# Durham E-Theses

## *On the Political Costs of Scientific Dissent*

NASH, ERIN,JADE

**How to cite:**

**Use policy**

# ON THE POLITICAL COSTS OF
# SCIENTIFIC DISSENT

Erin J. Nash

**Abstract**

Some speech acts about science can result in many non-experts in a political community holding false or inaccurate empirical beliefs.

But if people do hold false or inaccurate empirical beliefs, should we be concerned? If so, *why* should we be concerned? What's wrong, morally and politically speaking, with speech acts that misrepresent scientific testimony, or that otherwise count as misinformation about science?

Philosophers of science have largely neglected these questions, despite such questions often being tied to 'dissent' about science. Instead, they have almost exclusively focused on the epistemological implications and benefits of dissent for knowledge production and scientific progress. Even those philosophers of science who in recent years have displayed scepticism towards dissent have focused their arguments on its negative *epistemic* effects within, and on, the *scientific community* and *science* itself.

Instead, the aim of this thesis is to provide a better understanding of the nature of the *non-epistemic* consequences of dissent and the broader phenomenon of misinformation about science for *non-experts* and for *society* more broadly. To achieve this, I bring contemporary analytic social and political philosophy into conversation with current debates about problematic speech acts about science within philosophy of science.

I argue that some speech acts about science, including dissent, can interfere with and erode three core liberal democratic values: it can compromise *personal autonomy*; it can pre-emptively silence people, and thus constrain their *freedom of speech*; and it can undermine the *democratic legitimacy* of public policy. Moreover, I argue that these three consequences are interlinked.

I conclude my thesis with an argument for maintaining a focus on communicative ethics, and by offering a basic framework for reasoning through the highly context-dependent evaluations of, and judgments that need to be made about, scientific dissent within different parts of the public knowledge system.

# ON THE POLITICAL COSTS OF SCIENTIFIC DISSENT

A thesis submitted for the degree of

Doctor of Philosophy

by

Erin J. Nash

Department of Philosophy

Durham University

2018

# DECLARATION

I confirm that no part of the material contained in this thesis has previously been submitted for any degree in this or any other university. All the material is the author's own work, except for quotations and paraphrases which have been suitably indicated. The copyright of this thesis rests with the author. No quotation from it should be published without the author's prior written consent and information derived from it should be acknowledged.

Most of Chapter Five has been published as: Nash, E. J. 2018. In Defense of 'Targeting' Some Dissent about Science. *Perspectives on Science*, 26(3): 325-359. Thank you to *Perspectives on Science* and MIT Press for allowing me to republish this material here.

Some of text in the Introduction and Chapter Four appears in an article I co-authored with Robert Simpson on J.S. Mill's 'inseparability thesis', which is currently under review as Nash & Simpson: 'Practically Inseparable: Is Freedom of Speech Sufficient For Freedom of Thought?'

Erin J. Nash

# ACKNOWLEGEMENTS

How did I get here? When I look back on the various paths I took that eventually led me to completing a PhD, there are two things that stick out like sore thumbs: (i) Firstly, all of the radical contingencies along the way and the very different pathways I almost went down that were very unlikely to have also led here; and (ii) How I would not have made it to this point without the support—intellectual, emotional, and financial—of so many.

Firstly, and especially now knowing what I do about the extent to which the random circumstances a person is born into influences their future life trajectory, I feel very lucky to have such a wonderful immediate family—my parents, Judy and Peter Nash, and my sister, Tennille Nash. I'm sure that had my parents not not valued education as highly as they did, or self-sacrificed as much as they did to give Tennille and I the lives we have, that I wouldn't have written this thesis. I owe them all more than I can express.

I am also grateful to my step-family members in Washington State – Karen, Sadie, Nick, Jenna, and Toby – for their support of my work over the years.

I'm lucky that I had teachers at high school who encouraged me academically—especially Hilda Knight—and a few good friends who provided an oasis of sorts in the otherwise very anti-intellectual social environments we grew up in in the outer suburbs of Melbourne: thank you Natalie Roughan, Chris Roussos, Kacey Sinclair, and Julian Schokman.

During my first degree, and the years I spent working for the Department of Sustainability and Environment in Melbourne afterwards, I met many people who'd come to be some of my closest friends and greatest sources of support and encouragement: Heidi Zimmer, Tonia Cook, Rachael Newman, Cathy Ronalds, Dom Passaportis, and Alice Daly. I also miss working more closely with Jarom Stanaway, Marcus Salton, Mayumi

# TABLE OF CONTENTS

# CHAPTER ONE
# INTRODUCTION

In August of 1944, George Orwell attended a lecture by Oxford biologist John Baker at a PEN conference in London.[1] During this lecture, Orwell became aware of the growing influence of the anti-Mendelian genetics of Soviet biologist and agronomist Trofim Lysenko. In the U.S.S.R, Stalin was aggressively sponsoring an artificial expert *consensus* around Lysenko's specious science and engineering public *confidence* in its correctness (Caspari & Marshak 1965; Rossianov 1993; Stanchevici 2017). The resulting agricultural policies, based upon Lysenko's claims, led to famines that killed several million Soviets (deJong-Lambert 2012).[2]

