peculiar light, namely as a cooperative action.

3.4. Are We All on the Same Page? Sharing Narratives Through Disinformation

Piazza and Croce (2022) defended an interesting position regarding the possible non epistemic reasons for the sharing of false and misleading information. I have already mentioned the fact that there is no link of necessity between sharing certain news and being doxastically committed to the truth of the shared contents, and starting from this datum, Piazza and Croce (2022)'s idea is that, for certain people, what really motivates the act of sharing fake news is related to a desire of contribution to a common narrative (p. 66), and this for two deeply connected reasons. The first is making sense of a certain state of things, and in Chapter 2, I have mentioned other studies that focus on the impact that the need for sense making has on people' s decisional processes and states of belief (see 2.2.3). The second reason is deeply connected to the latter and regards the attempt to find psychological and personal relief out of a certain narration of negative facts and events. In analogy with the idea of Piazza and Croce (2022), Bergamaschi Ganapini (2021) argued that the action of sharing false contents is primarily motivated by pragmatic reasons, and in particular by the attempt to signal something to the audience, including one's social and group belonging which is necessary to solve "coordination problems". To coordinate around a common narrative allows to define a set of shared social norms, values and pieces of knowledge that makes it easier to orientate within a certain social environment, and sharing similar contents, even if false, allows different people to do so and become aware of their social group membership. The idea that sharing false contents can be part of a process of construction and progressive reinforcement of a common narrative fits very well with the idea of a dynamic and reflexive social identity formation process envisaged in the previous section. I have argued that the relation of dependence between the sharing of false contents and the endorsement of a particular social identity is not monodirectional, namely it does not go exclusively from a fixed social identity to the decision of sharing, but also on the other way around, meaning that the action of

sharing disinformation has effects on the sharer's social identity formation process as well. Moreover, the idea that sharing false contents is part of a process of attunement to a common narrative is compatible with the reported example of the pro-Leave rethoric during the Brexit campaign. Here, the use of disinformation was a component of a more general Brexitist narration of the referendum as a decision between being a true British or not.

However correct, the idea of Piazza and Croce (2022) is partially limited by the examples used to describe it. In particular, I believe that much subtler phenomena could be explained by referring to their interpretation of sharing false content as a way to attune to a common narrative, in combination with the idea of Bergamaschi Ganapini (2021) on the signalling function of false contents' sharing. Piazza and Croce (2022) mentioned the Pizzagate (Samuelson 2016) as an instance of a fake news that has been shared in the attempt to contribute to the consolidation of a certain narrative. Cases like these, where the absurdity of the facts included in the shared contents are blatant, it might be evident that what lies behind their sharing is very unlikely to be related with a doxastic commitment to the truth of the news, or equivalently, this might be the case for someone, but for a very meagre number of people. Even more importantly, sharing a content of this absurd kind is very likely to be related with the mechanisms of sense making and psychological relief previously recalled (2.2.3), which are indeed commonly associated with conspiracy theories like the Pizzagate. Considering less implausible cases, where a subtly false content is the object of sharing, even if it is more complicated to understand what lies beneath their spread, this does not mean that it could not be related with an attempt to contribute to a common narrative. On the contrary, I believe that Piazza and Croce (2022)'s argument remains solid, and that theirs is a promising explanation of false information sharing even in cases of less implausible contents, simply, in these cases, more effort is needed to understand the motivations and the features of the narrative around which sharers cluster. In the next subsection, I use another real-life example to better illustrate my point, and defend the one described by Piazza and Croce (2022)'s work.

3.4.1. Far-Right Disinformation on YouTube: The US Case

In this subsection, I consider the case of the rethoric used by the US far-right communication networks, which are platforms of discussion and disinformation sharing visited by millions of users. One common trait of almost all far-right political forces and movements is to distrust traditional media, encourage the emergence of a distrustful attitude in their supporters, and eventually form their own media channels and platforms. In a report on the emergence of far-right media networks in the US, Lewis (2018; 2020) analysed in particular the YouTube channels. What emerged from this work is that, instead of focusing on the repetition of certain contents, perhaps in the attempt to foster the emergence of an illusory truth effect, which is what Pennycook et al. (2018) considered a promising strategy (recall the discussion in 2.2.6), within these YouTube channels the main attempt is actually to convey a sense of authenticity and familiarity through direct engagement with the viewers, and more importantly a sense of belonging to a community (Kreiss 2018). There are channels where lack of care for the accuracy of the shared news is blatant, others where a scientific approach is mimicked while sharing false claims, others that invest more in the characterization of themselves and their supporters as outsiders, and also free and critical thinkers. All those different approaches are committed to the same idea, namely "rallying around a shared narrative and collective social identity" (Lewis 2018, p. 20). These alternative media channels invest a lot of effort to make certain items of disinformation widely spread. Considering disinformation about climate change and climate related issues, some of the most popular YouTube channels among those considered by Lewis (2018) are actively funded by interest groups. One of the main investors is the Koch Foundation, which has an unfortunate history of spreading and financing climate disinformation (see 1.1.4). However, the main goal of these platforms, as Lewis (2020) has pointed out, is not to convince those who are already prone, for political and ideological reasons, to accept certain contents like climate disinformation, but to create an environment where a specific narrative is progressively designed, and where people who are not already fully committed to a specific worldview or ideology could feel invested with an active role within a concrete community, and consequently with a group identity. As a consequence of this, people will signal their belonging and contribute to the enlargement of the narrative by sharing the elements that better characterize and define this particular narrative, which quite often are pieces of

disinformation[16]. In other words, in the attempt to cooperate with an open and apparently welcoming community, that is able to intercept the desires and feelings of many people who, by joining these groups, will end up contributing to the expansion of a narration that is based on false claims and disinformation, and in a way that unconsciously conditions their social and individual identities' formation process.

At this point, it is necessary to make some order among the various elements considered, and I believe that it is possible to offer a theory of disinformation sharing that is grounded on two main pillars, a dynamic conception of social identity formation, and a conceptualization of disinformation sharing as a cooperative action. In the next section, I will consider the socio-psychological literature on cooperation to show that the social factors that determine the emergence of cooperative attitudes, in groups and individuals, are widely present within the mentioned groups of disinformation creators, and possibly in others as well.

3.5. A New Look for Disinformation Sharing: Cooperation

I argue that it is possible to reinterpret the act of sharing misinformation and disinformation, online and on social media, as an instance of a cooperative or anti-cooperative behaviour, depending on the adopted reference point. More precisely, considering the sharing of scientific disinformation could simultaneously be an act of cooperation with the creators of the shared contents, and an act of anti-cooperation with the scientific institutions that struggle to debunk the same contents. I do not claim that all the instances of disinformation sharing should or could be read and explained in terms of cooperative behaviour. The proposed reinterpretation is perfectly compatible with the existence of all sorts of other reasons for disinformation sharing, both non-epistemic and epistemic, as being stuck into an epistemic bubble, or because of a strong information deficit, which has been rightfully indicated as a possible predictor of scientific disinformation acceptance and sharing by the proposers of the GBM (recall the discussion in 1.4.1).

[16] Unfortunately, the contents spread within these platforms often include also instances of hate speech and racist claims.

It is important to keep sharply separate the idea that I discuss in this section with the dynamics taking place within echo chambers, which might be understandably interpreted as instances of cooperative behaviours. As seen, echo chambers are very peculiar epistemic environments where the individual decision to share a certain content is essentially determined by the epistemic structure of the echo chamber, and by the related dynamics of belief maintenance and reinforcement. However, the one that I present and discuss here is a different form of cooperation, one that is not dependent on the endorsement of irrational states of belief, as in the case of the echo chambers. More explicitly, I claim that it is possible to share disinformation as a form of cooperation, or anti-cooperation, without being necessarily a member of a disinformation-based echo chamber. Nonetheless, I believe that the repeated endorsement of this general kind of cooperative attitude could eventually lead to joining an echo chamber. To consider the sharing of disinformation as a cooperative action can offer various research improvements and unveil new possible reasons and dynamics of disinformation proliferation. To realise this potential, it is necessary to understand what the predictors of cooperation are. I consider the socio-psychological literature on this topic, and in particular the group cooperation theory, which is related primarily to the work of Tom Tyler.

From the empirical data collected in various works, the overall result reported by Tyler (2010) is that cooperation is primarily predicted and determined by social motivations, and not by material gains or profit. More specifically, the social variables that condition people's cooperative behaviour are social identity related ones. The awareness of being a valuable, respected, and fairly treated member of a collective predicts the emergence of cooperative attitudes more than other kinds of motivations. This has reportedly been the case in studies conducted in different working environments, where Tyler and Blader (2000) noticed that cooperation among co-workers was predicted considerably more by status related variables, in particular pride and respect[17], than by economic reasons or other possible material benefits. Maybe more interestingly, even in conditions of poverty, where it was sound to expect that people would have been more

[17] Tyler (2010, p. 39) defines these two variables as follows: "Pride reflects judgments about the status of the group, […]. Respect reflects judgements about one's status within the group."

motivated by material and economic factors, Tyler (2010) has shown that people tended to be cooperative with the others, and with the institutions and the authorities, majorly because of social reasons of the mentioned kind[18]. In particular, one of the main predictors of cooperation is the belief of being part of a society where fair rules are applied during collective decision-making processes. Very similar results derive from studies regarding the predictors of cooperation with group authorities. When people are fairly treated by both political and legal authorities, they form the belief that they are members of a social environment that values procedural fairness. These socially related factors predict cooperation with legal and political authorities, even in conditions of difficulty, more than material reasons and motivations (Tyler 2010, Chapters 4 and 5).

Taking stock, what the reported results seem to have shown is that, when trying to explain an agent's or group's cooperative or anti-cooperative behaviour, the proper thing to do in order to understand the related motivations is focusing on the characteristics of the inhabited social environment, coherently with the discussion made with respect to social conflict. Moreover, from the elaboration of the reported results, Tyler (2010) and Tyler and Blader (2000) have offered fruitful insights regarding the relevance that cooperation has for social identity formation processes. What results from their studies is that the experiences and the modalities through which people are treated by the other members of their social groups, particularly by the authorities, and also by members of other groups, represent a decisive source of evidence in the process of social identity formation, in line with social identity theory (Ellemers, Spears, and Doosje 1999; Tajfel and Turner 1986). People develop cooperative attitudes and mature self-esteem, which are important components of one's social identity, because they are proud to be members

[18] If the reader is familiar with the literature on poverty traps, the latter result may not be very surprising. In conditions of extreme poverty and deprivation, people develop peculiar decisional patterns, called poverty traps, that fosters the permanence in a condition of deprivation instead of favouring its disappearance. The latter are deeply related with social factors and motivations. It is also interesting to notice that the data collected by Tyler and Blader (2000) show how much poor people value social factors and motivations that are also predictors of mutual trust. The latter point is aligned with recent studies in development economics showing that the emergence of trusting attitudes within poor communities limits the emergence of the behavioural patterns that constitute poverty traps. See: Banerjee and Duflo (2011), Haushofer and Fehr (2014), Mani, Mullainathan, Shafir, and Zhao (2013), Jachimowitz et al. (2017), Ausili and Martini (2023).

of a valued social group, and their status is determined by the respect received from other groups and by the perception of being fairly treated by their authorities. Then, and this is a crucial point, once developed because of the mentioned social motivations, cooperative behaviour itself works as a source of evidence. As a result of their cooperative attitude within the society, people infer that the latter is related with their identification with the group, which is due to the same reasons that trigger cooperation. In a way that efficiently resumes this intricate point, Tyler and Blader (2000, p. 146) explicitly claimed that: "By cooperating with the group, the person is also facilitating the creation and maintenance of his or her own favourable social identity", and later in the same text: "[…] people cooperate with groups because they identify with groups, *drawing their sense of themselves from them*." (my emphasis, p. 184).

Therefore, the already discussed reflexive nature of social identity formation processes appears evident in the context of cooperative behaviour. People's reasons to cooperate are related to their social identity, being these primarily predicted by social motivations, but social identity itself is formed and reinforced by the cooperative behaviour instantiated by the same individual who acquires knowledge of her social position out of it. In other words, people discover their social belonging, in part, from their own cooperative attitude, but it is their social belonging that explains their cooperative attitude.

This mechanism is no circular at all, at least not in a rationality related sense, and clearly not once a static idea of social identity, currently popular in the literature on disinformation, is dropped. If sharing is a form of cooperation, and if disinformation sharing is motivated by social identity related reasons, then once a dynamic notion of social identity is endorsed, it evidently results that the social act of sharing disinformation constitutes, for those who share it, new evidence regarding their own social identity and status. At this point, being in the position to know what factors influence cooperative behaviours, namely social motivations related to the perception of being respected and fairly treated, it is possible to claim that when some members of a social reality manifest anti-cooperative attitudes, or engage in social conflicts, this might be due to them not feeling respected or fairly treated. Moreover, it is possible to argue that if this anti-cooperative behaviour is prolonged, not only people's behaviour, but their social identity formation processes will be influenced by that, turning them into deeply anti-cooperative

agents on the long run. Echo chambers come to the rescue in clarifying this point. As previously claimed, it is possible to share disinformation as a form of cooperation without being part of an echo chamber, but as sharing goes on, it is very likely to fall into one, and with prolonged permanence, people could turn into completely different individuals, both in epistemic and social terms.

To conclude this chapter, consider again the discussion on the US far-right alternative media channels. Having at disposal the results of group cooperation theory, it is possible to see that what the network of far-right YouTube channels has been able to accomplish is not only to create an efficient narrative, but an environment where all the main predictive factors of cooperative behaviour are exploited to gain support, and to make their narratives spread all over the digital environment. Through the development of a sense of familiarity and belonging, of the kind described by Lewis (2018; 2020) and Kreiss (2018), subscribers became part of a social group designed to make them feeling involved, respected, fairly treated, and valuable members of an important and large community, hence prone to cooperate with them and contribute to the proliferation of their contents, ideas, and related narratives.

Therefore, by focusing again on scientific disinformation, and by inverting the reference point, it is possible to argue that the decision to be anti-cooperative with scientific institutions, by sharing items of scientific disinformation, could be evidence of the fact that the sharers are motivated by the discussed social reasons, including the feeling of not being respected, fairly treated or considered as members of valuable social groups within their societies. In the previous section, I have argued that sharing disinformation regarding scientific issues might be a way to engage in social conflict, and that it could be part of an attempt to question the legitimacy of the current relation between science and society. More concretely, I claim that it is worth considering the possibility that there are groups who believe to be unfairly treated by the authorities and the institutions, including the scientific ones, and for this reason share scientific disinformation, whereas others could feel culpably ignored by the society at large, and for this reason become vulnerable to every kind of anti-institutional disinformation, which may include anti-scientific ones. In the next chapter, I will face these issues and isolate the social groups that apparently are the more likely to develop anti-cooperative and distrusting attitudes that could eventually lead them to fall into the disinformation trap.

Conclusion of Chapter 3

The spread of disinformation is a phenomenon caused by the combination of epistemic and non-epistemic factors. Among the latter ones, motivated reasoning plays an important role in conditioning people's accuracy evaluations and sharing attitudes, and I have reported various empirical results that support this hypothesis. However, it would be a mistake to argue that the proliferation of disinformation is exclusively dependent on dynamics of group membership signalling, cheerleading, or other forms of social identity maintenance and reinforcement. The latter misperception is present in the literature on disinformation, and this is due to two entrenched reasons.

On the one hand, the digital environment is still widely considered to be subordinate to the one offline, and for this reason it is poorly considered as a space where social identities are formed and not only maintained or reinforced. I claim that the latter assumption is no longer tenable, and consequently, in this chapter, I have argued that the possibility that the spread of disinformation negatively influences people's social identity formation processes must receive more attention and research efforts.

On the other hand, I have criticized the notion of social identity considered in the literature on disinformation. In particular, social identities are often considered as sets of beliefs related to one's social group and individual characteristics, that are almost immutable, or at least very hardly questioned, in a way that is not justified by social identity theory. For this reason, some authors recommend communicating with groups that share disinformation for identity protective reasons, from a perspective that appears compatible, and not in contrast, with theirs. However, it is not always possible to sharply determine the features and beliefs of a certain social group, and so the related reasons for the behaviour of their members. Social identity theory has indeed shown that people are identity protective, but also that they constantly try to redefine themselves as socially situated individuals, and that they engage in social conflicts and processes of disidentification with their current group, in order to pursue a better social status. In this chapter, I have argued that the action of sharing disinformation could also be part of processes of the latter kind and thus be often motivated and conditioned by social identity formation processes, rather than maintenance and reinforcement ones. In subsection 3.3.1, I have supported this view with a case study. Following this line of inquiry, the

possibility, already discussed in literature, that the spread of disinformation is primarily a way to disseminate narratives and worldviews, more than specific contents, appears well supported, and I have defended it. In section 3.4, I have argued that the current spread of disinformation is in part due to certain social media users who share disinformative contents to contribute to the proliferation of specific narratives and worldviews. In other words, my claim is that the spread of disinformation is partially determined by a form of cooperation. This attitude is motivated by the fact that within certain digital environments and communities, that are characterized by the production and exchange of disinformation, the social factors that motivate cooperation are strategically and efficiently triggered. This research perspective allows to see that those who are prone to accept and share disinformation are not necessarily and always those who belong to specific social groups, whose ideas and beliefs are aligned with the pieces of disinformation that they end up sharing. Social identity theory has shown that, in conditions of social uncertainty, people are vulnerable to groups and rhetorics that are strongly identitarian and unambiguous in their prescriptions and claims (Hogg 2014). This is different from the need for sense making discussed in 2.2.3, which was related to the psychological need to make sense of a specific event. In the case of social uncertainty discussed by Hogg (2014), the one involved is a more general and deeper socio-psychological need to be aware of one's position and role in the society. Identitarian and extreme worldviews are unambiguous with this respect, and offer immediate answers and prescriptions, as seen in 3.3.1. Since the narratives and worldviews defended within antiscientific communities of disinformation spreaders are of the latter kind, it is possible that a non-trivial number of sharers of disinformation are people who felt socially insecure, anxious, and doubtful about their role and position in the society, at the moment in which they approached these groups that offered them clear worldviews, identities, and behavioural prescriptions, that could have sublimated their feelings, and offered them a sense of belonging to a community. Unfortunately, one that is founded on the sharing of disinformation, and other vicious practices.

4. The Army of the Left Behind: Social Exclusion in a Digital World

In this chapter I claim that the current spread of scientific disinformation is imputable, inter alia, to a diffused distrust in science. In particular, I argue that people distrust scientific institutions and experts in large majority due to a perceived lack of benevolence, more than competence. Therefore, mainly for non-epistemic reasons.

In literature, it is reported that there are certain social groups who tend to distrust science more than others. The main argument that I defend in this chapter is that alongside the ones already considered in literature, there is another social group whose members are prone to distrust science and consequently to become vulnerable to scientific disinformation. This is the group composed by those who feel socially excluded, left behind, and who have negative expectations about their future. In the difficult attempt to define the boundaries of this social reality, I will consider various empirical data.

In conclusion, I will claim that the belief of being socially excluded works as a form of higher order evidence that impairs the acceptance of evidence coming from scientific institutions. For an individual who believes to be left behind by higher social classes and institutions, the acceptance of pieces of evidence coming from these institutions is epistemically impeded, and the result of their reception is a condition of epistemic akrasìa. The act of sharing disinformation is then a result of the latter condition. The interesting upshot that derives from these considerations is that, in a condition of perceived social exclusion, it is possible to fall prey of scientific disinformation even when the vast majority of the contents to which one is exposed come from scientifically accurate sources. The chapter is structured as follows.

In section 4.1, I introduce the possibility that the current spread of scientific disinformation is related to a diffused distrust in science. In the related subsection 4.1.1, I argue for the possibility that the current distrust is science is related to the fact that various people impute a lack of benevolence, more than competence, to the scientific institutions. In subsection 4.1.2, I stress the difference between the distrust in science endorsed by minority groups and reactionaries. In section 4.2, I introduce the category of the socially excluded and in subsection 4.2.1 I consider some empirical evidence to support the claim that in societies where perceived social exclusion and negative

expectations are high, it is more likely for distrust in institutions, including the scientific ones, to settle, and hence for scientific disinformation to spread. In the conclusive section, 4.3, I claim that the position endorsed in the previous sections is supported by the possibility that the belief of being in a condition of social exclusion works as a form of higher order evidence.

4.1. Distrust in Science and its Impact in the Spread of Scientific Disinformation

The current spread of scientific disinformation can be interpreted as the sign of a friction between part of the civic society and the scientific institutions. More specifically, it is possible that those who accept and share false contents about science, online and on social media, do that because they distrust scientific institutions. In Chapter 2, I have mentioned the work of Aklin and Urpelainen (2014) where it is reported that the higher the scientific agreement on a certain issue, the more the support for policy regulation on that issue decreases among people who distrust scientists. More topic-specific evidence shows that, for instance, distrust in government and institutions is predictive of anti-vaccination attitudes and beliefs (Mitra et al. 2016, Hughes et al. 2021), and hence to the acceptance of scientific disinformation about vaccines. To determine to what extent distrust in science could contribute to the spread of disinformation, it is necessary to understand the possible reasons behind the emergence of a distrusting attitude towards scientific institutions.