According to Orwell's official biographer, the discovery that Stalin was persecuting scientists who dissented from the Lysenkoist orthodoxy, and that he was fabricating scientific results, was the catalyst for Orwell to begin working on *Nineteen Eighty-Four* (although the idea had first come to Orwell a year earlier) (Shelden 1991). Orwell remained deeply disturbed by Lysenkoism right up until his death in January 1950: the penultimate book he read was Julian Huxley's *Soviet Genetics and World Science: Lysenko and the Meaning of Heredity* (1949), and one of the final cuttings he pasted into his notebook was an article quoting Lysenko as maintaining that 'wheat can become rye' (Orwell 2010).

Curiously, half a century or so later, historians of science began to unearth the inverse of this phenomenon. In post-war western societies, scientific *dissent*[3] was being deployed

---

[1] As Orwell notes in a letter to C.D. Darlington, 14 March 1947.

[2] Lysenkoism was also adopted by Mao in China, leading to more famines and the deaths of millions of Chinese too (deJong-Lambert 2012).

[3] By 'dissent' I mean the act of not accepting, or of rejecting, a conclusion that is widely or officially held—i.e. an expert consensus—and/or the act of asserting a claim that is inconsistent with an expert consensus.

in a way that was not dissimilar to Stalin's campaign of misinformation: dissent was used to generate *confusion* and cast public *doubt* upon robust scientific findings. But in these cases, such strategies were being used not by the state but generally by large multinational corporations whose commercial products or services – tobacco, chemicals, waste, fossil fuels – harmed human health (e.g. Proctor 1996, 2011; Proctor & Schiebinger 2008; Michaels 2008; Oreskes and Conway 2010). For instance, for more than half a century, the tobacco industry and the scientists it employed publicly denied the existence of links between use of cigarettes and increases in morbidity and mortality. Moreover, the industry actively worked to obscure the knowledge of the harms that did exist, by claiming that the medical science and epidemiology were uncertain (Michaels 2008; Oreskes and Conway 2010). It also used advertising to draw positive associations between smoking and health.

According to the World Health Organisation (WHO) (2018), tobacco use kills one in two of its users. It is responsible for the premature deaths of more than 7 million people each year, and around 890,000 of these deaths are the result of non-smokers' exposure to smokers' second-hand smoke. The burden of tobacco-related illness and death is heaviest in low- and middle-income countries where 80% of the world's 1.1 billion smokers live (WHO 2018). Many tobacco users remain unaware of the health risks associated

---

I take dissent to differ from mere disagreement in that disagreement takes place when peoples' opinions or beliefs fail to accord on some matter 'X', but where no broad or official consensus can be said to exist regarding 'X'. I do not require dissent to involve an explicit rejection of the expert consensus, and I also understand dissent as encompassing cases where the speaker is not aware that an expert consensus exists, nor aware of the content of this consensus. My definition therefore departs from many ordinary usages of the term, e.g. those that hold that the mainstream consensus is known to the dissenter, and actively questioned or rejected. I depart from this way of understanding dissent, primarily because I am interested in actors who, even if they were not aware that they were making a dissenting speech act, 'should have known' that they were doing so, given epistemic and other obligations that they have. I want to make room for negligence and recklessness, as I think the literature to date overly focuses on deliberate acts of misleading others (e.g. in the work of Naomi Oreskes). This will become clearer in chapter two. It's also important to note at this stage that, given the focus on dissent within philosophy of science, this is the main speech act I investigate in this thesis. However, it's not the only speech act we might be interested in, and at times I talk about speech acts that misrepresent science or scientists' testimonies, or that otherwise misinform, including through omissions of information. These latter types of speech acts need not also be dissenting ones, but in this thesis I am interested in cases where they do tend to overlap.

with tobacco, and knowledge is often patterned by socio-economic class, with the poorest having the lowest levels of awareness.[4]

The stories of Lysenko and Big Tobacco have far more in common than it looks like at first blush. On the surface, these two stories involve very different phenomena. But the central agents play the same essential role: they work to distort people's perceptions of the epistemic landscape, to disorientate them, and ultimately, to manipulate their behaviour. These two strategies are different means to the same end: the goal of controlling mass opinion. Importantly, such manipulation does not require that *every* person be totally uncoupled from the relevant knowledge for it to be effective at achieving a political or financial goal; there need only be *a sufficient number of* people in a population who are detached and disorientated to an *adequate degree*.

Concerns about the power of public opinion, and of it being able to be captured and manipulated, have a long history. For instance, it was British philosopher J.S. Mill's concerns about the tyrannising potential of public opinion – rather than state power, as such – which primarily motivated his treatise *On Liberty* (1859/1991). In a letter to Harriet Taylor in January 1855, Mill says this work is urgently needed because 'opinion tends to encroach, more and more on liberty' (1972: 294). The sentiment appears in the essay itself, when he says that the greater danger to human freedom comes from a s*ocial* tyranny. Such social pressure is 'more formidable than many kinds of political oppression', because 'it leaves fewer means of escape, penetrating more deeply into the details of life, and enslaving the soul itself' (1859/1991: 8-9).

By the early years of the 20th century, scholars like Edward Bernays—Freud's nephew and the so-called 'father of public relations'—were conducting research into shaping public opinion and *teaching* these 'techniques' to practitioners to control human psychology in

---

[4] For example, in Indonesia, which is the fastest growing market for tobacco globally. Table 9.1 in WHO (2011: 83-85) presents the percentage of adults who believe that smoking causes serious illness, stroke, heart attack or lung cancer by current smoking status and selected demographic characteristics. For both smokers and non-smokers, the results are starkly patterned by education level. E.g. whilst 95% of those who have completed high school believe that smoking causes serious illness, only 66% of those who have completed less than primary school believe this. Similarly, whilst 75% of those who've completed a higher education degree believe that smoking can cause stroke, only 23% of those who haven't completed primary school do.

'desirable' ways (Bernays 1923, 1928a).[5] In the *American Journal of Sociology* in 1928 he wrote an article called 'Manipulating Public Opinion, the Why and the How'; in 1947 the *Annuals of the American Academy of Political and Social Science* featured an article by Bernays called 'The Engineering of Consent'. Chapter one of his book *Propaganda* (1928a), opens with the following assertion:

> The conscious and intelligent manipulation of the organized habits and opinions of the masses is an important element in democratic society. Those who manipulate this unseen mechanism of society constitute an invisible government which is the true ruling power of our country. We are governed, our minds are moulded, our tastes formed, and our ideas suggested, largely by men we have never heard of.... It is they who pull the wires that control the public mind.

It is well-known that Joseph Goebbels, the Minister of Propaganda of Nazi Germany from 1933 to 1945, was an avid follower of Bernays' theories.[6] Bernays learned that the Nazis were using his work in 1933 from a journalist. He later described receiving this news in his 1965 autobiography:

> They were using my books as the basis for a destructive campaign against the Jews of Germany. This shocked me, but I knew any human activity can be used for social purposes or misused for antisocial ones.

Given that the use of state-based propaganda and severe speech suppression helped to prop up the totalitarian regimes that blighted Europe in the 20th century, it's hardly surprising that the focus in western liberal democracies since then has been squarely on the threat of the state and the resources it possesses to control public opinions (Carey 1995; Stanley 2015). Over the past half a century or so, these societies have developed strong associations between dissent and truth-telling, and dissent and liberation. Today, dissent is celebrated across the political spectrum, and generally rightly so.

---

[5] During WW1 Bernays had worked for the government's Committee on Public Information to build support for the war. He called this work 'psychological warfare' and viewed it as having great potential for peace-time applications (Ewan 1996).

[6] Even despite the fact that Bernays was Jewish. That he was makes this an even more tragic story.

A willingness to challenge accepted ideas – that is, to dissent – is now commonly emphasised as a key component, if not the defining feature, of the scientific ethos too.[7] Dissent within and about science been valorised by philosophers of science for its *knowledge generating potential* (Mill [1859] 1999; Popper 1959, 1972: 352; Feyerabend 1975, Longino 1990, 2002); its capacity to help build and maintain *empirical successes* (Solomon 2001, 2006, 2008); and its ability to infuse scientists' conclusions with *objectivity* (Longino 1990, 2002). It has also been celebrated by science studies scholars for its role in fostering public confidence in science, in promoting trust between experts and the public (Beatty and Moore 2010; de Melo-Martin and Intemann 2018), in helping to incorporate 'lay expert' and indigenous knowledges into scientific inquiry (Epstein 1995), and in resisting technocracy (Jasanoff 1998). Some STS scholars seem to endorse Mouffe's (1996) contention that *consensus* is the real threat to democracy.

But in more recent years, several philosophers have started casting a more critical eye on dissent. These philosophical criticisms have tended to focus on the *epistemic consequences* of dissent *for science* and the impacts of dissent on *scientific communities*. For instance, Miriam Solomon (2001) was one of the first philosophers of science to question what we might even call the 'consensus' that had emerged within philosophy of science about the value of dissent in science and the disvalue of consensus — but her criticisms and prescriptions only accounted for *epistemic* values and risks:

> It follows that neither dissent nor consensus is invariably or intrinsically valuable for scientists; sometimes dissent will maximise empirical success and truth, and sometimes consensus will. In saying this, I am not only disagreeing with those who claim that consensus is intrinsically normative, but also with Mill, Feyerabend, Longino, etc., who see intrinsic merit in dissent. Obviously, *dissent will be appropriate when different theories have different empirical successes*. On the other hand, consensus *will be appropriate when one theory has all the empirical success*. (2001: 101).