4.1.1. A Perceived Lack of Benevolence Triggers Distrust in Science

Countless experiences, in our daily lives, are made possible by the fact the people trust each other. This is the result of a contingent process. Many societies have been, and still are, perfectly able to regulate themselves on different bases. A strong commitment between different individuals can result from prolonged relations that are not motivated by the parts trusting each other, but by a condition of long-term mutual dependence. In a social context where the possible partners of interaction are numerically limited, the mutual awareness of being necessary to each other is the only needed engine for the emergence and settlement of a collaborative behaviour (Banfield 1958 for founding

research on this social mechanism). In conditions of mutual and prolonged dependence, it is not appropriate to describe people's interactions as based on trust, since for the latter to possibly be the case, it is necessary that the parts involved are all uncertain about the outcome of their exchange, and that they can all potentially withdraw and change partners of interaction (Cook, Hardin, Levi 2005, Chap. 3). In a situation in which two parts have no alternative but interacting with each other, and cooperation is almost inevitable due to the social costs of deception, trust is not part of the picture. The latter scenario fades away with the emergence of complex societies, characterised by the possibility to entertain countless, short-term interactions with various and heterogeneous agents. When social opportunity costs are too high[19], and it is not possible to interact only with well known, long-term partners of interaction, trust comes into the picture as the regulatory factor of social exchanges (Luhmann 1980). Thus, trusting others could be interpreted as a way through which humans cope with their own cognitive and temporal limitations. To know who to trust is a way to avoid all sorts of time and energy consuming analyses regarding every component of the decisional process one is involved in. The benefits that derive from regulating the social life on the basis of trust are such that it eventually occurs automatically and on an unconscious level, in a way that Nguyen (2022) has defined "unquestioning". According to Yamagishi (2011) this sort of unquestioning attitude is the result of a socio-evolutionary process that has led to the development of a form of social intelligence that consists in the ability to spot trustworthiness and be able to know who to interact with, avoid deception, and make correct decisions. Given this cruciality of trust for social life, various accounts have theorized the underlying mechanisms that characterize it, trying to define what dynamics occur between the parts involved, the trustee and the truster.

Some accounts of trust that have been labelled as "predictive", and among these there is the one of Hardin (1993; 2002) who argued that trusting attitudes vary alongside attributions of "encapsulated self-interest". Hardin's idea, in a nutshell, is that for the truster to trust the trustee, it means nothing more than the truster being aware of the fact that the trustee has a share, or all, of truster's interests included (encapsulated) in theirs. In other words, trust is a form of rational prediction of the fact that the trusted will behave

[19] The cost in terms of missed opportunities and possibilities that result from not frequently changing partners of interaction.

in a self-interested way, and that it is in their self-interest to be trustworthy with the truster. For Hardin's idea to hold, it is necessary that both parts are able to understand each other's motivations, alongside the expectation that the trusted has the truster's self-interest encapsulated in theirs. For instance, it is in the bus driver's interest to drive me safely at work, because if he drives dangerously, I might complain with his company, or because we might both get involved in a car crash, therefore, it is sound for me to trust him. This proposal is grounded on an intuitive and sound idea, but it does not seem applicable to every situation or relation. As argued by Croce and Baghramian (2021) the risk that predictive accounts of trust run in general, and Hardin's one in particular, is to reduce trust to a mere form of reliance based on rational expectations. However, not every relation of dependence between a trustee and a truster can be rightfully represented in this way. One reason is that reliance, Croce and Baghramian (2021) argued: "requires no commitment on the part of the trustee to display an appropriate reaction to the trustor's attitude." (p. 452). There are situations where it is appropriate to state otherwise, namely, to argue that the trustee must feel committed to behave in accordance with the expectations of the truster, in virtue of the received trust and on the basis of the matter at stake. For instance, in the previous example of the bus driver, it is meaningful to interpret that relation of trust in terms of reliance. Being reliant that the bus driver will drive safely means that I expect him to act in a certain way. All my expectations are related with his actions, not the beliefs, reasons or values that motivate them, and I do not expect him to be furtherly committed to drive safely by the fact that, by stepping into his bus, I am somehow trusting him to drive safely. Differently, to trust the recommendations of the OMS during a pandemic, for instance, involves different and further aspects that could not be squeezed in a relation of reliance, since I expect the experts working at the OMS to feel committed to be trustworthy, and to do their best to report the truth, by the fact itself that people trust them and will behave on the basis of their recommendations.

In general, anyone is somehow epistemically forced to rely on the work and competence of others, and particularly on the work of experts and scientists. From the more trivial actions, like using alcohol to sanitize one's hands, to more demanding decisions, like getting a vaccine or undergo surgery, such a condition of dependence is related to the clear impossibility to personally verify the reliability of any scientific recommendation. The trust that scientifically laypeople have on scientists and experts is

primarily motivated by this condition of epistemic dependence. As stated by Croce and Baghramian (2021), scientifically laypeople's trust in scientists and experts is not related only to the expectation that them, the experts who work at the OMS in my example, will do something specific, as it was the case in the example of the bus driver. To trust an expert's recommendation means believing in the epistemic strength of the arguments, reasons and beliefs offered in support of that specific recommendation, whose truth and reliability laypeople could never directly test by themselves. Therefore, the one that regulates this peculiar relation, characterized by epistemic dependence on the side of the truster, is a different form of trust, namely a "testimonial" one (Croce and Baghramian 2021, p. 451).

The evaluation of a testimonial claim is of a peculiar kind. Being it the only form of acquisition of knowledge that is indirect (Martini 2015, p. 28), namely received by others and not directly acquired, its evaluation involves various epistemic and non-epistemic factors. People's evaluations of different testimonial claims can be simultaneously influenced by priors, biases, worldviews and in general by the social context they are embedded in. As Levy (2019) has stated, the epistemic component included in the evaluation of a testimonial claim is related to the assessment of the competence of the testifier. The latter dynamic can be severely influenced, and often compromised, by non-epistemic factors of the mentioned kind, including biases and prejudices (Fricker 2007). Similarly, the determination of the benevolence of the testifier, namely her commitment to honesty and sincerity, an already non-epistemic factor, can be influenced by the mentioned aspects. These two components, benevolence and competence, are necessary to determine the reliability of the testifier, and hence the acceptance of her claims. Levy (2019) argued that the current spread of disinformation regarding scientific issues, and the apparent increasing distrust towards science and institutions, are particularly related with people's evaluations in terms of benevolence. The idea is that, for a person in conditions of epistemic dependence, it is natural to assume higher benevolence on the side of those who endorse the same values and political positions, and consequently to consider them more reliable. Levy (2019)'s argument is based on the idea that to determine the reasons behind people's commitment to certain sources and their trusting attitudes, it is necessary to understand their attributions of

benevolence, and, to do so, it is necessary to systematically consider the social environment they are embedded in, and the ideas that proliferate within.

Desmond (2022) combined many of the aspects mentioned so far and argued that people's distrust in science is of a particular kind, namely a *social status*-based distrust, meaning that certain categories, including experts and members of the scientific community, are distrusted *in virtue* of their social status (Desmond 2022, p. 588). More precisely, the idea is that there are groups who believe to be socially disadvantaged in comparison to those composed by experts and members of the scientific institutions, and that the relative distance in terms of social status between them and the scientific community is so wide that there are no shared values, interests, and priorities between the two. The core of the distrusting attitude against science, Desmond (2022) claimed, is grounded on this belief. To move a step further, if the socially disadvantaged believe that scientists, as a social group, do not share any value or interest with them, the former have reasons to believe that their interests will not be adequately considered during collective decision-making processes, both regarding plain scientific issues, but also data-based political decisions, which will impact the society at large. There is also a more material and politically realist argument in support of this general one, that those who distrust scientists may use to motivate their distrusting attitude. If, within the society, divergences come up, and conflict arises, between a lower and a higher socially ranked group, the lower one is very likely to be poorly supported by the other social groups, due to its low social status. Moreover, conflict and contrast are very likely to occur between groups who share no interests and values, which makes the described scenario even more realistic. Therefore, from Desmond (2022)'s perspective, instances of distrust in science are in large majority a matter of perceived lack of benevolence. Distrusting attitudes in scientists result from the belief that the latter, through their work, pursue interests and goals that are proper and exclusive of those in their social status, ignoring the ones of those who belong to the lower social groups. This proposal is coherent with the actual disinformation strategies adopted by some interest groups and disinformation spreaders. Considering the disinformation about climate change, as mentioned in subsection 1.1.4, one of the principal attempts of those who create and spread disinformation is not to question scientists' competence in climate science, but their integrity and honesty, arguing that with their work, they pursue personal and group-specific interests that are independent

from the search for truth and knowledge. Therefore, certain disinformation strategies actively try to foster the emergence of distrust in science, and they do so by leveraging on non-epistemic factors, including the lack of benevolence and honesty of the scientists who are said to be motivated by interests that are completely detached from those of the rest of the society and of the collective at large.

4.1.2. Who Is More Likely to Develop a Social Status-Based Distrust in Science? Minorities, Reactionary and the Socially Excluded

Worldviews, political ideas and perspectives do have an influence in the acceptance or rejection of scientific facts, online and on social media (Pereira et al. 2021, Hornsey et al. 2016, Hornsey et al. 2018, Jost et al. 2018, Lewandowsky and Oberauer 2016; 2021) and worldviews are shaped by one's experiences as an individual and as a member of a collective and social group.

In the elaboration of the idea of social status based distrust, Desmond (2022) considered in particular minority groups as those that, due to a history of oppression and discrimination, in certain occasions perpetuated also by scientific authorities (e.g. the unfamous syphilis study conducted in the Tuskegee's community, see Brandt 1978), might be the ones who are more likely to develop the described form of socially grounded distrust in science. Indeed, there are results showing that minorities, in different societies, reflect above average levels of distrust in scientific institutions (Kennedy et al. 2007, Jaiswal and Halkitis 2019, Batelaan 2021). Distrust might have played a role, and studies show that it partially had, during the recent COVID-19 pandemic as well, where there has been a disproportion, particularly among different ethnic groups, in terms of vaccination rates (Paul et al. 2021, Na et al. 2023). Unfortunately, results showed that members of minorities have also been more likely to share false and misleading information about the virus and related issues (Roozenbeek et al. 2020). It is crucial to collect results like these, that point out the occurrence of a peculiar form of resistance to evidence, and which offer also a more general picture of the socially and culturally related dynamics of trust. However, the one expressed by minorities is only one of the possible forms of social status-based distrust in science which are currently conducive to the acceptance and share of scientific disinformation by large groups of people.

In Chapters 1 and 2, I have reported data showing that there is also a completely different form of distrust in science, which is manifested by individuals who reject scientific claims for identity protective reasons. Free market supporters and conservatives often manifest a distrusting attitude towards specific scientific claims, and particularly climate change, possibly due to the consequences that contrasting the climate crisis would have on the economy, and to the political characterization attributed to the topic, which has been historically associated to progressivism (see 1.1.4). On a deeper ideological level, science can also be rejected because of a clash between the values that regulate the scientific practices and the ones endorsed by specific groups of people. Lewandowsky and Oberauer (2021) tested the mutual relations between conservativism, belief in climate change, vaccines' safety, and the endorsement of two of the norms that regulate contemporary science, communism and universalism. The latter have been defined by Merton (1949) and indicated respectively the idea that scientific discoveries should be made available for the entire society, and that science should be impersonal and unbiased. The endorsement of the latter norms resulted positively correlated with high rates of belief in both climate change and vaccines' safety, while negatively correlated with conservativism. This result could be framed within a more general phenomenon, the anti-intellectualist tendency that many right-wing political movements have manifested in recent years, and not only in the US context. I will elaborate more on this topic in Chapter 5, but for the moment it is important to notice that a specific form of distrust in science is evidently related with political factors, independent of one's belonging to a minority or disadvantaged group, and that it is more diffused on the right side of the political spectrum. I argue that alongside these two socially motivated forms of distrust in science, the one manifested by minority groups and the one related with the anti-expert attitude of many right-wing movements, there is a third one, independent from the other two, and that possibly includes the majority of people who endorse a form of social status-based distrust in science.

In the attempt to define who might develop a form of social status-based distrust, it seems natural to consider those who struggle the most due to precarious social and living conditions. A continuous experience of precarity might be accompanied by the perception that the hiatus with the higher social classes is unbridgeable, and therefore some might form the belief that the interests of the latter are simply incompatible with

theirs. In Chapter 2, I have mentioned Goertzel (1994)'s discovery of a significant link between the acceptance of conspiratorial claims, whose credence is a significant predictor of the acceptance of scientific disinformation (Lewandowsky et al. 2013) and the social uncertainty caused by precarious working conditions, or by belonging to a minority. More recently, Angelucci and Prat (2023) reported that it is 23% more likely for a member of a high socio-economic class to distinguish true from false news (Angelucci and Prat 2023, p.6), which implies that it is more likely, for someone who belongs to lower socio-economic classes, to interpret an item of disinformation as an accurate piece of knowledge. However, lower socio-economic conditions seem to be a factor, but clearly not an explanation for differences of this kind. Angelucci and Prat (2023) argued that one possibility is that the reported results could be related to the fact that members of lower socio-economic classes have lower chances, throughout their daily lives, to develop a critical attitude, and related useful abilities to evaluate the accuracy of the encountered news (Angelucci and Prat 2023, §6). Further evidence is needed to support this possibility, but I believe that it goes in the correct direction.

It is not the fact itself of living in conditions of paucity that makes people more gullible or more prone to accept disinformation as true. Economic struggle, difficult living conditions, and negative expectations for the future have detrimental effects at a cognitive, emotional, and even epistemic level, which are the ones to consider when dealing with issues like trust in science and scientific disinformation. In a founding work on the topic of precarity, Standing (2011) has reported that precarious working conditions are accompanied by the development of psychological states that include: low self-esteem, lack of sense of belonging to a community, cynicism, perceived isolation, anxiety, and inadequacy to personal and social expectations. Precarious working conditions inevitably come along with a lack of work-based identity formation processes, and with the emergence of feelings like anxiety, anomie, powerlessness, fear and anger (Standing 2011, p. 19). If Desmond (2022)'s hypothesis is correct, and distrust towards science is largely motivated by differences in social status, then, precarious people might develop a form of distrust as a result of the mental state that characterizes their living conditions, and that is such to make constantly salient the existing social gap between them and the better off social groups. In other words, it is possible that this condition is a

predictor for the development of a socially based distrust in science, which could trigger various reactions, including the sharing of scientific disinformation.

Consider anger. There is evidence showing that anger influences for the worst the evaluation of items information. The latter effect might be indirect. For instance, it is possible that anger fosters motivated reasoning which conditions information evaluation. Weeks (2015) indeed reported that anger fosters motivated evaluations of items of misinformation, leading to the formation of inaccurate and partisan beliefs. Conversely, as already argued multiple times, the creators of items of disinformation use highly emotive language to make their contents going viral. This mechanism includes the loading of people's anger. Pennycook and Rand (2021) reported that among the emotionally charged false news that they considered in their work, those that were designed to foster anger and fear were also those that were more likely to be shared on social media. In a more recent study, mentioned also in 2.2, McLoughlin et al. (2024) considered a very large set of posted contents on different social media and showed that, on average, the misinformative and disinformative ones were more likely to foster angry reactions, and more importantly, that these contents were also more likely to be shared without being read than accurate ones. In a more domain specific study, Featherstone and Zhang (2020) showed that anger, among other negative emotions, is the one that predicts anti-vaccination attitudes the most. In their study, anger resulted strongly correlated with two further emotions, a feeling of being "fooled by the establishment" and a general sense of frustration. Similarly, in the work of Han et al. (2020) anger resulted positively correlated with false information sharing about COVID-19. The authors noticed that, among the other emotions considered, anger was also the more influent in guiding and incentivizing action. In other words, those who feel angry need to act somehow, to express it in a way or another, and sharing disinformation appears to be a possible way to do that. The latter is a very important result, because reflects an attitude that incentivises the desire to manifest one's feelings to the outside, which is a need that can be easily satisfied in a digital space, and particularly on social media. Clearly, anger is not a feeling that is exclusively or inherently associated with precarious living conditions, as no other feeling is, but the reported results support the possibility that there is a link between a specific social condition, that motivates anger, and a propensity to fall prey of disinformation.

Nonetheless, it would be unfruitful to focus on the category of precarious workers as the one that, alongside with minorities and far-right supporters, is more likely to include the vast majority of distrusters in science. The class of the precarious is too variable and heterogenous to be considered as a unicum, which is also part of the reason why forms of solidarity among precarious people struggle to develop (Sennett 2016, Preterossi 2011). Even in presence of general aspects, like the ones mentioned by Standing (2011), regarding the psychological, cognitive, and social effects of precarity, it would be very complicated to argue that precarious people manifest all of these in such a uniform way that would allow to support the argument that precarity is associated with a socially motivated distrust in science. However, I believe that it is possible to isolate a specific social group, that experiences most of the feelings and mental states commonly associated to precarity, and which could develop a social based distrust in institutions, including the scientific ones. This social group is represented by those who feel socially excluded, left behind by the institutions and by the other, better off, social groups, and that are characterized by the perception of having negative expectations ahead, and not exclusively in an economic sense. In the next section I delineate those who might fall within this category.

4.2. Social Exclusion and Distrust in Science

Sen (2000) used the term "social exclusion" not to indicate a category of individuals, social events, practices or dynamics, but a research approach, a methodology of sociological analysis. This approach is supposed to be beneficial for the analysis of a certain social issue, like capability deprivation deriving from poverty and unemployment in Sen's case, or socially motivated distrust in science and consequent spread of scientific disinformation in my case. According to Sen, the benefits of a research approach founded on the idea of social exclusion is to: "forcefully emphasizing the role of relational features in deprivation" (Sen 2000, p. 8). To reason in terms of social exclusion, Sen argued, allows to recognize and point out the link between the occurrence of a particular social dynamic with other aspects of the social life. Following this approach, Sen pointed out various hidden aspects and effects related to the social issue of prolonged unemployment, including the idea that unemployment compromises freedom. Losing work and income implies also the loss of the possibility to be part of various social activities, and

particularly the ones that define the sense of belonging to a community or society. On the long run, the latter effect could determine the progressive detachment of the individual from the community, with all sorts of detrimental effects on the emotional sphere. In other words, someone who feels socially excluded is also someone who feels to be deprived of a sense of belonging, and the possibility to be part of the most defining and characteristic activities of their society or community. Prolonged periods of unemployment, rhapsodic working commitments, and low disposable income, force many people to live in a short-termed perspective that does not allow to make projects and to freely define personal aspirations. Moreover, in times that are characterized by a structural form of uncertainty, that is determined by the uncountable variables that might influence one's life, possibilities, and prospects, the socio-psychological condition recognized by Sen can be particularly impactful and detrimental. This sense of lack of control over one's life is incentivized in a deeply connected and globalized reality, and it has found a sad confirmation with the 2008 financial crisis (Srnicek 2017, Chap. 1). In the short-term, these feelings might find a political outburst, as happened in recent years with the emergence of populist and reactionary forces, but it is also possible that these aspects crystalize and become a stable and constitutive component of various social groups, and thus relevant for the determination of trusting attitudes. To firmly believe that what someone has in front are negative and decreasing future expectations has a real effect on one's perspective on the world, and on the inhabited social reality (Graziosi 2023). Even if there is no proxy to measure the rate of diffusion of these feelings across different societies, I think that by considering the proper combination of variables and empirical results, it would be possible to show that there could be a very important connection between the spread of scientific disinformation and a diffused feeling of social exclusion, conceived as the perceived impossibility to take part in the most important social activities and hence to fully develop as members of the society. It is very important to keep the socio-economic status distinct from the idea of social exclusion. The former is not necessarily linked to the latter. Arguably, the best predictor of the development of a sense of social exclusion is the belief of having pessimistic and discouraging expectations for the future, which might be related or caused by economic factors, but remain distinct from mere and simple economic struggle. In particular, social exclusion is related to the perception that decision makers do not consider these worries and negative expectations

as priority, or do not have the ability to face them. For this reason, the belief of being socially excluded is closer to the idea of being left behind, politically abandoned, and for this reason it can have consequences on people's trusting attitudes that economic struggle alone does not necessarily have. For instance, Lübke (2022) reported that among EU citizens, those who manifested negative expectations for their socio-economic future were the more inclined to distrust climate science, denying the occurrence of anthropogenic climate change, contrary to those who are currently facing a bad economic situation, whose condition was not correlated with any specific belief about climate science. Clearly, this datum alone is not sufficient to argue that those who are afflicted by discouraging expectations for the future are more likely to distrust the institutions, including the scientific ones, and hence more prone to reject scientific claims. Nonetheless, results like these suggest that it is worth exploring the effect that socio-cultural factors that are underexplored in this debate, like people's expectations for the future, have on trusting attitudes toward the scientific community. In this chapter, my attempt is precisely to argue that it is necessary to enlarge the set of social factors that are considered impactful in the evolution and modification of trusting attitudes. In particular, I will consider three different variables, namely: social mobility, trust in government and institutions, and nationwide exposure to disinformation. Considering these data, it is already possible to envisage the possibility that a wide social class, including, but not limited to large numbers of young adults and precarious workers, and the intersection of the two, is in greater danger of developing a sense of social exclusion and consequently a distrusting attitude towards the institutions, including the scientific ones.