---

[7] This emphasis was perhaps bolstered in light of the roles of enforced consensus and group-think among scientists in the atrocities committed by the Nazis and in the Soviet Union. I'd like to stress that whilst there is no doubt that dissent in science, and about the empirical claims of scientists, can be extremely valuable, what motivates this thesis is that some dissent has almost certainly failed to contribute to, or has actively hindered, knowledge and scientific progress (Biddle and Leuschner 2015; Leuschner 2016; Biddle et al. 2017), and undermined non-epistemic goals and values too (Kitcher 2011). Some dissent is therefore *problematic*.

Similarly, Biddle and Leuschner's (2015; also see Biddle, Kidd and Leuschner 2017) recent critique of dissent about science is framed as a concern about *epistemically* detrimental dissent, where the detriment relates to the knowledge *of scientists* and progress *within science.* The first key difference between this thesis and the majority of the philosophical literature to date (with key exceptions being Kitcher 2011, Furman 2016, and Moore 2017), is that I focus on providing a better understanding of the *non-epistemic consequences* of *problematic* dissent for *non-experts* and for *society more broadly*, i.e. beyond the scientific domain.[8] To do so, I bring contemporary analytic political philosophy into conversation with the current debate about dissent within philosophy of science.

Other science studies scholars, whilst not denying that dissent can sometimes be problematic, and acknowledging that it sometimes has negative non-epistemic consequences, have for the most part argued that dissenting speech acts *as such* are not the 'real problem'. A recent article by de Melo-Martin and Intemann (2014), for example, argued that other scholars' attempts to 'target dissent' have been 'misguided and dangerous'. In their more recent book — *The Fight Against Doubt* (2018) — they urge us to shift our attention *away* from dissent and back to the conduct of those scientists who subscribe to the consensus and the institutional structures within which scientific research takes place. Although I agree that scholars ought to pay more attention to the background conditions within which dissent arises and takes hold, I think that de Melo-Martin and Intemann's directive is far too hasty. For one thing, finite resources force scientists, editors, journalists, politicians, librarians, social media managers and so forth to make judgments about which speech about science to prioritise. Where such decisions are inescapable, or shouldn't be avoided because of professional and moral obligations, philosophers still have a valuable role to play in interrogating the communicative ethics of different speech acts.

Moreover, given that the vast majority of science communication is mediated and controlled by private media corporations detached from democratic control, and that these

---

[8] For the purposes of this thesis, I understand problematic dissent to be a communicative act about science that we have good reasons think is unsubstantiated, or that misrepresents another's position, or first- or higher-order evidence, and that we reasonably believe constitutes a significant risk to some value(s).

entities often enjoy freedoms and powers that others' do not,[9] there is a reason to think that intermediaries deserve *more* scrutiny at this present moment, not less — including scrutiny of the impacts of their communicative practices on people and society. What is true of the threat that science and scientists pose to democracy is equally true, mutatis mutandis, of the owners and managers of the means of transmitting (mis)information to the public. Given that ordinary people often have no practical way outside the media of receiving expert testimony, and that scientific factual information is an important input into many decisions in our increasingly technologically-advanced societies, critiquing these intermediaries' practices seems like an especially important task.

My aim in this thesis is therefore to go some way towards a deeper understanding of these dynamics, and in particular, to identify some key *non-epistemic* consequences of problematic dissent. I argue that some dissent can interfere with and erode three core liberal democratic values: it can compromise personal autonomy; it can pre-emptively silence people, and thus constrain their rights to freedom of speech; and it can undermine the democratic legitimacy of public policy. Moreover, I argue that these three consequences are entwined, and that together they threaten to further exclude the public from decisions in technical domains.

I start, in Chapter Two, by defending the position that false beliefs are the sorts of things that can undermine our personal autonomy. In doing so, I illustrate two broad ways that false beliefs can curb an individual's autonomy: by undermining their control over their choice of values, and by undermining their control over whether or not their actions realise their values.

I then consider the conditions under which false beliefs can and cannot compromise personal autonomy. I critically evaluate Killmister's, Taylor's, and Mele's criteria, and find them all wanting. I then develop an alternative account based on the notion of voluntary and non-voluntary false beliefs, which I connect to the extant literature on 'voluntariness' from political philosophy and on the 'should have known' phenomenon

---

[9] Due to press freedoms and monopoly-like power over markets (e.g. the Murdoch-owned corporations control >70% of the print circulation in Australia; Facebook owns the three most popular social media apps, etc.).

within social epistemology. I argue that the voluntariness of our beliefs is crucial to whether we can subsequently be considered to have acted autonomously, and therefore also to what we are substantively responsible for.