Coherently with Desmond (2022)'s claim, and similarly with the distrust expressed by minority groups, I claim that those who feel socially excluded distrust science because do not believe that the public use of science takes their interests and perspectives into account, thus do not believe in the benevolence of scientific institutions. Meanwhile, and closer, in this case, to the form of distrust manifested by reactionaries, the socially excluded are particularly concerned with the future and by the possibility to inhabit a world that has not been designed considering their perspectives, ideas, and interests.

4.2.1. Where Future Expectations Are Low, Disinformation Is High

There are various countries in which, since many years now, it is possible to measure very low levels, or even detect a decline, of social mobility. This is evident in Italy, where Cobalti and Schizzerotto (1994) reported that the 60% of the Italian citizens born in the middle class were going to end up their lives in lower social classes. This is still the case, with Italy being one of the lower ranked countries in Europe in terms of social mobility expectations, very close to other Southern European countries like Spain and Greece, but also with the United States, whereas the countries with still high levels of social mobility include the Scandinavian ones and the Netherlands (World Economic Forum 2020). A low social mobility generates political and social resentment, which is associated with a disengagement from the social, political and economic life (Langenkamp 2021, Fehr and Fischbacher 2003). Then, by considering the few available data about national levels of disinformation spread alongside social mobility, a very interesting picture comes up.

Humprecht et al. (2020) defined a way to determine which countries are more (and less) resilient to disinformation exposure[20]. What resulted from their analysis was that the countries that are more vulnerable to disinformation are Spain, Italy, Greece and the US, and those which are less likely to be exposed to disinformation include the Scandinavian ones and the Netherlands. To be or not to be exposed to disinformation depends on specific factors. In particular, a diffused level of distrust in truth seeking institutions within the society, including scientific ones, is a significant predictor of the spread of disinformation, as it is the presence of populist forces in government and political debate, and high degrees of political polarisation and fragmentation. Another important factor is the distrust in traditional media, which is a predictor of a higher use of social media for information gathering, which is also the space of communication where more instances of disinformation circulate.

To enrich the picture, consider another set of data on the distribution of disinformation, limited to the EU area. Unzicker (2023) found out that Spain and Italy are reportedly the countries with some of the higher rates of self-reported encounters with items of disinformation, some of the higher reported levels of uncertainty regarding the reliability of the information found online and on social media, some of the higher rates

[20] In this study, data have been collected using both self-reported and not self-reported surveys.

of social media usage, among the highest variety of media used, and the highest rates of reported cases of mistakenly shared or liked pieces of disinformation. Moreover, and more importantly, these countries, with such high levels of reported encounters with disinformation, are also those where citizens are more distrustful towards traditional media (newspapers, television, radio) and more trustful in social media. In this report, Unziker (2023) considered also the levels of satisfaction, among EU citizens, regarding their respective national democratic systems. The lower levels of satisfaction were registered in Italy, Spain and Poland, and the higher ones in Belgium, Netherlands and Germany. Therefore, those who were satisfied with the democratic systems they were citizens of were also, on average, less uncertain about the veracity of the information found online, less likely to report that they have encountered disinformation, but they were also more likely to support interventions to regulate platforms in the attempt to limit the spread of disinformation (Unziker 2023, § 7). The citizens who, on the contrary, were dissatisfied with their democratic systems reported very opposite attitudes. These tended to be unsure about the veracity of the information found online, and in fact reported higher frequences of encounters with disinformation, but they were also at odds with a stronger regulation of social media platforms by the national governments. The clearest case is the Italian one, where people trusted the platforms but not the government, which is a possible indicator of the fact that citizens believe that the government itself is a spreader of disinformation, a hypothesis that would also explain the high numbers of reported encounters with disinformation in the country. Reading through these data, it is possible to see a correlation between distrust in political institutions and an ambivalent relation with the disinformation found online, which occurs in the same countries that are more exposed to disinformation, and where social mobility is significantly low, or even decreasing.

Clearly, this picture is dependent on data that are related to a specific time interval, and things can change in a relatively short amount of time, given the socio-political turmoil that characterizes the current historical period. Moreover, I acknowledge that to argue for a correlation of such a major scale is surely beyond the possibilities of this work. Nonetheless, I defend the underlying point and argue that ad hoc empirical work is necessary to test in a precise and statistically refined way the hypothesis that here I have only sketched. In particular, it is necessary to define a proxy for the condition of social

exclusion and test whether this is correlated to a propensity to accept and share disinformation, which could be mediated by a distrust in science and institutions, that is also very likely to be predicted by social exclusion, in my view. Indeed, the perception of being socially excluded could be determinant for the emergence of a social status-based distrust, that following Desmond (2022), can involve the scientific community, and consequently determine the spread of scientific disinformation. Further data seem to suggest that something of this kind is taking place.

As mentioned, Humprecht et al. (2020) reported that, inter alia, distrust in government and institutions is a predictor of disinformation sharing. The 2022 OECD Report: "Building Trust to Reinforce Democracy" published important results with this respect. Young adults, between 18 and 29 years of age, turned out to be those with the highest tendency to distrust their governments and institutions. This could be related to the fact that many countries that are currently characterized by low social mobility have also reported high levels of disinformation spread. Low social mobility is particularly concerning for the younger generations as they progressively become aware of the fact that they are very likely to have lower possibilities and resources than the previous generation. Moreover, women result to statistically distrust their governments more than men. The latter result could be related to the fact that, in many countries, women are still forced, by gender-based norms, to suspend their careers or terminate education programs to fulfil care duties, both of infants and of the elderly. In Italy, almost the 25% of women, at some point, had to interrupt their careers because of duties related with care, whereas only the 2.9% of men have to do the same (Istat 2019; see Granaglia 2022, Chap. 1). Considering Europe, the latter point is blatant when looking at the data about the gender distribution of the NEETs[21], that in many countries reflect a disproportion with much more women than men being members of this particular group. Italy again, has the second highest number of NEET citizens in Europe and, as for many other countries, the majority of people belonging to this category are women, a datum that is considered to be determined by gender-based social and family obligations of the mentioned kind (Odoardi et al. 2022). Clearly, being a NEET is a peculiar condition, which might be experienced in various ways, but it is evident that it is very likely to cause a competitive disadvantage

[21] Young People (15-29) Neither in Employment nor in Education and Training, 2023 data (Eurostat).

on the long run, and it is another indicator of the possibility that a feeling of social exclusion becomes pervasive among the younger generations.

Moreover, considering the effects that socio-economic variables have on trust in government, the OECD (2022) report shows that precarity, low income, financial vulnerability, territorial disparities, namely living in the poorer geographic areas of a territory that is characterized by considerable internal economic variability, are all significant predictors of lower levels of trust in the government and national institutions. These results could be better framed by adding two further variables, that are included in the report (OECD 2022), and that fit well with the reading offered so far. The belief of belonging to the lower social status of the society predicts a distrust towards the government 22.9 times more than financial struggle. Similarly, the belief of not being included in governmental and collective decision making is strongly correlated with distrust. In general, it is the feeling of insecurity that derives from the fear of being "left behind economically, socially and politically" (p. 56), which matters and determines distrusting attitudes. The attributed reasons, associated feelings, and the resulting self-image of oneself as a member of the society, that people infer from their current conditions of struggle, is what determines the emergence of distrusting attitudes. In a nutshell, those who feel ignored and not listened from the decision makers, and from the higher ranks of the society, are those who are more likely to distrust their government and relative national institutions. In the final section of this chapter, I will elaborate also on the epistemic configuration of these mechanisms and try to better define their relationship with the spread of disinformation. At this point, is apt to take stock of the presented results and arguments.

The belief of being socially excluded and left behind, namely currently deprived of the possibility to be part of the activities that characterize one's society, and persuaded by negative expectations about the future, can motivate distrust towards the institutions. The latter can be exacerbated in societies characterised by low levels of social mobility and eventually result in the development of a form of social status-based distrust of the kind described by Desmond (2022). The latter could then motivate the acceptance and sharing of pieces of disinformation. Coherently, data show that in countries where citizens self-report low levels of satisfaction with their democratic systems, they also self-report high rates of encounters with disinformation, which could indicate that, indeed, in these

countries, on average, more pieces of disinformation circulate, or, that citizens tend to be more distrustful towards all the information they come across, to the point that tend to interpret it as disinformation more often than those who live in other countries. (Unzicker 2023). In general, to feel socially disadvantaged and distant from higher social groups, and to firmly believe that one's standpoint is systematically ignored during collective decision-making processes, in combination with the pressure of having decreasing expectations in front, or due to the feeling of being trapped into precarity, is a condition that is very likely to determine the emergence of a sense of socio-political abandonment, and consequently result in a form of resentment. Over time, if nothing changes for the better, resentment could easily turn into distrust of the institutions, including the scientific ones. Therefore, I argue that Desmond (2022)'s idea is well supported and represents an important perspective to consider in a study of the possible determinants of the crisis of trust in science, and of the consequent spread of scientific disinformation. However, it is necessary to integrate it with more data and further considerations.

In particular, it is necessary to analyse the factors that might bridge the gap between the development of a form of socially based distrust in science with the resulting decision to share disinformation. In previous sections, I have mentioned the non-epistemic reasons for disinformation sharing, including motivated and biased reasoning processes. A distrustful attitude towards scientific institutions can be easily exacerbated by group dynamics, trigger motivated reasoning and the other mentioned mechanisms, and be conducive to the acceptance and sharing of disinformation. In the upcoming section, I argue that alongside these non-epistemic dynamics there is also a peculiar epistemic process through which the belief of being socially excluded could lead to the sharing of scientific disinformation.

4.3. Social Exclusion Is Higher Order Evidence

If the belief to be socially excluded is part of the cause of the development of a socially motivated distrust in science, then the awareness of one's social status can have an effect on the evaluation and collection of evidence that comes from scientific institutions. More specifically, I believe that social status constitutes a form of higher order evidence for those who feel socially excluded.

The firm belief of being a member of a social group whose interests are systematically ignored during collective decision-making processes by higher ranked social groups, including knowledge producers and brokers, has severe effects on the state of belief of those who endorse this credence, and particularly on the acceptance of pieces of information that are related to social issues. The possibility that social factors of the kind described in the previous section can be coherently interpreted as instances of higher order evidence is not an undisputed matter. A reading of this kind has been proposed by Levy (2022). In his work, any possible indicator of expertise, trustworthiness, and in general any piece of "evidence about our evidence" (Levy 2022, p. 136) is considered as definable as higher order evidence. As rightfully noticed by Dulith Novaes (2023), the one used by Levy is not a very orthodox notion of higher order evidence, however, as Dulith Novaes argued, this is not a reason to reject Levy's entire argument from scratch. The idea that various factors can represent instances of higher order evidence and, as such, condition people's trust in science is sound and worth exploring.

In particular, I argue that it is possible to accept Dulith Novaes (2023)'s point while defending Levy (2022)'s intuition, namely integrate the latter with a notion of higher order evidence that is more coherent with the one commonly considered in the related literature. Higher order evidence is usually interpreted as evidence about one's evidence, as Levy (2022) claimed, but more precisely, it is evidence regarding one's process of evidence collection, and surely it is not a sort of evidence which is, for some inherent reason, more relevant or important than the other ones. An item of higher order evidence is a piece of knowledge that, once acquired by an agent, rationally compromises the reliability of a considerable amount of evidence at her disposal (Horowitz 2013, Lasonen-Aarnio 2013). As DiPaolo (2018) has argued, this kind of evidence has, inter alia, two important features. First, it has retroactive force, namely, it shows to the agent that the endorsed state of belief is irrational from the moment of its very formation and primary endorsement. Second, higher order evidence is able to immediately disvalue the maintenance of any belief grounded on what turned out to be epistemically compromised evidence. Being aware of the fact that the evidence at disposal is no longer fully reliable can result in a condition that is commonly known as "epistemic akrasìa" (Horowitz 2013, Borgoni 2015). An agent is in an epistemically akratic state of belief if her evidence tells that '*p*' is the case, but an instance of higher order evidence indicates that on the contrary

it is 'not-*p*' to be the case, and hence the most rational proposition to endorse. The result is that, until a decision is taken between *p* or not-*p*, whoever is in an akratic condition, as long as she remains in that epistemic impasse, she is committed to a contradictory state of mind (Munroe 2023). The general point is that, to come across higher order evidence can be a very beneficial event to occur, but as soon as it happens, there arises an epistemically charged urgency to act somehow, to do something with one's state of belief to escape the resulting contradictory condition. Is this more orthodox notion of higher order evidence compatible with social factors like social status or group identity? In other words, is it possible that, for instance, the awareness of one's social status has the described epistemic effect, typical of higher order evidence, on the pieces of evidence collected regarding a certain issue? I believe that it is possible, and the spread of scientific disinformation can be the result of a mechanism of this kind.

The belief, resulting from daily experiences and related feelings, of being part of a social group that is left behind, namely that has its interests, worries, and priorities systematically ignored by decision makers, has a real effect on the state of belief of those doxastically committed to it. Considering those who live in precarious conditions, who struggle with the impossibility to fully develop as members of a society, it seems plausible that their very social condition works as a form of higher order evidence. However, this could be the case only in a specific situation. For a piece of knowledge to constitute higher order evidence it means that this works at the level of evidence collection, pointing out some flaws in that specific epistemic phase. In conditions of social exclusion, the related beliefs make the pieces of evidence deriving from socially higher ranked groups resulting inherently unreliable, since these are believed to be the result of processes that are laden with partisan values and interests that are incompatible with the ones of those who belong to the lower social status. This is a peculiar epistemic state that goes beyond the belief of a lack of benevolence by decision makers or scientists. It coincides with the credence of belonging to two completely different realities, each playing by their own rules, and pursuing their own goals, while following their own values. There is a sort of incommensurability of what is valuable in one condition and what is in the other, making certain sources of evidence inherently unreliable, and this latter informational content, regarding the intrinsic unreliability of the received evidence, is precisely what higher order evidence communicates and signals. For some agents and groups, it is their belief

of being socially excluded that a priori impairs the acceptance of any piece of evidence that has been produced by highly ranked social groups, which, as claimed by Desmond (2022), from some perspectives include the scientific institutions.

Therefore, and this is an important point, it is not necessary for a state of evidence to be poorly exposed to scientifically reliable sources of information to be vulnerable to scientific disinformation. On the contrary, in conditions of believed social exclusion, a broad exposure to scientifically accurate and institutionalized pieces of knowledge renders a condition of epistemic akrasìa very likely to occur. Indeed, it is more frequently that the scientific evidence at disposal, received from various sources, clashes with the higher order one that indicates, to those who believe to be socially excluded, that the collected evidence is the result of processes and interests that are incompatible with theirs, and hence inadequate to guide and condition their decision making. I argue that it is to escape this condition that many individuals end up rejecting scientific evidence, and the easiest way to externalize this decision is by sharing scientific disinformation, an action that works as an emblematic expression not of the rejection of the truth of a specific content, but of the distrust in the institutions who have produced it[22].

This hypothesis is coherent with the already discussed idea that the sharing of disinformation can be part of a process of clustering around a common narrative. In that case, the main aim behind the sharing of disinformation is to signal the adherence to a certain worldview, therefore, the specific content considered and spread plays a symbolic role for the most part. Similarly, I think that the attempt to escape the described akratic condition can motivate the sharing of scientific disinformation. In this case, the attempt to escape epistemic akrasìa is an end in itself, since this consists in a miserable condition, a sort of epistemic paralysis that the agent wants to overcome, but it could also be part of a process of expression of one's state of resentment and perceived social exclusion. Therefore, if my reconstruction is sound, it is possible to see that what sometimes appear as instances of disinformation sharing motivated by social reasons, like signalling one's group belonging or support for a certain position, hence non-epistemic factors, can hide

[22] In this work, I do not explicitly consider the possibility that sharing disinformation may be a form of protest, thus, this claim should not be interpreted as such. However, I consider this to be an interesting and possible line of research which is worth exploring, in particular, given the increasing research efforts invested in the epistemology of protest, see Medina (2023).

also an important epistemic explanation, which is related with the described attempt to escape the condition of epistemic akrasìa that results from the combination of the higher order evidence, that is one's social status, and the received scientific evidence, according to the dynamics described throughout this chapter.

Conclusion of Chapter 4

In this chapter, I have considered the possibility that the current spread of scientific disinformation is caused by a diffused form of distrust in science. In particular, I have discussed the possibility that distrust in science is largely motivated by non-epistemic reasons, including a perceived lack of benevolence on the side of the scientific institutions. I have considered the proposal of Desmond (2022), who argued that large numbers of people distrust scientific institutions in virtue of their members belonging to a higher social status, which is assumed as an indicator of the fact that these pursue goals and interests that are incompatible with those of the socially lower ranked groups. Elaborating on Desmond (2022)'s idea, I have argued that there is a particular group that is likely to develop a form of social status-based distrust in science, and that is composed by those who feel socially excluded and left behind. This notion has been introduced to represent the belief, that many people are currently very likely to endorse, to be left behind, in the sense that their interests, values and worries are largely ignored during collective decision-making processes both by decision makers and knowledge brokers. I have tried to keep this social group and related beliefs, as distinct as possible from those of minority and reactionary groups, who in mutually opposite ways, could be said to argue something similar to the mentioned idea of social exclusion. Indeed, it is possible that those who currently feel left behind, with time, become members of reactionary groups, or solidarise with members of minority groups, namely their feelings and beliefs can change and move in opposite directions. However, I believe that there is a stable and enduring sense of social exclusion, that is peculiar of the current historical moment, and that is characterized in particular by the fact that people are afflicted by negative and discouraging expectations about the future, and by a diffused sense of lack of control and powerlessness regarding it. In the difficult attempt to properly define the features of this social group, I have considered various empirical studies and variables. I have pointed out than in societies where social mobility is low or decreasing, and trust in government,

institutions, and democracy is low as well, disinformation is more likely to proliferate. I have argued that the latter is due to the fact that a diffused socially motivated distrust in institutions could eventually involve also the scientific ones, as Desmond (2022) pointed out, hence determining the spread of a diffused distrust in science. In conclusion, I have argued that the distrust in scientific institutions endorsed by those who feel left behind and socially excluded is not fully grounded on non-epistemic bases. On the contrary, I have claimed that a condition of social exclusion works as a form of higher order evidence that impedes the acceptance of pieces of evidence coming from institutional sources of information.

In this chapter, I have highlighted correlations and argued for positions that require further and ad hoc empirical studies to be better supported. Nonetheless, I defend the idea that lies beneath this chapter and entire thesis, namely that the spread of disinformation, and particularly of scientific disinformation, is primarily a socially motivated phenomenon, and that the set of social variables that are considered as potentially relevant for it should be expanded, within the literature, to include aspects and factors like the ones considered in this chapter.

5. Unexpected Drivers to Disinformation: The Shortcomings of the Communication Between Science and Society

In the first section of this chapter, I briefly review the evolution of the idea of objectivity in philosophy of science. I argue that, currently, objectivity is almost equivalent to unbiasedness, and for this reason scientific knowledge is characterised by two features, inter alia, transparency (section 5.2) and quantification (section 5.3). Focusing on the latter, I point out the communicative benefits of quantified knowledge. In sections 5.4 and 5.5, I consider the epistemic shortcomings that derive from an uncritical use of quantified knowledge in the political discourse, made by weak and distrusted political forces. In particular, my claim is that the latter contributes to the emergence of populist and reactionary forces, and to the loss of knowledge that derives from the diversity of perspectives that characterizes every plural society.

5.1. The Conceptual Evolution of Objectivity

As the main topic of this work shows, not everyone trusts science, and as I have shown in the previous chapter, there are various reasons why different people might be doubtful about the reliability of scientific data. However, it is safe to claim that in absence of strong beliefs or reasons to do otherwise, when people learn a new scientific discovery or datum, they tend to immediately interpret it as a fact, since it is indeed almost inevitable to intuitively associate scientific knowledge to objectivity. Behind this naïve, but nonetheless correct, intuition, there is an interesting and variegate conceptual history. In fact, it would be wrong to think that the idea of objectivity that we currently associate to scientific activities has never changed. For this reason, it is important to clarify the current sense of scientific objectivity and see whether it is possible to discover something useful to better understand the phenomenon of the spread of scientific disinformation.

There is a lively debate in philosophy of science regarding what would be the best way to conceive and accomplish objectivity (John 2021), and what role should scientists' values play in scientific production (Rudner 1953, Jeffrey 1956, Levi 1960, Douglas 2000, Kukla 2017) to avoid inductive risks (Hempel 1965) for instance. There is not enough space here for a complete discussion of all these points, but it is possible to safely

claim that, in current times, objectivity means primarily impersonality and unbiasedness. The idea is that scientific results ought not be influenced or determined by factors that are related to scientists' subjectivities (Koskinen 2018), as well as everything that composes them, like: biases, beliefs, ideas, perspectives, and personal goals. The emergence of a conception of objective knowledge as impersonal is the result of an historical process, imputable to some major changes that occurred in science (Daston 1992, Daston and Galison 2007). The necessity to make scientific knowledge interchangeable, namely easy to communicate between different research groups, has determined the necessity to overcome individualities and standardize scientific products. This need has become even more evident in very recent times, when international research groups started interacting from all over the world, using new communication technologies to exchange data and scientific results.

Another decisive factor that has contributed to the radicalization of the idea of objectivity as unbiasedness is the socialization of science. The idea that scientists have a social responsibility, and that they are bearers of a social role, is a contemporary acquisition, absent during Modern times, at least in the current configuration (Kitcher 2011, Chap. 16). The inextricable connection between science and society, consolidated in the 20th Century, has determined the formation and consolidation of new practices in science. Following Merton (1949)'s reading, the contemporary scientific enterprise is characterised by the values of universalism, communism, disinterest, and systematic doubt. In Merton's work, universalism is used to denote the "impersonal and unbiased character of science" (p. 1060) which strongly remarks the principle that science must be immune to personal interests, desires and biases. To accomplish the latter, complicated task, Longino (1990) argued that it is necessary to constantly control, and make explicit, the background assumptions of any scientific enterprise. According to Longino (1990)'s contextual empiricism, the contemporary connection between science and society is not only defined as tight, but really inextricable:

"Scientific knowledge is, […], social knowledge. It is produced by processes that are intrinsically social, and once a theory, hypothesis, or set of data has been accepted by a community, it becomes a public resource. It is available to use in support of other

theories and hypotheses and as a basis of action. Scientific knowledge is social both in the ways it is created and in the uses it serves" (Longino 1990, p. 75).

The idea of objectivity that it is now commonly associated to science is thus a direct consequence of the described social turn, which sees scientific knowledge a priori as a social product. Science in fact, as any other public activity, needs to be socially supported and justified as beneficial for the society at large. In other words, science must be useful for the collective. Particularly, during and immediately after World War II, scientists have started thinking of their work in terms of social risks and utility, and so did politics. Arguably, the symbolic moment of this shift is the publication of *Science: The Endless Frontier* by Vannevar Bush, in 1945. Here, Bush offers a purely utilitarian argument in defence of science. Scientific activity allows a country to have internal benefits, and also to be competitive in the international market, that is why it must be constantly supported from every point of view.

To take stock of these preliminary considerations, the contemporary configuration of an ideal scientific research is impersonal, unbiased, useful, and socially reflexive. It is important to notice that the evolution and settlement of the latter aspects has had meaningful effects not only in the way in which science is produced, but also in how it is communicated. Scientific results must reflect the described idea of objectivity, social utility, and being shareable with a global scientific community. Therefore, the scientific community has defined common communication systems, languages, and mechanisms of knowledge classification, and started representing scientific data accordingly. For all these reasons, two interrelated principles have majorly influenced the contemporary scientific communication, namely, quantification of scientific data and transparency. As I will argue in this chapter, the communicative relevance of these two principles is primarily related to their ability to foster public support in science and data-based policies. However, for transparency and quantification to accomplish their communicative task, it is necessary to maintain a critical attitude towards their possible effects on the public reception of scientific knowledge. To clarify the latter, maybe cryptical, point consider transparency first, and what could be the implications of an uncritically open and accessible system of knowledge production.

5.2. Transparency

Even if not exclusively focused on scientific production, one of the first voices to point out the possible risks of transparency was O'Neill (2002). In her work, O'Neill pointed out how the necessity to respect specific requirements of transparency might contribute to the constitution of a climate of constant anxiety and suspect towards the institutions, which could eventually cause the formation of a defensive attitude among the members of the institutions, the officials, and the experts who would constantly feel to be evaluated for their work by external agents. More importantly, O'Neill claimed that transparency could not simply determine a change of professional duties but ultimately distort the *raison d'être* of one's profession. If the necessity to comply with rules and requirements of transparency is no longer perceived as an external regulation of the modalities through which one's work should be communicated, but becomes a core and pervasive component of it, something that has a direct influence in the way in which people conceive their work and related duties, then there could be important and negative consequences on the efficiency and quality of the work accomplished. O'Neill (2002)'s analysis has been particularly relevant in this debate, and more recently Nguyen (2021) has considered and expanded it. The starting point of Nguyen's work is the claim that transparency is a form of surveillance, designed to gain public trust in exchange of a constant monitoring of the people and practices that constitute social and public activities. The latter trade, Nguyen argued, comes with a severe cost, namely that the practitioners of the monitored activities "internalize" the surveillance they are subject to. The unfortunate consequence of the internalization of the necessity to cope with the requirements of transparency, and the related feeling of being constantly monitored, becomes part of one's "own reasoning and values" (p. 25). The risk of losing one's profession's *raison d'être*, sketched by O'Neill (2002), appears in Nguyen's analysis as a serious possibility, which is very useful to consider in the context of the peculiar public enterprise that is scientific research.

The main risk related with the possibility that the scientists internalize the requirements of transparency is that this limits their creativity and reduces the variety of pursued research projects. Being aware of the fact that there are precise requirements of transparency to respect, early career scientists or others who struggle to cope with limited time and resources, could end up pursuing only those research enterprises whose results can be smoothly integrated and represented within the boundaries of the current

requirements of transparency. If the latter is the case, the trade between science and society could be represented as follows. Science gets public support and trust, because its work is openly accessible, and society gains the possibility to constantly monitor and evaluate the aptness and public utility of the scientific activities. If my interpretation of Nguyen (2021)'s intuition is correct, the described social trade comes with severe extra costs. The risk of the trade-off between accessibility and trust, which is at the core of transparency, is to take the epistemic component out of the picture, to the point that scientists appear as mere practitioners of predetermined standards of research, who should be trusted on the exclusive basis of their compliance with these norms, and not because of their competences, education and experience. The risk, thus, is a reduction in the relevance of competence in the attribution of trustworthiness to scientists. Moreover, for transparency to be efficient in incentivizing public trust in science, it has to be the case that having access to the methodologies and goals of science is something that actually inspires trust in science among the scientifically laypeople, an assumption that should not be given for granted at all.

As John (2018) has wisely noticed, trust in science might be fragile also as a consequence of the spread of what he calls a "folk philosophy of science" (p. 81), namely a set of diffused wrong beliefs and distorted ideas about how a proper scientific enterprise works. In presence of wrong ideas and perspectives, there is no reason to suppose that having open access to a scientific enterprise would generate public trust in science. On the contrary, the clash between the wrong expectations of someone who is misinformed about science, and the reality of the methodology of scientific research, could even trigger further distrust, which would be very difficult to repair.

A subtle point comes up from this brief introduction to the topic of transparency. In times where scientific objectivity is grounded on the ability of science to be reflexive and able to perform a constant monitoring of the background assumptions of any scientific research, and when the sensitivity for the possible social effects of a scientific inquiry is a core component of scientific research, a constantly critical attitude is necessary among scientists to produce results that could be safely considered objective and well grounded. However, in times where science and society are so deeply entrenched, the public trust in science is dependent on the efficiency of the communication between the two, and on the public and political use of scientific discoveries. Therefore, a critical and reflexive

attitude, that is crucial for scientists, should also be incentivized among those who make use of scientific results to motivate policies and public decisions, otherwise, the scientific effort invested to preserve objectivity would be vain. In the brief discussion made about transparency, it is already possible to see how the naive application and implementation of a communicative strategy uncritically based on transparency could damage both scientists' work and deprive the society of important resources that could derive from scientific research. Considerations of this kind are vital to avoid the settlement of suboptimal dynamics of knowledge production and public communication. The latter point results even clearer considering the literature on the quantification of knowledge.

5.3. Quantification

We can define as quantification the systematic representation of pieces of knowledge and scientific results with numbers and indicators. This practice has various benefits. First, independently of personal education and scientific literacy, numbers allow people to understand the main findings of very complex research processes. Quantification allows also to simplify the design of universal categorization systems for the classification and representation of scientific results, which guarantees to international research centres, spread all over the world, to uniformly interpret and communicate the results of their work. In other words, quantified knowledge works as a common language within a global scientific community.

Another important communicative aspect to consider is that numbers are automatically associated to very specific features. Quantified data are indeed very likely to be considered the outcomes of objective, rational, truth oriented and unbiased processes (Porter 1993; 1995, Desrosières 2015, Nelson Espeland and Stevens 2008). As discussed in 5.1, the current conceptualisation of the idea of objectivity associated with scientific research is primarily related to unbiasedness and impersonality, and quantified knowledge is inherently able to reflect both of these features, namely, to represent scientific data as the result of research processes that are truth oriented, and thus designed to cancel the influence of any bias, peculiar perspective, or goal, that could pollute the neutrality of the collected scientific results. Therefore, to order knowledge within a standardized system of indicators is particularly beneficial for many reasons, and not secondarily from a communicative point of view. Clearly, it would be epistemically

problematic to assume the objectivity of scientific data in virtue of them being represented in a numerical form, namely, to believe that numbers, as such, are impersonal and neutral.

Actually, as the sociology of quantification has shown, the processes of quantification have always been dependent on complex mechanisms that were primarily oriented to social and political monitoring, control, and management (Porter 1995, Hacking 1990). Measurement systems have historically been necessary to turn collective decision making into a mechanism of rational calculation, a process that has significantly involved scientific knowledge as well. The mentioned social turn of scientific research has not only determined the emergence of a form of social reflexivity among scientists, but also the necessity for the latter ones to motivate their research activities in terms of social utility. As previously seen, science must be useful to be financially supported, and hence its activities, and perspective results, must be calculable, and more importantly, commensurable with other indicators of social utility. The history of quantification and of its pervasive communicative role is thus primarily socio-political, and not only for the mentioned reasons. Indeed, the fact that numerical indicators automatically communicate a sense of objectivity, rationality, and impersonality is priceless from a communicative and rhetorical point of view, since numbers allow "to make decisions without seeming to decide" (Porter 1995, p.8). With this evocative formula, Porter (1995) aimed to point out that numbers and indicators allow to present every decision as the result of a calculation, of an objective analysis of the state of things, hence, as completely not arbitrary. This communicative usage of quantified knowledge is harmless in times of trust in democracy and in its systems. However, currently, many socio-political systems are characterised by a diffused popular distrust in democracy and in its institutions. This condition represents the perfect soil for the development of vicious communicative practices within the political discourse, and that involve also the use of numbers and indicators.

5.4. Numbers and Indicators in the Political Discourse

In Chapter 4, I have briefly mentioned the challenges posed by a globalized, connected and constantly changing world, how these negatively impact on people's future expectations and trust in governments. Indeed, representative democracy, hence elected political forces, are questioned, distrusted, and imputed of being unable to deal with the expectations of the civil society and to cope with the current challenges (Crouch 2004,

Landemore 2020). In this kind of socio-political milieu, communication becomes crucial, and numerical indicators turned out to be a very useful communicative tool. In fact, by grounding the decisions taken on numerical data, it is possible to represent them, in particular the most unpopular ones, as necessary, namely, not as the result of arbitrary decision processes but of rational calculations based on the objective state of things. Merry (2016) has efficiently described this point and argued that since quantified knowledge results homogeneous, decontextualized, impersonal, and centralized, it is useful in the political discourse: "[…] to legitimate political decisions as being scientific and evidence-based in a time when politics is questioned." (Merry 2016, p. 4). Therefore, the ability to select what numbers to use becomes crucial, since two different indicators could reflect two completely different states of things. Piketty (2021, Chap. 1) for instance, highlighted the communicative difference that follows from the use of net (NDP) or gross domestic product (GDP) to describe the future expectations of a certain economic policy.

The reiteration of this particular use of numbers in the political discourse leads to a paradoxical situation. As Rose (1991, p. 674) argued, the usefulness of numbers in politics turns out to be related to their ability to "de-politicize" politics, since they work as a guarantee that the discretional power of the decision makers has been minimized. The paradox is evident. In democratic systems, decision makers are elected precisely to make decisions that are impactful for the collective, therefore, the adoption of the described practice appears antithetic with the sense and meaning of their political activity. In analogy with the risks related to an uncritical adoption of practices of transparency in science (5.2), an uncritical use of numbers in the political discourse can compromise the *raison d'être* and work of elected representatives. Clearly, a data-based approach is epistemically desirable in itself, since it allows to deal with very complex situations and design well informed and scientifically supported policies. Similarly, the attempt to limit the discretionary power of the decision makers is a virtuous practice in itself. However, in conditions like the current ones, where democratic political forces are widely distrusted (Quilter-Pinner et al. 2021, Pew Research Centre 2024, OECD 2021) the risk is that this form of "calculative expertise" (Miller 1992; 2001) is uncritically and almost unconsciously set as standard. If this uncritical approach becomes uncontrolled over time, it would be extremely difficult to duly consider the contingent nature of numerical

indicators and categorization systems, and to constantly question their descriptive adequacy and efficiency.

This scenario can be conducive to an important shortcoming. In a context of the described kind, where an uncritical use of numbers in the political discourse has become standard, any piece of knowledge and information that cannot be parametrized into the current categorization system, namely, that does not satisfy the properties of decontextualization and impersonality mentioned by Merry (2016), would be deemed inoperable during collective decision-making processes. Consequently, any form of local knowledge, the acquisition of which depends on the particular standpoint occupied by members of different communities and social groups, can hardly be considered in the political discourse. Newfield (2022) made an analogous and more general point, arguing that: "Quantification tends to strip out history, context, and local knowledges, thus partially falsifying any particular situation." (p. 49).

In complex, numerous, and diverse societies, it is inevitable that some forms of local knowledge are ignored during collective decision-making processes. It is not possible to make a cost-benefit analysis, relative to the implementation of a new policy for instance, without including in the analysis only commensurable variables. Clearly, a commensuration process can be arbitrary and cause categorical mistakes (Nelson Espeland and Stevens 1998), but, by definition, a common measure must be general and independent from contingences like group specific variables. Therefore, given the inevitable fact that knowledge comes in dispersed form, Hayek (1945) argued that the duty of every society is to find the most rational mechanism that limits this dispersion, and that allows to properly optimise the available knowledge. In other words, to make better decisions for the collective, it is necessary to constantly improve from a descriptive point of view, namely, to be able to reconsider and redefine the categorization and quantification systems currently in use, which is a process that requires an unceasing critical approach[23]. I argue that the latter possibility is severely prevented when decision makers do not consider numbers primarily for their descriptive efficiency, but to hide behind them and to make political decisions appear to be necessitated by a supposedly

[23] See Desrosières (2015) and Thernstrom (1987) for two different but emblematic examples of critical approaches to the topic of quantification.

objective reality. When the latter is the case, there are no conditions for critical thinking, and therefore for the redefinition of the indicators and categorization systems in use.

From a political point of view, the resulting picture could be even more merciless when further possible scenarios are considered. In this section, I have argued that, in times where global challenges cause a crisis of trust in democracy, one way to justify political decisions as objective, rational, and impartial is to use numerical data to motivate them. When this reliance on numbers becomes blind and uncritical, the exclusion of various forms of local knowledge becomes systematic, uncontrolled, and irrational. On the long run, the most problematic effect of this mechanism is the settlement of a form of incommunicability between politics and society. In these conditions, when people experience major difficulties or worries that cannot be translated into numbers, they end up being largely ignored or totally overlooked[24]. This form of incommunicability could be conducive to the development of a form of distrust towards the democratic institutions, which could then be exploited by reactionary and populist forces. Concretely, a scenario of the described kind has occurred during the 2016 US Presidential elections. According to Newfield (2022), the Democratic Party, in the years before the mentioned elections, had developed a tendency to overlook "the qualitative, personal, local, community-based experience that said the Democrats' model wasn't working" (p. 63), which was paired with a naturalizing attitude in international economy. During times of financial struggle, as the ones that have followed the 2008 subprime mortgages crisis, the Democratic Party's communication was largely oriented to point out that negative macroeconomic events were normal in a globalized economy. The latter consideration might have been correct, but hardly functional in mitigating the anger and anti-system feelings that started spreading within the public debate, and particularly among those who experienced major financial losses. The described communicative scenario has represented the perfect ground for the emergence of populism and particularly for the proliferation of an anti-

[24] With this respect, it is emblematic the exchange, reported by Newfield (2022), between President Obama and a representative of Steelworkers Union. Newfield highlights how President Obama, in response to statements that reflected the problems and fears related to missing job opportunities, replied "by invoking the quantitative fact of its opposite, a manufacturing boom" (Newfield 2022, p. 61). The point is not whether the "boom" was real or not, but that, since the experience reported by the representative of the steelworkers was not directly reflected by numbers and indicators, it resulted impossible to have a conversation about it, even with a political interlocutor who was careful of the claims of workers, as President Obama was.

intellectualist attitude, maximally represented by the Republican candidate Donald Trump, who, according to Newfield (2022, p.63), has been able to personify the desire to overcome experts' data and opinions that was present in part of the North American society. This political event has marked an important milestone also from a socio-epistemic point of view, since it has contributed to settle the ground for the proliferation of disinformation in the US based debate, and consequently in the rest of the world. In the next subsection, I defend this point.

5.4.1. From Quantification to Disinformation

Müller (2023) argued that populist ideologies often endorse the principle of epistemic optimism. This principle states that, within the public and political discourse, common people have the epistemic competence to discern the truth due to their proximity to every day's life. It is an inherent component of many populist rhetorics to claim that the knowledge that derives from daily experiences is more reliable than the experts' one, due to the mentioned epistemic optimism, and also because, from the same populist perspectives, academic elites, scientific institutions and knowledge brokers are corrupt, and not really committed to the discovery of truth (Mede and Schaffer 2020). The risk related to the proliferation of this kind of populist claims is not only to impair people's trust in scientists, but to compromise the perception of the epistemic status of experts, since it allows scientifically laypeople to claim that their personal experiences are no less valuable pieces of knowledge than the ones deriving from scientific institutions. In a context where the described populist claims proliferate, and where anger and resentment arise, it is possible that epistemic optimism enables many people to vindicate the relevance and importance of their knowledge, which they consider to be systematically set aside within the public discourse.

The development of the US political debate in 2016, and the events that took place immediately afterwards, are coherent with the described epistemic dynamics. As Chatterjee (2022) has noticed, in the US, the far-right populist movement developed for various factors, including the desire to react against the incommunicability with part of the institutions, that was the result of the described number-based political discourse. The paradoxical turn of this political process was that part of the far-right supporters and representatives themselves started using numbers and indicators to defend their theses

and ideas. Moreover, they used indicators of a particular kind that Chatterjee (2022, p. 40) has defined "plebiscitary". During the 2016 campaign, candidate Trump, and many of his supporters, publicly used various numerical indicators with the specific intent to highlight the candidate's popularity, closeness to the will of the people, and his ability to properly understand their needs. Thus, every number that could have worked as an indicator of public support was used as a plebiscitary number: "[…] rally size, referendum and poll results, Twitter followers, television ratings, and even the stock market" (Chatterjee 2022, p. 40).

At this point, it is possible to notice that the political and public debate, that resulted from the combination of all the mentioned factors, represented the perfect soil for the spread of false and misleading information, for two interrelated reasons.

On the one hand, the picture described in this section suggests that part of the US society has been persuaded by populist and anti-intellectual claims. These have determined the development of a form of distrust in science grounded in the belief that scientific institutions do not lack only in benevolence, as discussed in Chapter 4, but also in competence, and that direct experience and proximity to everyday life are sufficient to make correct decisions. Therefore, the proliferation of scientific disinformation within far-right groups might have been more epistemically motivated than usually thought, and particularly, it could have been motivated by the belief that one's scientific competence is by no means inferior to that of the experts. From the perspective of the supporters of far-right populist positions, the combination of the assumed moral corruption of experts and scientists with the idea of epistemic optimism made both distrust in science epistemically legitimated and also the use of plebiscitary numbers. These have been used to contrast both scientific data and the claims of their opponents, in a way that reflects an attitude to consider one's personal experience as the most reliable sources of knowledge available (Zoonen 2012).

On the other hand, the inclusion of plebiscitary numbers inevitably polluted the public debate. In an environment where anything that works as an alleged indicator of a candidate's public support is shared, contents' accuracy becomes broadly ignored and almost irrelevant from a communicative point of view. Therefore, large quantities of news are shared, and groups that are not committed to information accuracy become vulnerable to disinformation, or even creators of it. In fact, not only candidate Trump himself, but

also his supporters have reportedly been a community of heavy spreaders of disinformation and fake news (Alcott and Gentzkow 2017), and in previous chapters I have showed in multiple occasions that, currently, US conservatives, and particularly Trump's supporters (Pretus et al. 2023), are generically more likely to accept and share disinformation on social media (Pereira et al. 2021), and to trust anti-scientific sources of information (Guess et al. 2020). I believe that this intrinsic tendency to share more disinformation than other political supporters is dependent on the described mechanisms, and thus, it can be rightfully interpreted, at least in part, as an indirect effect of the suboptimal and uncritical use of numbers made in the political discourse.