Building on this, in Chapter Three I explore the implications of compromising a person's autonomy on their freedom of speech. My aim is to challenge the background assumption held by some philosophers of science in debates about scientific dissent that the value of freedom of speech is necessarily only, or best, served via a thorough-going toleration of all forms of dissent about science in all parts of the public knowledge system.

To make my argument, I draw on feminist free speech theory developed by Catherine MacKinnon, Rae Langton, Jennifer Hornsby, Caroline West, Mary-Kate McGowan, and Lorna Finlayson. Although most of this scholarship has developed in the context of pornography and hate speech, I use its insights into 'silencing' and apply them to social relationships that have not (to my knowledge) been viewed through this lens: the relationships between scientific experts, intermediaries who communicate science, and the public.

My contribution here is to develop an account of a new type of silencing that captures and explains one of the non-epistemic wrongs and harms of problematic scientific dissent. First, I survey existing accounts of silencing and argue that the taxonomy is incomplete because the accounts all share two common background assumptions that often do not hold. From there I present a novel conception of silencing that I call 'pre-emptive' silencing. As we shall see, there are compelling reasons in favour of regarding other types of silencing as free speech violations; so too, I will argue, for pre-emptive silencing. Although additional work would need to be done to show that any instance of pre-emptive silencing is in fact a free speech violation, I demonstrate that this type of silencing is both philosophically interesting and deserving of further attention.

In Chapter Four, I argue that because of the impacts of some scientific dissent on the autonomy and freedom of speech of the members of a political community, it follows that this dissent has the ability to undermine the democratic legitimacy of public policy.

I first argue that other scholars' concerns about the impacts of scientific dissent on public policy and democratic values (such as Philip Kitcher's and Naomi Oreskes') accord most closely with Rational Epistemic Proceduralist accounts of democratic legitimacy. However, using Fabienne Peter's criticisms of such accounts, I argue that these scholars make unnecessary and controversial assumptions. I then introduce Peter's own account of democratic legitimacy, Pure Epistemic Proceduralism, and argue that some dissenting speech acts about science interfere with the degree to which conditions of political fairness, including epistemic fairness, are met; hence, such dissent can block or erode the democratic legitimacy of public policy. This argument sheds light on the democratic responsibilities of both scientists and those actors within the public knowledge system who mediate experts' testimonies, such as journalists and politicians. If we accept that, as Moore (2017) argues, expert authority depends on the exercise of public judgment and live possibilities for non-experts to protest, oppose, and scrutinise the factual claims of scientists, we must attend more closely to the ways that intermediaries convey information between experts and non-experts.

In Chapter Five I present an argument against the claim that certain ways of responding to dissent are 'misguided' and 'dangerous'. I then offer a basic framework for reasoning through the highly context-dependent evaluations of, and judgments that need to be made about, scientific dissent within different parts of the public knowledge system.

I conclude, in Chapter Six, by providing a summary of the arguments presented in this thesis and pointing to some potential future research avenues.

# CHAPTER TWO
# (MIS)INFORMATION, FALSE BELIEFS,
# & AUTONOMOUS AGENCY

*Introduction*

One of the primary *epistemic* consequences of problematic dissent about science is that such dissent can, and often does, result in many non-experts holding false or inaccurate factual beliefs about the empirical world. But should we be concerned about people holding false or inaccurate beliefs, and if so, why? In other words: what are the *non-epistemic* consequences of misinformation-induced false beliefs?

The first part of my tripartite answer to this question is that false and inaccurate factual beliefs can compromise our *personal autonomy*; they can undermine our ability to govern our own lives through our deliberations and actions.[10] This view has been defended by Alfred Mele (1995/2001: 179-182), James Stacey Taylor (2009), and more recently, by Suzy Killmister (2013a). Other social and political philosophers have also made passing remarks which indicate they agree. Sisela Bok (1978: 20–21), for one, suggests that a key reason why the victims of liars feel resentful is that they 'see that they were manipulated, that the deceit made them unable to make choices *for themselves* according to the most adequate information available, unable to act *as they would have wanted to act* had they known all along' (my emphases). Although Bok and others focus on *lies*, and thus on false beliefs brought about by *deliberate* deception, I show here how we can also be wronged when our false and inaccurate beliefs are the product of others' negligence and recklessness. I think victims of negligent and reckless misinforming speech acts are likely to—rightly—feel a similar sense of resentment, because these behaviours also affect their autonomy (i.e. their ability to make choices for themselves) as a consequence of the speaker's failing to meet certain obligations or to take due care.

---

[10] I provide a more detailed definition of personal autonomy in the next section.