In this section, I have argued that in times of distrust in the democratic systems, the political class may react by radicalizing its trust in numbers and uncritically basing its communication on numerical data and indicators. The loss of local knowledge that inevitably derives from it, if uncontrolled, could have major socio-political and socio-epistemic consequences including the settlement of a form of incommunicability between certain social groups and the institutions. This condition, if prolonged, can facilitate the emergence of anti-system rhetorics and populist forces. These would then exploit people's resentment and foster the emergence of anti-intellectual attitudes which could be conducive to the spread of false and misleading contents. This is clearly the result of vicious practices of discontent exploitation by interest groups, but also the consequence of the fact that those with relevant, but unquantifiable, experiences end up feeling marginalized if the institutions are unable to communicate with them. The problem is that, as populist movements grow, the democratic forces will feel even more questioned and unstable, and as a consequence of this, they may furtherly radicalize their trust in numbers. The result is a political debate waged with indicators that comes at the expenses of the epistemic quality of the shared contents. In the next section, I argue that the described shortcomings are related to a more general one, namely the systematic loss of the diversity of perspectives that compose the society. This loss is caused by the fact that an uncritical trust in numbers has the main purpose of limiting the emergence of conflicts, and even of the kinds that are fruitful for democracy.

5.5. The Epistemic Benefits of Diversity

Considering the issues presented in the previous section from a slightly different perspective, it is possible to notice that the communicative use of numbers and quantified knowledge, that part of the literature on quantification has criticized, is ultimately finalized to the mitigation of conflicts within the society. Rational decisions, made on the basis of numerical indicators, should indeed limit the emergence of conflicting attitudes, since data-based decision processes are supposed to be automatically associated to objectivity, rationality and neutrality. However, certain forms of conflict are actually very beneficial in every democratic system. More specifically, I argue that it is during conflicting interactions that the society benefits from the diversity of perspectives that inhabits the public and political space (see also Muldoon 2018). It is conflict that allows different forms of local knowledge to emerge, and to be considered, and possibly included, within the public debate. Of course, it is a clear responsibility of the ruling forces to limit violent and harmful forms of conflict, but not those that allow different values, worldviews and experiences to be expressed and publicly defended. These kinds of interactions, where diversity of perspectives arises, are the ones that Sunstein (2017) considered crucial for democracy and that as such must be preserved also within the digital environment. The risk that I aim at pointing out is analogous to the one mentioned in section 5.4, namely that if marginalized and ignored perspectives find no space of expression within the boundaries of the democratic discourse, they might find an outlet in anti-democratic spaces and, as a consequence of this, become vulnerable to anti-democratic ideas. An uncritical trust in numbers, and in their communicative effects, is one of the factors that could cause this unfortunate outcome. To efficiently elaborate and defend this idea, it is necessary to motivate two claims. First, that a diversity of perspectives is beneficial for the society, and second, that moments of conflict are necessary for the emergence of the benefits related to diversity.

5.5.1. Group Decision Making Benefits from Diversity

Various empirical results have shown that diversity improves collective decision-making processes and the accomplishment of goal-oriented tasks. Scholars have collected empirical results of this kind even considering very different notions of diversity,

including demographic diversity (Watson et al. 1993, Levine et al. 2014) diversity of skills (Alaoui et al. 2022), and diversity that derives from randomization among people with different competences and abilities (Hong and Page 2004). In the reported studies, the benefits of diversity resulted within performative contexts, namely during experiments where it was tested group's decision making or ability to accomplish a certain task. However, it is important to point out, as Steel et al. (2021) have done, that the benefits of diversity do not arise only in contexts when the diverse standpoints included in a group, and the related competences and skills, are germane to the task at hand. In other words, diversity is beneficial for collective decision making, independently from the goal to accomplish. Steel et al. (2021) have argued that what makes diversity so beneficial for group reasoning is that this improves the collective "information elaboration process". In particular, on the one hand, diversity forces the members of a group to pay more attention to others' claims and to "assess information more carefully" (Steel et al. 2021, p. 1294). In homogeneous groups, interactions are more automatic, meaning that each member is more likely to assume that the others share certain experiences and views with them. On the contrary, in a diverse group, there is no reason to assume that a certain idea or experience, that is typical of one's particular group or category, is equally shared by everyone. Actually, this is very unlikely to be the case. On the other hand, in homogenous groups, members are more reluctant to express their dissent, or arguing against a common position, since homogeneity is also related with higher levels of trust and reliance in others (Steel et al. 2021) whereas the latter dynamic is less likely to occur in a heterogeneous group. These two aspects highlight the fact that, within internally diverse groups, each member can give very few beliefs, ideas, and experiences for granted. As a result, interactions and exchanges of ideas require more effort from each member than under conditions of group homogeneity. This hypothesis, defended by Steel et al. (2021), is both simple and counterintuitive. Indeed, the benefits of diversity, in terms of decision making, appear related to the fact that diversity causes a preliminary difficulty to jump to conclusions, since it forces everyone to consider a more variegate set of perspectives, and related pieces of knowledge. In other words, diversity is beneficial for collective decision making because makes the entire process more complicated, at least during the preliminary phases of interaction and exchange of information (see also Zollman 2010 for a similar conclusion). Steel et al. (2021) argued

that the accounts of diversity that they have considered failed to reach their same conclusion because most of them were reluctant to accept the idea that the benefits of diversity come at the price of clashes of views and perspectives that often, but necessarily, Steel et al. (2021) claimed, are conducive to forms of intragroup conflicts.

This is a very realist perspective that, nonetheless, in the configuration proposed by Steel et al. (2021), is defined considering ideal conditions of group deliberation, namely interactions during which every member is both allowed and willing to put their knowledge at disposal of the community, and every contribution is gathered and equally considered during decision-making processes. I argue that it would have been much more appropriate to discuss this proposal within a different framework, namely a non-ideal one, which could have allowed to point out further relevant aspects. As Shavit (2021) has argued, in non-ideal conditions of group interaction, the mere presence of demographic diversity within a collective is not sufficient for the emergence of any benefit. The point defended by Shavit (2021) is that dispersed knowledge, once manifested, needs to be stimulated, and work needs to be done to constantly integrate it within a certain group or community; otherwise, there is no reason to suppose that diversity alone would spontaneously and continuously stimulate the group to produce new pieces of knowledge, and consider anyone's contribution as equally valuable. On the contrary, in conditions of marginalization and discrimination of certain groups, it is possible that interactions among diverse members generate forms of epistemic injustice and silencing, or reinforce already existing ones, that impede the production and collective use of certain pieces of local knowledge, namely the ones that the marginalized groups could have offered (Shavit 2021, p. 529).

Therefore, considering non-ideal conditions of group interaction, it is possible to argue that the mere presence of a heterogeneous composition of members is not sufficient to ameliorate group thinking and collective decision making. Moreover, even if Steel et al. (2021)'s account is grounded on the idea that a preliminary conflict of views is necessary to benefit from group diversity, I argue that their proposal fails to fully recognise the crucial *epistemic* role that conflict plays in making the benefits of diversity to emerge, and again, this depends on their ideal assumptions about group interactions. The members of a collective, whose voices remain unheard or ignored during collective decision-making processes, can only overcome this condition of marginalization by

pointing out the relevance and epistemic value of their perspectives, which is an action that often requires going through moments of openly expressed dissent. In other words, in conditions of diversity, where certain perspectives are silenced or ignored, conflict has an epistemic role, namely opening a breach within the existing conditions of interaction, through which marginalized perspectives can point out their knowledge, experience, and related relevance. In the next section, I defend this view on the relevance of conflict for the emergence of the epistemic benefits of diversity. If this argument is correct, then it is clear why uncritically preventing the emergence of every form of conflict within the society has suboptimal consequences, namely, because it would also prevent the occurrence of those forms of conflict that are necessary for marginalized voices to vindicate their position and relevance, also epistemic, in the society.

5.5.2. Diversity Needs Conflict

To better state and defend the point discussed at the end of the previous subsection, it is useful to consider some remarks that Anderson (2006) made regarding Agarwal (2000; 2001)'s research with local communities, based in India and Nepal, called community forest groups (CFGs).

In these communities, due to the gender-based norms that regulate the division of labour, women have overseen the collection and conservation of wood for a very long time. When new social practices were introduced for the latter tasks, the women who had always been in charge of them, considered the new practices inefficient. Anderson (2006) pointed out that, since women were not allowed to take part in the communities' assemblies and decision-making processes, the entire community struggled to include women's knowledge and consequently failed to re-evaluate the newly implemented practices of wood's collection. It can be easily noticed that the exclusion of women's perspectives, from the ones considered during decision-making processes, constituted a form of social and epistemic injustice that was not only ethically problematic, but that indeed made the entire community losing the knowledge that came from their experiences and competence.

To point out the recognized inefficiencies, with respect to wood gathering and storage, it was necessary for the women of the CFGs to contrast their exclusion from the collective moments of discussion and decision making. In some of these communities,

women publicly dissented and gained important accomplishments, including the possibility to attend assemblies and finally put their competences at disposal of the whole community. The important point, that Anderson (2006) rightfully highlighted, is that these results were possible because, through dissent, women gained a space, a public platform, from which they pointed out not only the injustice of their exclusion, but particularly the severe loss of knowledge that would occur if it continued. Without this kind of risky and dangerous moments of public dissent, it is impossible for silenced groups to make the rest of the society realising the unfairness of their marginalization, and the potential fruitfulness of their contribution to the common decision-making processes, particularly when these aim to be democratic.

In section 5.4, I have mentioned some of the reasons behind the use of numbers in politics, included their ability to make a decision-making process looking unbiased, objective, and rational. These communicative features are highly valued by questioned and distrusted political forces, because they can limit the emergence of public objections to the decisions taken and hence reduce the situations of conflict between the political class, that would struggle to handle them, and the dissenting members of the society. However, in this section, I have argued that, in diverse societies, the occurrence of certain forms of conflict are propaedeutic to the public fruition of the epistemic benefits that come from the heterogenous perspectives of their members (see also Bohman 2007). Therefore, the loss of local knowledge, that is an inevitable consequence in a communication system based on quantification, can also be attributed to something other than the impossibility of taking into account those experiences that are not reflected in numerical indicators. Indeed, it is possible that questioned political forces, by settling an uncritically number-based political discourse, limit the emergence of those conflicting exchanges that are necessary in a diverse society, and consequently prevent the emergence of heterogeneous perspectives, and related pieces of knowledge, within the democratic debate. Then, these unexpressed conflicting attitudes can be exploited by anti-democratic forces, and lead to the consequences described in section 5.4. Therefore, an uncritical use of numbers within the political discourse can trigger a chain of effects that include the systematic loss of diversity of perspectives and competences that compose the society, and to the consequent emergence of resentment and social anger.

Conclusion of Chapter 5

In this chapter, I have indicated some possibly unexplored paths to scientific disinformation, namely factors that are not commonly considered in the literature as possibly influential in the development and spread of disinformation. In particular, I have focused my attention on the risks related with the settlement of a viciously uncritical use of quantification in the political discourse. I have argued that, when the latter occurs, a form of incommunicability arises between the political class and the social groups who have unquantifiable pieces of knowledge to share. This incommunicability can make some groups feel politically marginalised, to the point where they become vulnerable to populist and anti-intellectual claims. This paves the way to the proliferation of disinformation, as happened during the 2016 US Presidential elections, following the dynamics that I have described in 5.4 and related subsections.

In section 5.5, and related subsections, I have elaborated more on the idea that an uncritical approach to quantification could prevent the society from benefitting of its diversity and heterogeneity of perspectives. The latter occurs because the described use of numbers is uncritically conceived to prevent all forms of conflict from arising, including those that are necessary for the emergence of marginalised or ignored perspectives. Therefore, if socio-political conflicts are not mediated within the democratic space, they will naturally find an outburst in non-democratic, or even anti-democratic contexts, leading to the dynamics described in the previous sections.

My overall conclusion is that an uncritical and blind use of numbers in politics limits the possibilities of dialogue, dissent and the necessary forms of conflict within the society. The case is the same for transparency. As seen in the dedicated section (5.2), an uncritical approach to transparency, particularly in science, can have harmful consequences for the scientists themselves, and compromise the production of results that would be beneficial for the society at large. In analogy with the discussion on quantification, a questioned and distrusted political class would run the risk to conceive transparency as the trade-off between monitoring and trust that Nguyen (2021) argued for, and which reflects the same dedication to the suppression of every form of conflict seen with respect to quantification. To restore moments of fruitful discussion, and even conflict, politics should vindicate its *raison d'être* and manage dissent in a rational (Hayek 1945) and just way. To do this, it is necessary, inter alia, to recover a critical

attitude towards the use of numerical indicators in the political discourse. A communication characterised by a transparent and number-based approach is clearly fruitful, but not when it is motivated solely by the fear of losing consensus and the desire to hide behind numbers, and hence “de-politicize” the work of democratically elected figures.

Conclusive Remarks

In this thesis, I have analysed the spread of disinformation, particularly scientific disinformation, online and on social media. I have considered the possible reasons and causes of this phenomenon that are already discussed in the literature and proposed new and original ones.

In Chapters 1 and 2, my goal was to offer a broad and detailed picture of part of the literature on disinformation, framed as a socio-epistemic issue. Particularly, in Chapter 1, my goal was to define disinformation and take a position in the debate regarding its conceptualization. In this chapter, I have argued that a propositional content is a piece of disinformation if it has been spread with misleading intentions, namely as part of an attempt to convince the audience of something false. I have also claimed that it is not necessary that a piece of information is factually false to be defined as disinformation. A factually true content can be disinformative when shared with a specific audience that is very likely to derive false implicatures from it, or if shared in a decontextualized and misleading way. In this chapter, I have also argued that the misleading intentions of the creators of disinformation are inherently related to considerations regarding the attitudes and beliefs of the targeted audience. My general conclusion is that the spread of disinformation is not primarily the result of people's gullibility, or of the casual dissemination of large amounts of false contents, but it is the result of the exploitation, by the creators and spreaders of disinformation, of the possibilities offered by the digital environment, in the attempt to target those groups that are more likely to accept the received contents as true and behave accordingly. For this reason, a complete analysis of the phenomenon of disinformation requires to systematically take into account the attitudes and beliefs of the audience targeted by the producers of disinformation, since these condition the process of elaboration and creation of the contents of disinformation.

In Chapter 2, I have analysed the dynamics through which disinformation proliferates. Several scholars have argued that the digital environment, and social media platforms in particular, have contributed significantly to the spread of disinformation. This is because online it is easier to share contents with large numbers of users, and because the current mechanisms of content recommendation tend to segregate people into

groups of like-minded users, commonly defined as echo chambers. In this chapter, I have defended an intermediate position. On the one hand, I have recognised that the communication within social media platforms is designed in a way that facilitates the proliferation of inaccurate and misleading contents, and also the formation of segregated groups that could easily radicalise and polarise. However, on the other hand, I have also pointed out that empirical results show that segregation online is not automatic or inevitable, and that the vast majority of social media users have balanced information diets. Therefore, I have claimed that the digital environment is a fertile soil for the proliferation of disinformation, but its spread is decisively dependent on people's decision making and cannot be fully imputed to the current mechanisms of content recommendation. To support this claim, in the second part of this chapter, I focused on the topic of echo chambers. I have argued that their use in literature has been confusing, and that their meaning has been poorly defined. Therefore, I have proposed a new definition of echo chambers as epistemically vicious and irrational decisional patterns.

In Chapters 1 and 2, I have reviewed various positions, and considered the possible reasons, causes, and dynamics of the spread of disinformation that are already discussed in the literature. Differently, in the rest of the thesis, I have introduced other possible explanations for the spread of disinformation and considered further related risks that are currently underexplored.

In Chapter 3, my aim was to point out that the proliferation of disinformation online can be caused by factors that have been largely ignored in the literature, and which can have severe long-term effects that, similarly, have not been properly considered. More specifically, I have argued that, even if empirical results show that disinformation sharing is predicted by social and group identity protective attitudes, it would be a mistake to ignore the possibility that it is also related to identity formation processes. By taking the latter possibility into account, I have been able to look at the act of sharing disinformation from an original perspective, namely as a form of cooperation. This reframe allowed me to argue that, in certain cases, disinformation sharing is not motivated by social identity protective attitudes, but by their very opposite, namely group disidentification, conflict, and consequent social identity formation.

In Chapter 4, I have considered the possibility that the proliferation of scientific disinformation is as a consequence of the fact that large numbers of people currently

distrust science. I have focused on the possibility that those who distrust science, and consequently are more vulnerable to accept and share scientific disinformation, are in large majority motivated by non-epistemic factors. More specifically, I have argued that there is a social group, which is composed by those who feel socially excluded, left behind, and who have negative expectations about their future, that are likely to develop a form of socially motivated distrust in science. In support of this claim, I have combined different empirical results to point out that in societies that are highly exposed to disinformation, social mobility is low and people manifest distrust towards the democratic system and related institutions, which could include also the scientific ones. The picture described in Chapter 4 needs future ad hoc empirical studies to be better tested, since for the moment it is based on correlations, but still, I think that it points in a fruitful direction, namely one that considers the spread of disinformation as a useful indicator of hidden socio-epistemic issues that must be faced. To conclude this chapter, I have claimed that those who feel socially excluded may also have epistemic reasons to spread scientific disinformation. In particular, I have argued that one's social condition can work as a form of higher order evidence that impedes the acceptance of information coming from certain sources, which in some cases include also the scientific ones.

In Chapter 5, I have considered the possibility that the spread of scientific disinformation is related to possible inefficiencies in the current scientific communication, and conversely in the communication between science and society. I have considered two specific characteristics, transparency and quantification, and focused mostly on the latter one. In the rest of the chapter, I have argued that the spread of scientific disinformation has been indirectly facilitated by the exploitation of the communicative features of numbers and indicators within the political discourse. In this chapter, I aimed at showing that suboptimal practices and uncritical attitudes endorsed by members of the political institutions, that do not appear to be directly related to the spread of disinformation, can in fact be conducive to it. More generically, these can also determine the development and consolidation of anti-scientific and anti-democratic forces within the society, and to other suboptimal consequences that are not only political, but also epistemic.

I believe and hope that, throughout this thesis, I have pointed out some interesting and useful aspects, and it seems appropriate to conclude it by mentioning some possible future scenarios and related research perspectives.

First, it is important to stress that the current digital environment may change in the near future, and for the worst. Thus, the positions that I have defended in Chapter 2, and the empirical results that I have mentioned, could rapidly become pointless and outdated. In particular, it is important to avoid that people are automatically clustered into groups of like-minded users as soon as they start sharing information on a social media platform. Indeed, if the wrong decisions are made, both by governments and social media companies, the claims and fears of many scholars regarding the proliferation of echo chambers, and the formation of an inherently polarising digital environment, may turn into reality in the near future. For this reason, it is important to support the research on the topic of online segregation and keep on stressing that, for the safety of every democratic system, it is crucial to preserve and improve the heterogeneity of the positions and views that the users are exposed to on social media. Alongside, a critical research approach must be preserved to identify the suboptimal, uncritical, and outdated practices that jeopardise the communication and the delicate relationship between the civil society and the institutions, and particularly between science and society, as in the case discussed in Chapter 5.

In conclusion, it is important to point out that, even if the spread of scientific disinformation is a problematic phenomenon that should be studied with the goal of neutralizing it, in this work, I have argued that it also indicates something more than the mere existence of interest groups who purposedly pollute the public debate and compromise people's acceptance of scientific claims. Indeed, the spread of disinformation, and in particular of scientific disinformation, is possible because of the activity of common people who, for various reasons, are motivated to share false and misleading contents with others. Therefore, the spread of disinformation can also be interpreted, and consequently used, as an important indicator of hidden social dynamics and intergroup conflicts. In fact, sharing disinformation seems to be a way, for large numbers of people, to manifest their anger, fear, distrust, and disappointment towards the institutions, but also a way to feel connected to a community or to a social group. For instance, in this work, I have argued that those who believe to be left behind and socially

excluded are more likely to share disinformation, for both epistemic and non-epistemic reasons (Chapter 4). This is a hypothesis that, as said, needs further empirical work to be better defined and supported, but again, I endorse both the general conclusions drawn and, particularly, the research approach advocated in that chapter. My aim, in Chapter 4 and in general throughout this thesis, was to go beyond the hypothesis that the spread of scientific disinformation is related to a diffused distrust in science, or that it is caused by biased and motivated reasoning. I firmly believe that it is time for this literature to investigate the deeper socio-epistemic and socio-political dynamics that might be involved in the spread of disinformation. In particular, it is necessary to consider issues like the one discussed in Chapter 3, namely the possibility that what are often considered instances of motivated reasoning are indeed the result of social identity formation processes. By following a research approach of this kind, it would be possible to pursue important and long-term oriented research projects, and also to re-interpret already collected results in a new light. To offer a concrete example, the case of the disinformation on climate change is perfect. The available results show that certain groups tend to share disinformation on climate change more than others, and also to be more sceptical about climate science. These are, in the vast majority of cases, conservatives, voters of conservative parties, and endorsers of a free market ideology. However, it is possible that to the spread of disinformation about climate change contribute also those who believe to be socially excluded and left behind. In particular, since climate change can be easily, but still wrongfully, interpreted and narrated as a long-term issue, those who currently feel left behind and socially abandoned may be inclined to react negatively to the investment of public effort in preventing what they think, or have been convinced to be, a long-term problem. Thus, in response to what is interpreted as a further piece of evidence of their social condition, they may become vulnerable to scientific disinformation about climate change. Consequently, following the dynamics described in Chapters 3, 4 and 5, as some of them start sharing disinformation, their social identity changes accordingly, and with time, they may end up joining anti-scientific positions and related groups. Therefore, for large numbers of people, the primary reason for climate scepticism and for a propensity to share disinformation may not be related to political affiliation, but as both the exposure to disinformation and its sharing progressively modify social identities, partisanship and group belonging may later become the main predictors of the described attitudes. Thus,

from a research perspective, it is crucial to investigate the possibility that the spread of disinformation is partially caused by the fact that, for certain groups, to share disinformation works as an outburst of a deep socio-epistemically and socio-politically motivated resentment. With time, this mechanism may lead to the modification of entire societies, since large numbers of people can be attracted by anti-scientific and anti-democratic forces who try to exploit their resentment. This research approach would allow both to unveil unconsidered social issues that are urgent to face, and to understand to what extent the spread of disinformation is conceived and designed not only to condition people's decision making regarding specific issues, but at a deeper level, to manipulate the formation of their political sophistication (Christiano 2022) and their relationship with the democratic and scientific institutions.

In other words, to conclude this thesis as it started, it is urgent to better delineate the features of the "breeding ground", to quote Bloch (2014) one last time, that are socio-epistemic and socio-political in nature, and that represent the condition of possibility for the spread of false and misleading contents that are dangerous for large numbers of people, and for our planet.

References

Abenathy, P. (2018) Report: The Expanding News Deserts, Center for Innovation and Sunstainability in Local Media, UNC School of Media and Journalism. Available at: https://www.cislm.org/wp-content/uploads/2018/10/The-Expanding-News-Desert-10_14-Web.pdf

Agarwal, B. (2000) Conceptualizing environmental collective action: Why gender matters, *Cambridge Journal of Economics*, 24, pp. 283–310, https://doi.org/10.1093/cje/24.3.283

Agarwal, B. (2001) Participatory exclusions, community forestry, and gender: An analysis for South Asia and a conceptual framework. *World Development,* 29(10), pp. 1623–1648, https://doi.org/10.1016/S0305-750X(01)00066-3

Aklin, M., Urpelainen, J. (2014) Perceptions of scientific dissent undermine public support for environmental policy, *Environmental Science & Policy*, 38, pp. 173-177, https://doi.org/10.1016/j.envsci.2013.10.006

Alaoui, L., Janezic, K.A., Penta, A. (2022) Coordination and sophistication. BSE Working Paper 1372. http://bse.eu/research

Alcott, H., Gentzkow, M. (2017) Social Media and Fake News in 2016 Elections, *Journal of Economic Perspectives*, 31(2), pp. 211–236.

Alonso, W., Starr, P. (1987) *The politics of numbers*. New York: Russell Sage Foundation.

Altay, S., Berriche, M., Acerbi, A. (2023) Misinformation on Misinformation: Conceptual and Methodological Challenges, *Social Media + Society*, pp. 1-13, https://doi.org/10.1177/20563051221150412

Anderson, E. (2006) The epistemology of democracy. *Episteme* 3(1-2), pp. 8–22. https://doi.org/10.3366/epi.2006.3.1-2.8

Angelucci, C., Prat, A. (2024) Is Journalistic Truth Dead? Measuring How Informed Voters Are about Political News, *American Economic Review*, 114 (4), pp. 887–925**,** https://doi.org/10.2139/ssrn.3593002

Ash, E., Galletta S., Hangartner, D., Margalit, Y., Pinna, M. (2024) The Effect of Fox News on Health Behavior During COVID-19, *Political Analysis*, 32(2), pp. 275-284.

Ausili, L., Martini, C. (2023) Incertezza e trappole di povertà, Sistemi Intelligenti, 1/2023, pp. 105-126, DOI:10.1422/106989

Avin, C., Daltrophe, H., Lotker, Z. (2024) On the impossibility of breaking the echo chamber effect in social media using regulation. Sci Rep., 14(1), 1107. https://doi.org/10.1038/s41598-023-50850-6

Baghramian, M., Croce, M. (2021) Experts, Public Policy and the Question of Trust. In Michael Hannon & Jeroen de Ridder (eds.), *The Routledge Handbook of Political Epistemology*, pp. 446-457. New York: Routledge.

Baines, D.L., Elliott, R.J. (2020) Defining misinformation, disinformation and malinformation: An urgent need for clarity during the COVID-19 infodemic. Available online at: https://repec.cal.bham.ac.uk/pdf/20-06.pdf

Bakke, A. (2020), Everyday Googling: Results of an observational study and applications for teaching algorithmic literacy. Computers and Composition, Volume 57, pp. 1-16. https://doi.org/10.1016/j.compcom.2020.102577

Bakshy, E., Messing, S., Adamic, L.A. (2015) Exposure to ideologically diverse news and opinion on Facebook, *Science*, 348 (6239), pp. 1130–1132. https://doi.org/10.1126/science.aaa1160

Ballardini, B. (1995) *Manuale di Disinformazione*, Roma: Castelvecchi.

Ballew, M., Marlon, J., Goddard, E., Carman, J., Verner, M., Rosenthal, S., Maibach, E., Kotcher, J., Leiserowitz, A. (2024) Public understanding of climate change has grown in the US. Yale University and George Madison University. New Heaven CT: Yale Program on Climate Change Communication. Available at: https://climatecommunication.yale.edu/publications/ccam-explorer-2023/

Banerjee, A.V., Duflo, E. (2011) *Poor Economics*, New York: Public Affairs.

Banfield, E.C. (1958) *The Moral Bases of a Backward Society*, New York: Free Press.

Barbera, P. (2020) Social Media, Echo Chambers, and Political Polarisation, In: Persily, N., Tucker, J. A. (eds.), *Social Media and Democracy,* pp. 34–55, Cambridge: Cambridge University Press.

Bastos, M., Farkas, J. (2019) "Donald Trump is my president!": The internet research agency propaganda machine, *Social Media and Society*, 5(3). https://doi.org/10.1177/2056305119865466

Batelaan, K. (2022) 'It's not the science we distrust; it's the scientists': Reframing the anti-vaccination movement within Black communities. Glob Public Health, 17(6), pp. 1099-1112. doi: 10.1080/17441692.2021.1912809.

BBC (2024), *Timeline of how online misinformation fuelled UK riots*, available at: https://www.bbc.co.uk/bitesize/articles/zshjs82

Begby, E. (2021) Evidential Preemption, *Philosophy and Phenomenological Research*, 102(3), pp. 515-530, https://doi.org/10.1111/phpr.12654

Begby, E. (2024). From Belief Polarisation to Echo Chambers: A Rationalizing Account. *Episteme*, *21*(2), pp. 519–539, https://doi.org/10.1017/epi.2022.14

Benkler, Y., Faris, R., Roberts, H. (2018) *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*, Oxford Academic, https://doi.org/10.1093/oso/9780190923624.001.0001

Bennett, W. L., & Livingston, S. (2018). The disinformation order: Disruptive communication and the decline of democratic institutions. *European Journal of Communication*, *33*(2), pp. 122-139. https://doi.org/10.1177/0267323118760317

Bergamaschi Ganapini, M. (2021) The signalling function of sharing fake stories, *Mind and Language,* 38(1), pp. 64-80. https://doi.org/10.1111/mila.12373

Betsch, C., Sachse, K. (2013) Debunking vaccination myths: Strong risk negations can increase perceived vaccination risks. *Health Psychology*, 32(2), pp. 146–155. https://doi.org/10.1037/a0027387

Bizel, G., Singh, A. K. (2023) Political Polarisation, Misinformation, and Sentiments: A Social Media Analysis About 'Capitol Hill 2021 Attack' by Tweets, *Journal of Social, Humanities and Administrative Sciences*, 9(61), pp. 2257-2266. DOI: http://dx.doi.org/10.29228/JOS HAS.64755

Bloch, M. (2014) *La Guerra e le false notizie*, Roma: Fazi Editore

Bohman, J. (2007) Political communication and the epistemic value of diversity: Deliberation and legitimation in media societies, *Communication Theory*, 17, pp. 348–355. https://doi.org/10.1111/j.1468-2885.2007.00301.x

Bolsen, T., Druckman, J. N. (2018). Do partisanship and politicization undermine the impact of a scientific consensus message about climate change? *Group Processes & Intergroup Relations*, *21*(3), pp. 389-402. https://doi.org/10.1177/1368430217737855

Bontridder, N., Poullet, Y. (2021) The role of artificial intelligence in disinformation. *Data & Policy*, *3*, https://doi.org/10.1017/dap.2021.20

Borgoni, C. (2015) Epistemic Akrasia and Mental Agency. *Rev.Phil.Psych.* **6**, pp. 827–842 https://doi.org/10.1007/s13164-014-0205-4

Boutyline, A., Willer, R. (2016) The Social Structure of Political Echo Chambers: Variation in Ideological Homophily in Online Networks, *Political Psychology*, 38(3), 551-569. https://doi.org/10.1111/pops.12337

Bradley, R. (2017) *Decision Theory with a Human Face*, Cambridge: Cambridge University Press.

Brady, W.J., Crockett, M.J., Van Bavel, J.J. (2020) The MAD Model of Moral Contagion: The Role of Motivation, Attention, and Design in the Spread of Moralized Content Online, *Perspective on Psychological Science*, 15(4), pp. 978-1010. https://doi.org/10.1177/1745691620917336

Brandt, A. M. (1978) Racism and Research: The Case of the Tuskegee Syphilis Study. *The Hastings Center Report*, *8*(6), pp. 21–29. https://doi.org/10.2307/3561468

Breakstone, J., Smith, M., Connors, P., Ortega, T., Kerr, D., & Wineburg, S. (2021a) Lateral reading: College students learn to critically evaluate Internet sources in an online course, *Harvard Kennedy School Misinformation Review*, 2(1), pp. 1–17. https://doi.org/10.37016/mr-2020-56

Breakstone, J., Smith, M., Wineburg, S., Rapaport, A., Carle, J., Garland, M., & Saavedra, A. (2021b) Students' Civic Online Reasoning: A national portrait, *Educational Researcher*, 50(8), pp. 505-515, https://doi.org/10.3102/0013189X211017495

Brennen, J. S., Simon, F. M., Howard, P. N., Nielsen, R. K. (2020) Types, sources, and claims of COVID-19 misinformation (RISJ Factsheets). Reuters Institute for the Study of Journalism.

Brulle, R.J. (2014) Institutionalizing delay: foundation funding and the creation of U.S. climate change counter-movement organizations, *Climatic Change*, 122, pp. 681-694, https://doi.org/10.1007/s10584-013-1018-7

Bursztyn, L., Rao, A., Roth, C., Yanagizawa-Drott, D. (2020) Misinformation During a Pandemic, Becker Friedman Institute for Economics, Working Paper No. 2020-44.

Bush, V. (1945) *Science: The Endless Frontier*, Princeton, NJ: Princeton University Press.

Cardenal, A. S., Aguilar-Paredes, C., Cristancho, C., Majó-Vázquez, S. (2019) Echo-Chambers in Online News Consumption: Evidence from Survey and Navigation Data in Spain, *European Journal of Communication*, 34, pp. 360–376, https://doi.org/10.1177/0267323119844409

Carrasco-Farré, C. (2022) The fingerprints of misinformation: how deceptive content differs from reliable sources in terms of cognitive effort and appeal to emotions, *Humanit Soc Sci Commun,* 9(162), https://doi.org/10.1057/s41599-022-01174-9

Chatterjee, E. (2022) Numbers without experts, In: Newfield, C., Alexandrova, A., John, S. (eds.) *Limits of the numerical: The abuses and uses of quantification*, Chicago: Chicago University Press, pp. 23–45.

Chigwedere, P., Seage, G.R. 3rd, Gruskin, S., Lee, T.H., Essex, M. (2008) Estimating the lost benefits of antiretroviral drug use in South Africa, *J Acquir Immune Defic Syndr*, 49(4), pp. 410-415. https://doi.org/10.1097/QAI.0B013E31818A6CD5

Christiano, T. (2022) Algorithms, Manipulation, and Democracy, Canadian Journal of Phiosophy, 52(1), pp. 109-124. https://doi.org/10.1017/can.2021.29

Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., Starnini, M. (2021) The echo chamber effect on social media, *Proc. Natl. Acad. Sci. U.S.A.*, 118 (9), https://doi.org/10.1073/pnas.2023301118

Clark, A., Fleche, S., Layard, R., Powdthavee, N., Ward, G. (2018) *The Origins of Happiness: The Science of Well-Being over the Life Course*, Princeton, NJ: Princeton University Press.

Cobalti, A., Schizzerotto, A. (1994) *La Mobilità Sociale in Italia*, Bologna: Il Mulino.
consequences, Time, Available at: https://time.com/4590255/pizzagate-fake-news-what-to-know/

Cook, J., Oreskes, N., Doran, P. T., Anderegg, W. R. L., Verheggen, B., Maibach, E. W., Carlton, J. S., Lewandowsky, S., Skuce, A. G., Green, S. A., Nuccitelli, D., Jacobs, P., Richardson, M., Winkler, B., Painting, R., Rice, K. (2016) Consensus on consensus: A synthesis of consensus estimates on human-caused global warming, *Environmental Research Letters*, *11*(4), Article 048002. https://doi.org/10.1088/1748-9326/11/4/048002
Coser, L. A. (1956). *The functions of social conflict.* Free Press.

Cota, W., Ferreira, S.C., Pastor-Satorras, R. *et al.* (2019) Quantifying echo chamber effects in information spreading over political communication networks. *EPJ Data Sci.,* 8(35). https://doi.org/10.1140/epjds/s13688-019-0213-9

Crockett, M.J. (2017) Moral outrage in the digital age. *Nat Hum Behav* **1**, pp. 769–771, https://doi.org/10.1038/s41562-017-0213-3

Crouch, C. (2004) *Post-Democracy*, Polity Press.

Daston, L. (1992) Objectivity and the Escape from Perspective, *Social Studies of Science*, 22(4), 597-618, https://doi.org/10.1177/030631292022004002

Daston, L., Galison, P.L. (2007) *Objectivity*, Princeton, NJ: Princeton University Press.

De Leonardis, O. (2022) Quantifying Inequality: From Contentious Politics to the Dream of an Indifferent Power, In: Mennicken, A., Salais, R. (eds.), *The New Politics of Numbers: Utopia, Evidence and Democracy*, Executive Politics and Governance, Palgrave MacMillan. pp. 135-166.

Del Vicario, M., Scala, A., Caldarelli, G., Stanley, H.E., Quattrociocchi, W. (2017) Modeling confirmation bias and polarisation. *Scientific Reports*, 7, 40391, https://doi.org/10.1038/srep40391

Desmond, H. (2022) Status Distrust of Scientific Experts, *Social Epistemology*, *36*(5), pp. 586–600. https://doi.org/10.1080/02691728.2022.2104758

Desrosières, A. (2015) Retroaction: How indicators feed back onto quantified actors. In: Rottenburg, R., Merry, S., Park, S., Mugler, J. (eds.) *The world of indicators: The making of governmental knowledge through quantification.* Cambridge: Cambridge University Press, pp. 329–353. https://doi.org/10.1017/CBO9781316091265.013

Di Cristofaro, C. (2020). Adolescenti sui social più di 4 ore al giorno. Non riescono a ridurre tempo online. IlSole24Ore. Available at: https://www.ilsole24ore.com/art/adolescenti-social-piu-4-ore-giorno-non-riescono-ridurre-tempo-online-AD9Y7w5

Diaz Ruiz, C. (2023). Disinformation on digital media platforms: A market-shaping approach. *New Media & Society*. https://doi.org/10.1177/14614448231207644

DiPaolo, J. (2018). Higher-Order Defeat is Object-Independent, *Pacific Philosophical Quarterly*, 99(2), pp. 248-269, https://doi.org/10.1111/papq.12155

Douglas, H. (2000), Inductive Risk and Values in Science*, Philosophy of Science*, 67(4), pp. 559-579, https://doi.org/10.1086/392855

Dretske, F. (1983) Precis of Knowledge and the Flow of Information, *Behavioral and Brain Sciences*, 6(1), pp. 55-63, https://psycnet.apa.org/doi/10.1017/S0140525X00014631

Dretske, F. (2008) Epistemology and Information, In: Pieter, A., Van Benthem, J. (eds.) *Handbook of the Philosophy of Science*, Volume 8. Philosophy of Information, Elsevier-North Holland, pp. 24-47.

Dulith Novaes, C. (2023) The (higher-order) evidential significance of attention and trust—comments on Levy's *Bad Beliefs, Philosophical Psychology*, 36(4), pp. 792-807, https://doi.org/10.1080/09515089.2023.2174845

Dunlap, R.E., Brulle, R. J. (2020) Sources and Amplifiers of Climate Change Denial, In: Holmes, D. C., Richardson, L. M. (eds.) *Research Handbook on Communicating Climate Change*, Cheltenham: Edward Elgar, pp. 49-61.

Dunlap, R.E., McCright A., (2010) Climate change denial, sources, actors, and strategies, In: Lever-Tracy, C. (ed.) *Routledge Handbook of Climate Change and Society*, Abingdon, UK: Routledge.

Dunlap, R.E., McCright A., (2015) Challenging Climate Change: the denial countermovement, In: Dunlap, R.E., Brulle R.J. (eds.), Climate Change and Society: Sociological Perspectives, New York: Oxford University Press, pp. 300-332.

Durante, M. (2017) Ethics, Law and the Politics of Information: A Guide to the Philosophy of Luciano Floridi, *The International Library of Ethics, Law and Technology,* Volume 18, Springer.

Ecker, U.K.H., Lewandowsky, S., Cook, J. et al. (2022) The psychological drivers of misinformation belief and its resistance to correction, *Nat Rev Psychol,* **1**, pp. 13–29, https://doi.org/10.1038/s44159-021-00006-y

EDMO (2022) Disinformation about climate change – Main narratives in June at the European level. https://edmo.eu/publications/disinformation-about-climate-change-main-narratives-in-june-at-the-european-level/

Egan, P.J. (2020) Identity as Dependent Variable: How Americans Shift Their Identities to Align with Their Politics, *American Journal of Political Science*, 64(3), pp. 699-716, https://doi.org/10.1111/ajps.12496

Ellemers, N., Spears, R., Doosje, B. (1999). *Social identity: Context, commitment, content.* Blackwell Science.

Fallis, D. (2014) A Functional Analysis of Disinformation, In: *iConference 2014 Proceedings,* pp. 621–627, https://doi.org/10.9776/14278

Fallis, D. (2016) What is Disinformation? *Library Trends*, 63, pp. 401-426, https://doi.org/10.1353/LIB.2015.0014

Fazio, L.K., Brashier, N.M., Payne, B.K., Marsh, E.J. (2015) Knowledge does not protect against illusory truth, *J Exp Psychol Gen*, 144(5), pp. 993-1002. https://doi.org/10.1037/xge0000098

Featherstone, J. D., & Zhang, J. (2020). Feeling angry: the effects of vaccine misinformation and refutational messages on negative emotions and vaccination attitude. *Journal of Health Communication*, *25*(9), pp. 692–702, https://doi.org/10.1080/10810730.2020.1838671

Fehr, E., U. Fischbacher (2003) The Nature of Human Altruism, *Nature*, 425, pp. 785–791, https://doi.org/10.1038/nature02043

Fetzer, J.H. (2004) Disinformation the use of false information, *Minds and Machines,* 14, pp. 223- 229, https://doi.org/10.1023/B:MIND.0000021683.28604.5b

Fisher, K.A., Nguyen, N., Fouayzi, H., Singh, S., Crawford, S., Mazor, K.M. (2023) Impact of a physician recommendation on COVID 19 vaccination intention among vaccine hesitant individuals, *Patient Educ Couns*, 106, pp. 107-112, https://doi.org/10.1016/j.pec.2022.09.013

Flaxman S, Sharad Goel, Justin M. Rao (2016) Filter Bubbles, Echo Chambers, and Online News Consumption, *Public Opinion Quarterly*, Volume 80, Issue S1, 2016, Pages pp. 298–320, https://doi.org/10.1093/poq/nfw006

Floridi, L. (2007) In defence of the veridical nature of semantic information. *European Journal of Analytic Philosophy (EUJAP), 3*(1), pp. 31–41.

Floridi, L. (2011). *The philosophy of information*. Oxford: Oxford University Press.

Floridi, L. (2013) *Ethics of Information*, Oxford Academic.

Floridi, L. (2014) *The Fourth Revolution: How the Infosphere Is Reshaping Human Reality*. Oxford: Oxford University Press.

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

Fukuyama, F., Grotto, A. (2020). Comparative Media Regulation in the United States and Europe. In: Persily, N., Tucker, J. A. (eds.), *Social Media and Democracy,* pp. 199–219, chapter, Cambridge: Cambridge University Press.