The conclusion that false beliefs have the potential to undermine personal autonomy is, however, contested — particularly by those philosophers who are *internalists* about personal autonomy. An internalist holds that personal autonomy depends entirely on the person's mental state or structure at the point of action. On internalist accounts, personal autonomy is independent from external reality, i.e. a person's past circumstances that cause her to hold some mental state; her present circumstances in response to which she is moved to act; or the future circumstances that come about as a result of her act. Michael McKenna (2005), for example, rejects the contention that false beliefs can undermine the autonomy of a person's actions. Likewise, Nomy Arpaly says she doubts 'whether anyone wishes to claim… that an ill-informed decision cannot be an instance of autonomous agency' (Arpaly 2005: 175). For Arpaly, 'ill-informed' even includes those decisions that are linked to deception.

Given the pride of place personal autonomy has within contemporary moral, social, and political philosophy, as well as applied ethics,[11] this dispute has particularly far-reaching consequences for both theory and practice. Autonomy is invoked in defence of a wide range of rights, including the right to free speech (or, put differently, the freedom to dissent over empirical claims — a right I explore in the next chapter). Many of those defences of free speech that appeal to uninhibited public discourse as being essential for a well-functioning democracy can also be viewed as grounded in an ideal of autonomy (e.g. see Post 1993, a value I explore in Chapter Four).

Autonomy is also importantly related to judgments about *responsibility*, often via the notion of *voluntariness*. A necessary condition for holding people substantively responsible[12] for their choices and actions is that those choices and actions were voluntary (Olsaretti 2008: 119). And as Olsaretti (2004) suggests, a concern for voluntary choice can be grounded in respect for autonomy. Several important questions in moral, social and political philosophy are therefore likely to turn on, firstly, who is correct about

---

[11] Exemplified, perhaps, by the focus on 'informed consent' in medical ethics.

[12] The term 'substantive responsibility' is Thomas Scanlon's (1998, ch. 6). Roughly, substantive responsibility has to do with the obligations people have towards each other' and holding agents liable for their actions (p. 290).

whether false and inaccurate beliefs compromise autonomy, and secondly, if false beliefs can compromise autonomy, whose account of the conditions under which this can occur is correct.[13]

I do two things in this chapter. First, I defend the position that false beliefs are the sorts of things that *can* undermine personal autonomy (§2.1.). In doing so, I provide a sketch of two broad ways false beliefs can curb a person's autonomy: by undermining their control over their choice of values, and by undermining their control over whether or not their actions realise their values. Second, I aim to provide a better understanding of the conditions under which false beliefs can and cannot compromise personal autonomy, with the view to identifying when, and why, the transmission of misinformation through knowledge and information systems is problematic. In §2.2 I argue that that none of the accounts currently on offer are satisfactory: I critically evaluate Killmister's (2013), Taylor's (2009), and Mele's (1995/2001) criteria for when false beliefs compromise autonomy and find them all wanting. Then, in §2.3, I offer an alternative, positive account based on the notion of voluntary and non-voluntary false beliefs, which I connect to the extant literature from political philosophy on voluntariness, as well as that from epistemology on the 'should have known' phenomenon. I then briefly discuss the conclusions my framework would yield about personal autonomy in the examples presented in earlier sections (§2.4), as well as the implications of my account for the debates within political philosophy on voluntariness and responsibility (§2.5). I argue that the voluntariness of our beliefs is crucial to whether we can subsequently be considered to *act* voluntarily or not, and therefore for what we are substantively responsible for. These findings have important implications for debates in democratic societies about relationships between experts, intermediaries and the broader public.

---

[13] Additionally, although autonomy has generally been associated with liberalism, concepts similar to autonomy can be found in other comprehensive ideologies too, such as: the importance of non-domination, or independence from arbitrary power, in contemporary civic republicanism; the alienation of the worker from their *Gattungswesen* ('species-essence') in Marxism; and concerns about socialisation in feminism (Killmister 2013b). It's therefore possible that the implications of my analysis will be of relevance beyond the liberal political paradigm too, though I cannot explore this possibility further here.

## 2.1. Two Ways False Beliefs Can Undermine Autonomy

### 2.1.1. What is Personal Autonomy?

Although the definition of personal autonomy is contested, all definitions share a common idea: that of *self-governance* (MacKenzie and Stoljar 2000: 5, Colburn 2010: 4, Killmister 2017). According to one standard view (Frankfurt 1971; Watson 1975), autonomy is simply a matter of our second-order and first-order desires displaying coherence. But this can't be sufficient for self-government, because these formal conditions could be satisfied in a case in which a person had been systematically indoctrinated or brainwashed into the attitudes that they identify with and that structurally cohere. Intuitively, we wouldn't want to say that such a person is autonomous with respect to attitudes that formed because another agent ran roughshod over the person's independence of mind.

I therefore favour Alfred Mele's (1995/2001) dual-aspect view of autonomy, in which a person requires two distinct forms of control for self-government.[14] First, they must have control over their values, desires, preferences, and related intentions. I'm going to call this type of control *reflective autonomy*.[15] Second, Mele requires a person to have control over *whether or not* their actions succeed in realising their values, desires, preferences and intentions in the world. I call this component *practical autonomy*. Practical autonomy

---

[14] I am adopting Mele's (1995/2001) account of autonomy, with some minor modifications, for my purposes in this chapter. Mele's account of autonomy is one of the most well worked-out and respected accounts of autonomy. As Mele is interested in what makes a person autonomous or non-autonomous, he requires someone to have 'sufficient' control over both functions. But like Killmister (2017), I take autonomy to be a scalar concept, and so I'm primarily interested in what renders someone more or less autonomous (i.e. what enhances or curtails autonomous agency). I have therefore dropped the sufficiency clause in adopting Mele's conception of autonomy. My concern in this chapter is also with more 'local' types of autonomy, as opposed to 'global' (or lifetime) autonomy. I am concerned with whether in, a particular instance, an individual believes, desires, intends, or acts autonomously, rather than whether she counts more generally as an autonomous person.

[15] Mele calls this 'psychological autonomy'. On my account, as with Killmister's (2013a), reflection can be affective, involve the emotions and the imagination, and take place below the level of conscious awareness.

refs to the extent to which a person causally determines *whether* they succeed or fail to realise their will in the world.[16]

I take Mele's account to share important similarities with two other accounts of autonomy that I will also be drawing on in this chapter: those of political philosophers Ben Colburn and Suzy Killmister. Colburn defines autonomy as 'an ideal of people deciding for themselves what is a valuable life, and living their lives in accordance with that decision' (2010: 19). The first part of Colburn's definition ('people deciding for themselves what is a valuable life') aligns with Mele's requirement that people have control over their values, desires, preferences, and related intentions (reflective autonomy). The second part of Colburn's definition ('living their lives in accordance with that decision') accords with Mele's requirement that a person has control over whether or not their actions succeed in realising their ends (practical autonomy). Like Mele, putting one's choices about what is valuable into action is part of what it means to be autonomous for Colburn. It is also part of what it means to be autonomous for Killmister too. Killmister defines personal autonomy as being '*in control of* one's interactions with the world' and it is 'to have those interactions *correspond to* one's self' (2013b: 110).[17] Thus, the second part of Killmister's definition aligns with the first aspect of Mele's and Colburn's definitions, and the first part of Killmister's definition is similar to the second component of Mele's and Colburn's. It is also worth noting that all three accounts of autonomy broadly map onto what J.S. Mill, in *On Liberty*, called 'the only freedom which deserves the name', which is 'that of pursuing *our own good* in *our own way*' (1859/1999: 55).

Suggesting that people should have reflective and practical autonomy is not to say that people are, or should be, maximally self-sufficient and causally isolated from others. As

---

[16] Note that as I understand autonomy, whether one succeeds or fails does not, in itself, bear on one's degree of personal autonomy; one could fail to realise their values or intended actions but still be autonomous, even optimally or maximally so. What is required is just that it is a person's own agency that determines *whether* they succeed or fail. For example, I might intend to sink at least one of my free throws to win the game but miss them both. But my failure to do so was the product of my autonomous agency.

[17] All three scholars are therefore externalists about autonomy, in that they all reject that an agent's autonomy depends entirely on their mental state or structure at a point in time.

socially and historically embedded beings, we are at least partially constituted by historical contingencies and social relations. Rather than self-sufficiency being the hallmark of autonomy, feminist analytic philosophers in particular have convincingly argued that our autonomy, and our very capacity to be autonomous, *requires* interpersonal connections, and even sometimes relations of dependency, with others (Mackenzie and Stoljar 2000). Social relationships and socialisation can enhance an agent's capacity for autonomy when these relationships and processes assist us to develop capacities for critical reflection, or to cultivate self-trust, self-confidence, and so on (Friedman 2003: 97). Therefore, what is important is not that we are partly dependent on others, but that the nature of our interpersonal relationships are *appropriate*, especially in terms of the *degree* to which these relations shape our lives, and the *kinds* of ways in which they do so. We therefore need to be able to distinguish between legitimate and illegitimate ways of influencing each other (Dworkin 1988: 11).

Next, I provide a sketch of two key ways false beliefs have the potential to undermine personal autonomy: first, by undermining a person's control over their choices about their values and ends, i.e. their reflective autonomy (§2.1.2), and second, by undermining a person's control over whether or not their acts succeed in realising the values and ends they have adopted, i.e. their practical autonomy (§2.1.3).[18]

### 2.1.2. False Beliefs & Reflective Autonomy

I take 'values' to refer to those things or qualities that we deem to be favourable or pursuit-worthy. Our values are sometimes referred to as our 'ends' or are described as those things that constitute our conception of 'the good'. Our valuing practices involve instrumental and non-instrumental valuation, and include — but are also broader than — the formation and revision of *preferences.*[19]

---

[18] As will be made clear in due course, the presence of certain types of information (and the absence of certain types and/or amounts of misinformation) within a person's socio-epistemic environment are, on my account, constitutive of these 'control conditions'.