Fung, B. (2024) UK riots show how social media can fuel real-life harm. It's only getting worse, CNN. Available at: https://edition.cnn.com/2024/08/09/tech/uk-protests-social-media/index.html

Galeotti, E. (2020) Political Disinformation and Voting Behavior: Fake News and Motivated Reasoning, *Notizie di Politeia,* 37(142), pp. 64-85.

Garrett, R.K. (2017) The Echo Chamber Distraction: Disinformation Campaigns are the Problem, Not Audience Fragmentation, *Journal of Applied Research in Memory and Cognition*, 6, pp. 370-376, https://psycnet.apa.org/doi/10.1016/j.jarmac.2017.09.011

Gasser, U., Cortesi, S., Malik, M., Lee, A. (2012) Youth and digital media: From credibility to information quality, Berkman Klein Center for Internet & Society. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2005272

Goertzel, T. (1994) Belief in conspiracy theories. *Political Psychology, 15*(4), pp. 731–742. https://doi.org/10.2307/3791630

Gottfried, J., Shearer, E. (2016) News use across social media platforms, Pew Research Centre, Available at: https://www.pewresearch.org/journalism/2016/05/26/news-use-across-social-media-platforms-2016/

Granaglia, E. (2022) *Uguaglianza di Opportunità. Sì, ma quale?*, Bari: Laterza.

Graziosi, A. (2023) *Occidenti e Modernità. Vedere un mondo nuovo*, Bologna: Il Mulino.

Grebe, E., Nattrass, N. (2012) AIDS conspiracy beliefs and unsafe sex in Cape Town, PubMed, 16(3), pp. 761-773, https://doi.org/10.1007/s10461-011-9958-2

Greenspan, R.L., Loftus, E.F. (2020) Pandemics and Infodemics: Research on the effects of misinformation on memory, *Hum. Behav. & Emerg. Tech.*, 3, pp. 8-12, https://doi.org/10.1002/hbe2.228

Guess, A.M., Lyons B.A. (2020) Misinformation, Disinformation, and Online Propaganda. In: Persily, N., Tucker, J.A. (eds.) *Social Media and Democracy*. SSRC Anxieties of Democracy. Cambridge University Press, pp. 10-33.

Guess, A.M., Lyons, B.A., Nyhan, B., Reifler, J. (2018) Avoiding the echo chamber about echo chambers: Why selective exposure to like-minded political news is less prevalent than you think, *Knight Foundation Report*.

Guess, A.M., Nyhan, B., Reifler, J. (2020) Exposure of untrustworthy websites in the 2016 US election, *Nature Human Behaviour*, 4, pp. 472–480. https://doi.org/10.1038/s41562-020-0833-x

Gunn, H. (2021) Filter bubbles, Echo Chambers, and Online Communities. In: Hannon, M., de Ridder, J. (eds.), *The Routledge Handbook of Political Epistemology*. New York: Routledge. pp. 192-202.

Hacking, I. (1990) *Taming of Chance*, Cambridge: Cambridge University Press

Hallin, D.C., Mancini, P. (2004) *Comparing Media Systems: Three Models of Media and Politics*, Cambridge: Cambridge University Press.

Hameleers, M. (2023) Disinformation as a context-bound phenomenon: toward a conceptual clarification integrating actors, intentions and techniques of creation and dissemination, *Communication Theory*, pp. 1-10, https://psycnet.apa.org/doi/10.1093/ct/qtac021

Hameleers, M., van der Meer, T. G. (2020). Misinformation and polarisation in a high-choice media environment: How effective are political fact-checkers*? Communication Research*, *47*(2), pp. 227–250. https://doi.org/10.1177/0093650218819671

Han, J., Cha, M., & Lee, W. (2020). Anger contributes to the spread of COVID-19 misinformation, Misinformation Review: Harvard Kennedy School.

Hardin, R. (1993) The street level epistemology of trust, *Politics & Society*, 21(4), pp. 505-529, https://doi.org/10.1177/0032329293021004006

Hardin, R. (2002) *Trust and Trustworthiness*, Russell Sage Foundation.

Hargittai, E., Fullerton, L., Menchen-Trevino, E., Thomas, K. Y. (2010). Trust online: Young adults' evaluation of Web content, *International Journal of Communication*, 4, pp. 468–494.

Harvey, J.A., van der Berg, D., Ellers, J., Kampen, R., *et al.* (2018) Internet Blogs, Polar Bears, and Climate Change Denial by Proxy, *BioScience*, 68(4), pp. 281-287. https://doi.org/10.1093/biosci/bix133

Haushofer, J., Fehr, E. (2014), On the psychology of poverty, *Science*, 344(6186), pp. 862–867, https://doi.org/10.1126/science.1232491

Hayek, F.A. (1945) The Use of Knowledge in Society, *The American Economic Review*, 35(4), pp. 519-530.

Hempel, C.G. (1952) *Fundamentals of Concept Formation in Empirical Science*, University of Chicago Press.

Hempel, C.G. (1965) Science and Human Values, Aspects of Scientific Explanation and other Essays in the Philosophy of Science, New York: The Free Press, pp. 81-96.

Hess, D. J. (2014) When Green Became Blue: Epistemic Rift and the Corralling of Climate Science, Political Power and Social Theory, 27, pp. 123-153, https://doi.org/10.1108/S0198-871920140000027012

Hmielowski, J.D., Matthews, A., Chu, H. (2023) Going the distance for COVID-19: Relationships Among News Use, Psychological Distance, Risk Perceptions, and Behavioural Intentions, Science Communication, 45(4), pp. 460-484, https://doi.org/10.1177/10755470231188694

Hogg, M.A. (2014) From Uncertainty to Extremism: Social Categorization and Identity Processes, *Current Directions in Psychological Science*, 23(5), pp. 338-342, https://doi.org/10.1177/0963721414540168

Hong, L., Page, S.E. (2004) Groups of diverse problem solvers can outperform groups of high-ability problem solvers, *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 101(46), pp. 16385–16389, https://doi.org/10.1073/pnas.0403723101

Hornsey, M.J., Fielding, K.S. (2017) Attitude roots and Jiu Jitsu persuasion: Understanding and overcoming the motivated rejection of science, *American Psychologist*, 72 (5), pp. 459-473, https://doi.org/10.1037/a0040437

Hornsey, M.J., Harris, E.A., Bain, P. *et al.* (2016) Meta-analyses of the determinants and outcomes of belief in climate change, *Nature Clim Change*, **6**, pp. 622–626, https://doi.org/10.1038/nclimate2943

Hornsey, M.J., Harris, E.A., Fielding, K.S. (2018) The psychological roots of anti-vaccination attitudes: A 24-nation investigation. *Health Psychol*, 37(4), pp. 307-315. https://doi.org/10.1037/hea0000586

Horowitz, S. (2013) Epistemic Akrasìa, *Nous*, 48(4), pp. 718-744, https://doi.org/10.1111/nous.12026

Huddy, L. (2002) From Social to Political Identity: A Critical Examination of Social Identity Theory, *Political Psychology*, 22(1), pp. 127-156, https://doi.org/10.1111/0162-895X.00230

Hughes, B., Miller-Idriss, C., Piltch-Loeb, R., Goldberg, B., White, K., Criezis, M., Savoia, E. (2021) Development of a Codebook of Online Anti-Vaccination Rhetoric to Manage COVID-19 Vaccine Misinformation, *International Journal of Enviornmental Research and Public Health*, 18, 7556.

Humprecht, E., Esser, F., Van Aelst, P. (2020) Resilience to Online Disinformation: A Framework for Cross-National Comparative Research, *The International Journal of Press/Politics*, *25*(3), pp. 493-516, https://doi.org/10.1177/1940161219900126

Istat (2019) Conciliazione tra lavoro e famiglia, available at: https://www.istat.it/it/files//2019/11/Report-Conciliazione-lavoro-e-famiglia.pdf

Jachimowicz, J. M., Chafik, S. Munrat, S., Prabhu, J. C., Weber, E. U. (2017) Community trust reduces myopic decisions of low-income individuals, *Proceedings of the National Academy of Sciences*, 114(21), pp. 5401-5406. https://doi.org/10.1073/pnas.1617395114

Jacobo, J. (2024) The presidential debate proved that climate change is not at center of 2024 election, experts say, *ABC News*. Available at: https://abcnews.go.com/US/presidential-debate-proved-climate-change-center-2024-election/story?id=113585193

Jaiswal, J., Halkitis, P.N. (2019) Towards a More Inclusive and Dynamic Understanding of Medical Mistrust Informed by Science, *Behav Med*., 45(2), pp. 79-85, https://doi.org/10.1080/08964289.2019.1619511

John, S. (2018) Epistemic trust and the ethics of science communication: against transparency, openness, sincerity and honesty, *Social Epistemology*, 32(2), pp. 75-87, https://doi.org/10.1080/02691728.2017.1410864

John, S. (2021) *Objectivity in Science*, Cambridge: Cambridge University Press.

Jost, J.T., Van der Linden, S., Panagopoulos, C., Hardin, C.D. (2018) Ideological asymmetries in conformity, desire for shared reality, and the spread of misinformation. *Curr Opin Psychol.*, 23, pp. 77-83, https://doi.org/10.1016/j.copsyc.2018.01.003

Kahan, D.M. (2015) Climate-Science Communication and the Measurement Problem, *Advances in Political Psychology*, 36(1), https://doi.org/10.1111/pops.12244

Kahan, D.M., Peters, E., Wittlin, M. et al. (2012) The polarizing impact of science literacy and numeracy on perceived climate change risk, *Nature Clim Change*, 2, pp. 732-735, https://doi.org/10.1038/NCLIMATE1547

Kahneman, D. (2011) *Thinking Fast and Slow*, FSG.

Kahneman, D., Tversky, A. (1979) Prospect Theory: An analysis of decision under risk, *Econometrica*, 47(2), pp. 263-291, https://psycnet.apa.org/doi/10.1017/CBO9780511609220.014

Kennedy, B., Tyson, A. (2024) How Republicans view climate change and energy issues, Pew Research Centre, Available at: https://www.pewresearch.org/short-reads/2024/03/01/how-republicans-view-climate-change-and-energy-issues/

Kennedy, B.R., Mathis, C.C., Woods, A.K. (2007) African Americans and their distrust of the health care system: healthcare for diverse populations, *J Cult Divers*, 14(2), pp. 56-60. https://doi.org/10.35455/BRK1232021

Kitcher, P. (2011) *Science in a Democratic Society*, Prometheus Books.

Koskinen, I. (2018) Defending a risk account of scientific objectivity, *The British Journal for the Philosophy of Science*, 71(4), pp. 1187-1207.

Kreiss, D. (2018) The Media Are about Identity, Not Information, In: *Trump and the Media*, Boczkowski, P,J., Papacharissi, Z. (eds.) Cambridge, MA: MIT Press, pp. 93–99.

Kukla, R., (2017), *Infertility, epistemic risk, and disease definitions,* Synthese 196**,** pp. 4409– 4428, https://doi.org/10.1007/s11229-017-1405-0

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*(3), pp. 480–498. https://doi.org/10.1037/0033-2909.108.3.480

Lackey, J. (2021) Echo Chambers, Fake News, and Social Epistemology, In: Bernecker, S., Flowerree, A.K., Grundmann, T. (eds.), *The Epistemology of Fake News*, Oxford Academic.

Landemore, H. (2020) *Open Democracy: Reinventing Popular Rule in the Twenty-First Century*, Princeton University Press.

Langenkamp, A. (2021). Lonely hearts, empty booths? The relationship between loneliness, reported voting behavior and voting as civic duty, *Social Science Quarterly,* 102(4), pp. 1239–1254. https://doi.org/10.1111/ssqu.12946

Lasonen-Aarnio, M. (2013) Higher-Order Evidence and the Limits of Defeat, Philosophy and Phenomenological Research, 88(2), pp. 314-345, https://doi.org/10.1111/phpr.12090

Lau, A.Y.S., Coiera, E.W. (2007), Do People Experience Cognitive Biases while Searching for Information? *Journal of the American Medical Informatics Association*, 14 (5), pp. 599-608, https://doi.org/10.1197/jamia.m2411

Leiserowitz, A. (2007) Communicating the Risks of Global Warming: American Risk Perceptions, Affective Images, and Interpretive Communities, In: Susanne, C., Dilling, M., Dilling, L., *Creating a Climate for Change: Communicating Climate Change and Facilitating Social Change*, eds., New York: Cambridge University Press, pp. 44–63.

Leiserowitz, A., Edward Maibach, E., Roser-Renouf, C., Geof Feinberg, G., Seth Rosenthal, S. (2014) *Climate Change in the American Mind*, New Haven, CI: Yale Project on Climate Change Communication.

Levi, I. (1960), Must the Scientist Make Value Judgments?, *The Journal of Philosophy*, 57(11), pp. 345-357.

Levine, S., Apfelbaum, E., Bernard, M., Bartelt, V., Zajac, E., Stark, D. (2014) Ethnic diversity deflates price bubbles, *Proceedings of the National Academy of Sciences of the United States of America* (PNAS) 111(52), pp. 18524–18529. https://doi.org/10.1073/pnas.1407301111

Levy, N. (2019) Due Deference to Denialism: Explaining Ordinary People's Rejection of Established Scientific Findings, *Synthese*, 196(1), pp. 313–27, https://doi.org/10.1007/s11229-017-1477-x

Levy, N. (2022) *Bad Beliefs: Why They Happen to Good People*, Oxford: Oxford Academic, https://doi.org/10.1093/oso/9780192895325.001.0001

Lewandowsky, S, Oberauer K, Gignac GE. (2013) NASA faked the moon landing-therefore, (climate) science is a hoax: an anatomy of the motivated rejection of science, *Psychol Sci.*, 24(5), pp. 622-33, https://doi.org/10.1177/0956797612457686

Lewandowsky, S, Oberauer K. (2021) Worldview-motivated rejection of science and the norms of science. Cognition, 215, 104820. https://doi.org/10.1016/j.cognition.2021.104820

Lewandowsky, S., Cook, J. (2016) Rational Irrationality: Modeling Climate Change Belief Polarisation Using Bayesian Networks, *Topics in Cognitive Science,* (8), pp. 160–179, https://doi.org/10.1111/tops.12186

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012) Misinformation and Its Correction: Continued Influence and Successful Debiasing. *Psychological Science in the Public Interest*, *13*(3), pp. 106-131, https://doi.org/10.1177/1529100612451018

Lewandowsky, S., Oberauer, K. (2016). Motivated Rejection of Science. *Current Directions in Psychological Science*, *25*(4), pp. 217-222, https://doi.org/10.1177/0963721416654436

Lewis, R. (2018) Alternative Influence: broadcasting the reactionary right on YouTube, *Data & Society*. Report available at: https://datasociety.net/library/alternative-influence/

Lewis, R. (2020) "This Is What the News Won't Show You": YouTube Creators and the Reactionary Politics of Micro-celebrity, *Television & New Media*, 21(2), pp. 201-217, https://doi.org/10.1177/1527476419879919

List, A., Grossnickle, E. M., & Alexander, P. A. (2016). Undergraduate students' justifications for source selection in a digital academic context, *Journal of Educational Computing Research*, 54(1), pp. 22–61. https://doi.org/10.1177/0735633115606659

Loftus, E.F., Klemfuss, J.Z. (2023) Misinformation – past, present, future, *Psychology Crime Law*, 30(4), pp. 312-318, https://doi.org/10.1080/1068316X.2023.2219813

Longino, H.E. (1990) *Science as Social Knowledge: Values and Objectivity in Scientific Inquiries*, Princeton NJ: Princeton University Press.

Lübke, C. (2022) Socioeconomic Roots of Climate Change Denial and Uncertainty among the European Population, *European Sociological Review*, 38(1), pp. 153–168, https://doi.org/10.1093/esr/jcab035.

Luhmann, N. (1980) *Trust and Power,* New York: Wiley.

Mahtani, A. (2020) Awareness growth and dispositional attitudes, *Synthese*, 198(9), pp. 8981-8997, https://doi.org/10.1007/s11229-020-02611-5

Mani, A., Mullainathan, S., Shafir, E., Zhao, J., (2013), Poverty impedes cognitive function, *Science*, 341(6149), pp. 976–980, https://doi.org/10.1126/science.1238041

Marchlewska, M., Cichocka, A., Kossowska, M. (2017) Addicted to answers: Need for cognitive closure and the endorsement of conspiracy beliefs, *European Journal of Social Psychology*, 48(2), pp. 109-117, https://psycnet.apa.org/doi/10.1002/ejsp.2308

Marsili, N. (2020) Retweeting: its linguistic and epistemic value. *Synthese* 198, pp. 10457–10483, https://doi.org/10.1007/s11229-020-02731-y

Martini, F. (2015) *Conoscenza Digitale: l'attendibilità delle informazioni in rete*, Roma: Carocci Editore.

Martzoukou, K., Fulton, C., Kostagiolas, P., & Lavranos, C. (2020) A study of higher education students' self-perceived digital competences for learning and everyday life online participation. Journal of Documentation, 76(6), pp. 1413-1458. https://doi.org/10.1108/JD- 03-2020-0041

Mattavelli, S., Bianchi, C., Brambilla, M., Motterlini, M. (2024) True and moral by repetition: unveiling the impact of exposure on positive stereotypes perception. *International Review of Social Psychology*, 37(1), pp. 1-10, https://doi.org/10.5334/irsp.933

McCright A., Dunlap, R.E. (2011) Cool dudes: the denial of climate change among conservative white males in the United States, *Global Environmental Change*, 21(4), pp. 1163-1172. https://doi.org/10.1016/j.gloenvcha.2011.06.003

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(2), pp. 165–193. https://doi.org/10.1080/00933104.2017.1416320

McLoughlin et al. (2024) Misinformation exploits outrage to spread online, *Science,* 386, pp. 991-996, https://doi.org/10.1126/science.adl2829

Mede, N.G., Schaffer MS (2020) Science-related populism: Conceptualizing populist demands toward science. Public Understanding of Science, 29(5), pp. 473-491, https://doi.org/10.1177/0963662520924259

Medina, J. (2023) *The Epistemology of Protest: Silencing, Epistemic Activism, and the Communicative Life of Resistance*, Oxford: Oxford University Press.

Mennicken, A., Nelson Espeland, W. (2019) What's new with numbers? Sociological approaches to the study of quantification, *Annual Review of Sociology*, 45, pp. 223–245, https://doi.org/10.1146/annurev-soc-073117-041343

Mercier, H. (2020) *Not Born Yesterday: The Science of Who We Trust and What We Believe*, Princeton NJ: Princeton University Press.

Merry, S.E. (2016) *The seductions of quantification: Measuring human rights, gender violence, and sex trafficking*, Chicago: The University of Chicago Press.

Merton, R. (1949) *Social Theory and Social Structure*, The Free Press.

Miller, P. (1992) Accounting and objectivity: The invention of calculating selves and calculable spaces, *Annals of Scholarship*, 9(1/2), pp. 61–86.

Miller, P. (2001) Governing by numbers: Why calculative practices matter, *Social Research*, 68(2), pp. 379–396, https://doi.org/10.1002/9780470774274.CH10

Mitra, T., Counts, S., Pennebaker, J.W. (2016) Understanding Anti-Vaccination Attitudes in Social Media, *Proceedings of the International AAAI Conference on Web and Social Media*, 10(1), pp. 269-278, https://doi.org/10.1609/icwsm.v10i1.14729

Mozur, P. (2018) A Genocide Incited on Facebook, with Posts from Myanmar's Military, *The New York Times*. Available at: https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html

Muldoon, R. (2018) The paradox of diversity, *The Georgetown Journal of Law and Public Policy*, 16, pp. 805–818.

Müller, J.F. (2023) An Epistemic Account of Populism, *Episteme*, pp. 1-22, https://doi.org/10.1017/epi.2023.39

Munroe, W. (2023). Misleading Higher-Order Evidence and Rationality: We Can't Always Rationally Believe What We Have Evidence to Believe. *Episteme*, pp. 1–27, https://doi.org/10.1017/epi.2023.45

Na, L., Banks, S., Wang, P.P. (2023) Racial and ethnic disparities in COVID-19 vaccine uptake: A mediation framework, Vaccine, 41(14), pp. 2404-2411, https://doi.org/10.1016/j.vaccine.2023.02.079

Nelson Espeland, W., Stevens, M.L. (1998) Commensuration as a social process. *Annual Review of Sociology,* 24, pp. 313–343. https://doi.org/10.1146/ANNUREV.SOC.24.1.313

Nelson Espeland, W., Stevens, M.L. (2008) A sociology of quantification, *European Journal of Sociology,* 49(3), pp. 401–436. https://doi.org/10.1017/S0003975609000150

Nelson, J.L., Webster, J.G. (2017) The Myth of Partisan Selective Exposure: A Portrait of the Online Political Audience, *Social Media + Society*, 3(3), https://doi.org/10.1177/2056305117729314

Newfield, C. (2022) The role of the numerical in the decline of expertise. In: Newfield, C., Alexandrova, A., John, S. (eds.) *Limits of the numerical: The abuses and uses of quantification*, pp. 47-66. Chicago: Chicago University Press.

Nguyen, C. Thi (2020) Echo Chambers and Epistemic Bubbles, *Episteme,* 17, pp. 141–61, https://doi.org/10.1017/epi.2018.32

Nguyen, C. Thi (2021) Transparency is Surveillance, *Philosophical and Phenomenological Research,* 105(2), pp. 331-361, https://doi.org/10.1111/phpr.12823

Nguyen, C. Thi (2022) Trust as an unquestioning attitude, In: Gendler, T.S., Hawthorne, J., Chung, J. (eds.) *Oxford Studies in Epistemology Volume 7*, Oxford Studies in Epistemology 7, pp. 214-244, https://doi.org/10.1093/oso/9780192868978.003.0007

Nickerson, R. (1998) Confirmation bias: A ubiquitous phenomenon in many guises, *Review of General Psychology,* 2(2), pp. 175–220. https://psycnet.apa.org/doi/10.1037/1089-2680.2.2.175

Norgaard, K.M. (2011) *Living in Denial: Climate Change, Emotions, and Everyday Life,* Cambridge (Massachusetts): MIT Press.

Nyhan, B. (2021) Why the backfire effect does not explain the durability of political misperceptions, *Proc Natl Acad Sci*, 118(15), https://doi.org/10.1073/pnas.1912440117

Nyhan, B., Reifler, J., Richey, S., Freed, G.L. (2014) Effective messages in vaccine promotion: a randomized trial, *Pediatrics*, 133(4), https://doi.org/10.1542/peds.2013-2365

Nyhan, B., Settle, J., Thorson, E. et al. (2023) Like-minded sources on Facebook are prevalent but not polarizing, *Nature*, 620, pp. 137-144, https://doi.org/10.1038/s41586-023-06297-w

O'Neill, O. (2002), *A Question of Trust*, The Bbc Reith Lectures 2002, Cambridge: Cambridge University Press.

Odoardi, I., D'Ingiullo, D., Quaglione, D. (2022) Gender disparities between young and adult NEETs: do we need a more refined policy approach? *Applied Economics*, *55*(56), pp. 6685–6699. https://doi.org/10.1080/00036846.2022.2161991

OECD (2022) Building Trust to Reinforce Democracy: Main Findings from the 2021 OECD Survey on Drivers of Trust in Public Institutions, Building Trust in Public Institutions, OECD Publishing, Paris. Available at: https://www.oecd.org/en/publications/building-trust-to-reinforce-democracy_b407f99c-en.html

Oreskes, N., Conway, E.M. (2010) *Merchants of Doubt*, Bloomsbury.

Osmundsen, M., Bor, A., Vahlstrup, P. B., Bechmann, A., Petersen, M. B. (2021) Partisan Polarisation Is the Primary Psychological Motivation behind Political Fake News Sharing on Twitter, *American Political Science Review*, *115*(3), pp. 999–1015, https://doi.org/10.1017/S0003055421000290

Oysterman, D., Dawson, A. (2020) Your fake news our facts: identity-based motivation shapes what we believe share and accept, In: Greifeneder, R., Jaffe, M., Newman, E., Schwarz, N. (eds.) *The Psychology of Fake News: Accepting, Sharing, and Correcting Misinformation* (1st ed.). Routledge.

Panizza, F., Ronzani, P., Martini, C., Mattavelli, S., Morisseau, T., Motterlini, M. (2022) Lateral reading and monetary incentives to spot disinformation about science. *Scientific Reports,* **12**(5678), https://doi.org/10.1038/s41598-022-09168-y

Paul, E., Steptoe, A., Fancourt, D. (2021) Attitudes towards vaccines and intention to vaccinate against COVID-19: Implications for public health communications, *Lancet Reg Health Eur*, https://doi.org/10.1016/j.lanepe.2020.100012

Pennycook, G. (2025) *'Politics, not policy': Meta ending fact-checking program*, Cornell Chronicle, Cornell University Media Relation Office. Interview available at: https://news.cornell.edu/media-relations/tip-sheets/politics-not-policy-meta-ending-fact-check-program

Pennycook, G., Cannon, T.D., Rand, D.G. (2018) Prior exposure increases perceived accuracy of fake news, *Journal of Experimental Psychology General*, 147(12), pp. 1865-1880, https://doi.org/10.1037/xge0000465

Pennycook, G., Rand, D. (2019) Lazy not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than motivated reasoning, *Cognition*, 188, pp. 39-50, https://doi.org/10.1016/j.cognition.2018.06.011

Pennycook, G., Rand, D.G., (2021). The Psychology of Fake News, *Trends in Cognitive Sciences*, 25(5), pp. 388-402, https://doi.org/10.1016/j.tics.2021.02.007

Pereira, A., Harris, E., Van Bavel, J. J. (2021) Identity concerns drive belief: The impact of partisan identity on the belief and dissemination of true and false news, *Group Processes & Intergroup Relations*, 26(1), pp. 24-47, https://doi.org/10.1177/13684302211030004

Peters, U. (2022) What Is the Function of Confirmation Bias? *Erkenntnis*, 87, pp. 1351-1376. https://link.springer.com/article/10.1007/s10670-020-00252-1

Pew Research Center (2024) Public Trust in Government 1958-2024. Available at: https://www.pewresearch.org/politics/2024/06/24/public-trust-in-government-1958-2024/

Piazza, T., Croce, M. (2022) *Che cosa sono le fake news*, Roma: Carocci Editore.

Piketty, T. (2021) *A Brief History of Equality*, Harvard University Press.

Pluviano, S., Watt, C., Della Sala, S. (2017) Misinformation lingers in memory: failure of three pro-vaccination strategies, *PLoS ONE*, 12(7), https://doi.org/10.1371/journal.pone.0181640

Pooley, E. (2010) *The Climate War: True Believers, Power Brokers, and the Fight to Save the Earth*, New York: Hyperion.

Popper, K.R. (1972) *Conjectures and Refutations*, 4th edition, London and Henley, Routledge Kegan Paul.

Porter, T. (1993) Statistics and the politics of objectivity. *Revue de Synthèse* (114): pp. 87-101, https://doi.org/10.1007/BF03181156

Porter, T. (1995) *Trust in numbers: The pursuit of objectivity in science and public Life*, Princeton University Press.

Power, M. (1997) *The Audit Society: Rituals of Verification*, Oxford: Oxford University Press.

Preterossi, G. (2011) *La Politica Negata*, Bari: Laterza.

Pretus, C., Servin-Barthet, C., Harris, E.A., Brady, W.J., Vilarroya, O., Van Bavel, J.J. (2023) The Role of Political Devotion in Sharing Partisan Misinformation and Resistance to Fact-Checking, *Journal of Experimental Psychology: General*, 152(11), pp. 3116-3134, https://doi.org/10.1037/xge0001436

Quattrociocchi, W., Scala, A., Sunstein, C. R., (2016) *Echo Chambers on Facebook*. Available at: http://dx.doi.org/10.2139/ssrn.2795110

Quilter-Pinner, H., Statham, R., Jennings, W., Valgarõsson, V. (2021) Trust Issues: Dealing with Distrust in Politics, *Institute for Public Policy Research*.

Rathje, S., Roozenbeek, J., Van Bavel, J.J., van der Linden, S. (2023) Accuracy and social motivations shape judgements of (mis)information, *Nature Human Behaviour*, 7, pp. 892-903, https://doi.org/10.1038/s41562-023-01540-w

Rathje, S., Van Bavel, J.J., van der Linden, S. (2021) Out-group animosity drives engagement on social media, *Proc Nat Acad Sci USA*, 118(26). https://doi.org/10.1073/pnas.2024292118

Rideout, V., Robb, M. B. (2019) The Common Sense census: Media use by tweens and teens, San Francisco, CA: Common Sense Media.

Ronzani, P., Panizza, F., Martini, C., Savadori, L., Motterlini, M. (2022) Countering vaccine hesitancy through medical expert endorsement, Vaccine, 40(32), pp. 4635-4643, https://doi.org/10.1016/j.vaccine.2022.06.031

Roozenbeek, J., Schneider, C.R., Dryhurst, S., Kerr, J., Freeman, A.L.J., Recchia, G., van der Bles, A.M., van der Linden S. (2020) Susceptibility to misinformation about COVID-19 around the world. *R Soc Open Sci.*, 7(10), 201199, https://doi.org/10.1098/rsos.201199

Roozenbeek, J., van der Linden, S. (2024) *The Psychology of Misinformation*, Cambridge: Cambridge University Press.

Rose, N. (1991) Governing by numbers: Figuring out democracy, *Accounting, Organization and Society*, 16(7), pp. 673–692, https://doi.org/10.1016/0361-3682(91)90019-B

Rosentiel, T. (2009) Fewer Americans see solid evidence of climate change, Pew Research Center. Available at: https://www.pewresearch.org/science/2009/10/22/fewer-americans-see-solid-evidence-of-global-warming/

Rudner, R., (1953) The Scientist Qua Scientist Makes Value Judgments, *Philosophy of Science*, 20(1), pp. 1-6, https://psycnet.apa.org/doi/10.1086/287231

Samuelson, K., (2016) *What to know about Pizzagate, the fake new story with real*

Scarantino, A., Piccinini, G. (2010). Information without truth, *Metaphilosophy, 41*(3), pp. 313–330, https://doi.org/10.1111/j.1467-9973.2010.01632.x

Schaffner, B. F., Roche, C. (2017). Misinformation and motivated reasoning: Responses to economic news in a politicized environment. *Public Opinion Quarterly,* 81(1), pp. 86–110, https://doi.org/10.1093/poq/nfw043

Sen, A. (2000) Social Exclusion: Concept, Application and Scrutiny. Office of Environment and Social Development, Asian Development Bank, Social Development Papers.

Sennett, (2016) *L'Uomo Flessibile: Le conseguenze del nuovo capitalismo sulla vita personale*, Feltrinelli.

Shavit, A. (2021) Diverse populations are conflated with heterogeneous collectives, The Journal of Philosophy, 118(10), pp. 525–548, https://doi.org/10.5840/jphil20211181037

Shelby, A., Ernst, K. (2013) Story and science: How providers and parents can utilize storytelling to combat anti-vaccine misinformation, *Human Vaccines & Immunotherapeutics*, 9(8), pp. 1795-1801, https://doi.org/10.4161/hv.24828

Simchon, A., Edwards, M., Lewandowsky, S. (2024) The persuasive effects of political microtargeting in the age of generative artificial intelligence, PNAS Nexus, 3, pp. 1-5, https://doi.org/10.1093/pnasnexus/pgae035

Simion, M. (2023) Knowledge and Information, *Episteme*, pp. 1-12, https://doi.org/10.1017/epi.2023.25

Simmel, G. (1908) *Sociologia*, Milano: Meltemi.

Simon, H.A., (1955), A Behavioral Model of Rational Choice, *The Quarterly Journal of Economics*, 69(1), pp. 99-118.

Simonov, A., Sacher, S., Dubé, J., Biswas, S. (2020) The Persuasive Effect of Fox News: Non-Compliance with Social Distancing During the COVID-19 Pandemic, *Political Communication eJournal,* https://doi.org/10.2139/ssrn.3600088

Singer, D.J., Bramson, A., Grim, P., et al. (2019), Rational social and political polarisation, Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition, 176(9), pp. 2243-2269, https://doi.org/10.1007/S11098-018-1124-5

Sisco, M.R., Bosetti, V., Weber, E.U. (2017) When do extreme weather events generate attention to climate change? *Climatic Change*, 143, pp. 227-241, https://doi.org/10.1007/s10584-017-1984-2

Søe, S.O. (2021) A unified account of information, misinformation, and disinformation, *Synthese*, 198, pp. 5929-5949, https://doi.org/10.1007/s11229-019-02444-x

Srnicek, N. (2017) *Platform Capitalism*, Polity.

Standing, G. (2011) *The Precariat: The new dangerous class*, London: Bloomsbury Academic.

Starr, P. (2020) The Flooded Zone: How We Became More Vulnerable to Disinformation in the Digital Era, In: Persily, N., Tucker, J. A. (eds.), *Social Media and Democracy*, pp. 67–91, Cambridge: Cambridge University Press.

Steel, D., Fazelpour, S., Crewe, B., Gillette, K. (2021) Information elaboration and epistemic effects of diversity, *Synthese*, 198, pp. 1287–1307, https://doi.org/10.1007/s11229-019-02108-w

Stieglitz, S., Dang-Xuan, L. (2014) Emotions and Information Diffusion on Social Media-Sentiments of Microblogs and Sharing Behaviour, *Journal of Management Information Systems*, 29(4), pp. 217-248, https://doi.org/10.2753/MIS0742-1222290408

Sunstein, C.R. (2017) *#Republic*, Princeton NJ: Princeton University Press.

Sunstein, C.R., Vermeule, A., (2009) Conspiracy Theories: Causes and Cures, *The Journal of Political Philosophy,* 17(2), pp. 202–227, https://doi.org/10.1111/j.1467-9760.2008.00325.x

Swire-Thompson, B., Ecker, U.K.H., Lewandowsky, S., Berinsky, A.J. (2019) They Might Be a Liar But They're My Liar: Source Evaluation and the Prevalence of Misinformation, *Political Psychology*, 41(1), pp. 21-34. https://doi.org/10.1111/pops.12586

Swire, B., Ecker, U.K.H., Lewandowsky, S. (2017) The role of familiarity in correcting inaccurate information, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(12), pp. 1948-1961. https://psycnet.apa.org/doi/10.1037/xlm0000422

Tajfel, H. (1970) Experiments in Intergroup Discrimination, *Scientific American*, 223(5), pp. 96-102, https://doi.org/10.1038/scientificamerican1170-96

Tajfel, H., Turner JC (1986) The Social Identity Theory of Intergroup Behaviour, In: Worchel, S., Austin, W.G. (eds.), *Psychology of intergroup relations*, pp. 7-24, Chicago: Nelson Hall.

Tajfel, H., Turner, J. C. (2004) The Social Identity Theory of Intergroup Behavior, In: Jost, J.T., Sidanius, J. (Eds.), *Political psychology: Key readings,* pp. 276–293.

Tajfel, H., Turner, J.C. (1979) An Integrative Theory of Intergroup Conflict, In: Hatch, M.J., Schultz, M. (eds.), *Organizational Identity: A reader*, pp. 56-65, Oxford: Oxford University Press.

Thernstrom, A. (1987) Statistics and the politics of minority representation: The evolution of the Voting Rights Act since 1965. In: Alonso, W., Starr, P. (eds) *The Politics of Numbers*, pp. 303-327, New York: Russell Sage Foundation.

Treen, K.M., Williams, H.T.P., O'Neill, S.J. (2020) Online misinformation about climate change, *WIREs Climate Change*, 11(5), https://doi.org/10.1002/wcc.665

Tucker, J.A., Guess, A., Barbera, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D., Nyhan, B. (2018) Social Media, Political Polarisation, and Political Disinformation: A Review of the Scientific Literature. Available at: https://dx.doi.org/10.2139/ssrn.3144139

Tyler, T.R. (2010), *Why People Cooperate? The Role of Social Motivation*, Princeton University Press.

Tyler, T.R., Blader, S.L. (2000), *Cooperation in Groups: Procedural Justice, social identity and behavioural engagement*, Psychology Press, Philadelphia PA: Taylor & Francis.

Unzicker, K. (2023) Disinformation a challenge for democracy. Attitudes and Perceptions in Europe. DOI10.11586/2023044

Uscinki, J., Douglas, K., Lewandowsky, S. (2017) Climate Change Conspiracy Theories, *Oxford Research Encyclopedia of Climate Science, https://doi.org/10.1093/acrefore/9780190228620.013.328*

Van Bavel, J.J., Rathje, S., Harris, E., Robertson, C., Sternisko, A. (2021) How social media shape polarisation, *Trends Cogn Sci*., 25(11), pp. 913-916, https://doi.org/10.1016/j.tics.2021.07.013

Van der Linden S., Panagopoulos C., Roozenbeek J. (2020) You are Fake News: political bias in perceptions of fake news, *Media, Culture and Society*, 42(3), pp. 460-470, https://doi.org/10.1177/0163443720906992

Van der Linden, S., Leiserowitz, A., Rosenthal, S., Maibach, E. (2017) Inoculating the Public against Misinformation about Climate Change, *Global Challenges* 1(2), https://doi.org/10.1002/gch2.201600008

Van Prooijen, J.W., Douglas, K.M. (2017) Conspiracy theories as part of history: The role of societal crisis situations*, Memory Studies*, 10(3), pp. 323–333, https://doi.org/10.1177/1750698017701615

Vellani, V., Zheng, S., Ercelik, D., Sharot, T. (2023) The illusory truth effect leads to the spread of misinformation, *Cognition*, 236(105421), https://doi.org/10.1016/j.cognition.2023.105421

Vosoughi, S., Deb, R., Sinan, A. (2018) The spread of true and false news online, *Science,* 359(6380), pp. 1146–1151, https://doi.org/10.1126/science.aap9559

Walker, M. (2021) U.S. newsroom employment has fallen 26% since 2008, Pew Research Center. Available at: https://www.pewresearch.org/short-reads/2021/07/13/u-s-newsroom-employment-has-fallen-26-since-2008/

Walsh, D., Rashwan, N. (2019) "We're at a war": A covert social media campaign boosts military rulers, The New York Times, available at: https://www.nytimes.com/2019/09/06/world/middleeast/sudan-social-media.html

Wang, W., Brashier, N.M., Wing, E.A., Marsh, E.J., Cabeza, R. (2016) On Known Unknowns: Fluency and the Neural Mechanisms of Illusory Truth, Journal of Cognitive Neuroscience, 28(5), pp. 739-746, https://doi.org/10.1162/jocn_a_00923

Watson, W., Kumar, K., Michaelsen, L. (1993) Cultural diversity's impact on interaction process and performance: Comparing homogeneous and diverse task groups, *Academy of. Management Journal*, 36(3), pp. 590–602, https://psycnet.apa.org/doi/10.2307/256593

Weber, E. U., Johnson, E. J. (2015). Can we think of the future? Cognitive barriers to future-oriented decision making, In: *Global Cooperation and the Human Factor in International Relations,* pp. 139-154, Taylor and Francis, https://doi.org/10.4324/9781315691657-14

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), pp. 699-719, https://doi.org/10.1111/jcom.12164

Williams, J. (2018) *Stand Out of Our Light: Freedom and Resistance in the Attention Economy*, Cambridge University Press.

World Economic Forum (2020) The Global Social Mobility Report 2020: equality, opportunity and the new economic imperative. Available at: https://www3.weforum.org/docs/Global_Social_Mobility_Report.pdf

World Economic Forum (2024) The Global Risks Report 2024. Available at: https://www.weforum.org/publications/global-risks-report-2024/

Yamagishi, T. (2011) *Trust: The evolutionary game of mind and society,* Springer. https://doi.org/10.1007/978-4-431-53936-0

Yang, J.Z., Liu, Z., Wong, J.C.S. (2022) Information seeking and information sharing during the COVID–19 pandemic, *Communication Quarterly*, 70(1), pp. 1–21, https://doi.org/10.1080/01463373.2021.1995772

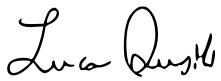
Young, L. D., Fitz, E. B. (2022) Who are the 3 Per Cent? The Connections Among Climate Change Contrarians. *British Journal of Political Science*, *52*(4), pp. 1503–1522. https://doi.org/10.1017/S0007123421000442

Zarouali, B., Dobber, T., De Pauw, G., & de Vreese, C. (2022) Using a Personality-Profiling Algorithm to Investigate Political Microtargeting: Assessing the Persuasion Effects of Personality-Tailored Ads on Social Media. *Communication Research*, 49(8), pp. 1066-1091, https://doi.org/10.1177/0093650220961965

Zhou, Y., Shen, L. (2021) Confirmation Bias and the Persistence of Misinformation on Climate Change, *Communication Research*, pp. 1–24, https://doi.org/10.1177/00936502211028049

Ziegler, A. (2017) Political orientation, environmental values, and climate change beliefs and attitudes: An empirical cross country analysis, *Energy Economics*, 63, pp. 144-153. https://doi.org/10.1016/j.eneco.2017.01.022
Zollman, K. (2010) The epistemic benefit of transient diversity, *Erkenntnis*, 72(1), pp. 17–35, https://doi.org/10.1007/S10670-009-9194-6

Zoonen van, L. (2012) I-pistemology: Changing truth claims in popular and political culture, *European Journal of Communication*, 27(3), pp. 56-67, https://doi.org/10.1177/0267323112438808



UNIONE EUROPEA
Fondo Sociale Europeo
Ministero dell'Università e della Ricerca
PON RICERCA E INNOVAZIONE
REACT EU
UniSR Università Vita-Salute San Raffaele

The PhD Fellowship was co-financed with funds from the National Operational Programme for Research and Innovation 2014-2020, by FSE REACT-EU funds
Action IV.4 "PhD Courses and Research contracts on innovation topics" and Action IV.5 "PhD Courses on Green related topics"