[19] I count preferences as types of values here. A preference is the outcome of an overall or total comparative evaluation in which someone has compared and ranked alternative courses of action with respect

False and inaccurate beliefs can lead an individual to valuing something, or not valuing something, or under- or over-valuing something, that they would or would not value, or would value to a greater or lesser degree, if they had true or accurate beliefs. False and inaccurate beliefs can also lead to someone to interpreting and understanding their values differently, even if in some sense they would remain unchanged upon the formation of accurate beliefs.[20] This in turn means that false beliefs can cause individuals to make different choices than they otherwise would have made if they had true or accurate beliefs. Consider the following example:

---

**Tennille 1**

Tennille lives in the U.S. in the 1950s. She highly values the pleasure she derives from smoking cigarettes (value 1, $V_1$). Whilst Tennille also values her bodily health ($V_2$), she does not believe that there are any harmful health risks or consequences associated with smoking (belief 1, $B_1$). Tennille's $B_1$ belief has been significantly influenced by the tobacco industry's misinformation campaign, which has featured prominently within Tennille's socio-epistemic environment. Her beliefs, mixed with her values, mean that Tennille has a strong preference for smoking ($V_3$) vis-à-vis not-smoking ($V_4$).

However, by the 1970s, the tobacco industry's misinformation campaign has waned — in part due to increasing government regulation of speech by the tobacco industry — and citizens are now better informed about the links between smoking and cancer. Tennille has come to accept that smoking tobacco is a significant risk factor for many diseases and avoidable premature mortality ($B_2$). Although Tennille still values both her health ($V_2$) and the pleasure of smoking ($V_1$), her revised empirical beliefs (i.e. $B_2$) prompt her to reconsider her all-things-considered judgement about the value of smoking ($V_3$) vis-à-vis not smoking ($V_4$). This leads to a reversal of Tennille's preferences because she now attributes more value to not smoking than she does to smoking (i.e. she now perceives not smoking to be more worthy of pursuit than smoking). Furthermore,

---

to most (in cases of an overall comparative evaluation) or all (in cases of a total comparative evaluation) of what matters to them (Hausman 2012: 1-9). The inputs into this evaluation include (other) types of values (including desires), as well as beliefs about various facts that pertain to the decision problem (Hausman 2012: 5-6).

[20] E.g. even if 'the equality of all persons' remains someone's fundamental value, more accurate empirical beliefs could plausibly lead them to revising their understanding of what exactly that value amounts to.

her preference for not smoking is now so weighty that she decides to quit smoking alto-
gether.

It's plausible that had the tobacco industry's misinformation campaign not featured in Tennille's socio-epistemic environment, or had it not featured as prominently as it did, Tennille would have been more aware of, and would have more readily accepted, the links between smoking, disease and premature death, perhaps even as early as these links were known in the 1950s. Since Tennille's preferences reversed upon being apprised of this knowledge, in the possible world in which misinformation was absent (or reduced), it's likely that Tennille would have come to the decision to quit smoking much earlier in her life. So misinformation, through a causal influence on Tennille's beliefs, not only altered Tennille's beliefs, but also altered some of her values — in this case, her prefer-ence for smoking vis-à-vis not-smoking, as well as how strongly she held that value.

Because Tennille's misinformed state was the product of others' deceptive agency — as it has been shown that the tobacco industry was aware of the link between smoking and cancer as early as the mid-1950s, but that it continued to publicly dispute this link (Proc-tor 1996, Michaels 2008, Oreskes and Conway 2010) — in this case it's clear that exter-nal forces (other agents) played a significant role in controlling and sculpting the con-tours of Tennille's beliefs and values. Misinformation, through the false beliefs it in-duced, eroded *the extent to which it was Tennille who had control* over her choice of values; Tennille's reflective autonomy was therefore undermined.

Of course, there will have been others whose preference for smoking, and the strength of that preference, would have been robust between misinformation-laden and misin-formation-free worlds (i.e. the difference in informational and doxastic states would not have altered their preferences).[10] However, the steady decline of both the proportion of adults who are smokers and cigarette consumption in most Western societies since the 1970s (Gallup 2016) is evidence that the value people place on smoking is declining. As such, it is reasonable to think that there were many 'Tennilles' — that is, many people

---

[10] In these cases, whilst the presence of misinformation did not harm these people, it may have still wronged them.

whose autonomy was harmed — in societies in which the tobacco industry's misinformation campaign flourished.

### 2.1.3. False Beliefs & Practical Autonomy

To be autonomous, Mele requires that a person not only have (sufficient) control over their choice of values, but also that they have sufficient control over their capacity to make progress towards bringing about a state of the world in which those values are instantiated. To achieve this, a person generally needs to move from some situation A, where value V is not instantiated, to some situation B, where V is instantiated. To get from A to B, they will partly rely on information to help them construct a viable path between A and B that is also adequately in harmony with other values they hold.

Misinformation can frustrate a person's ability to realise their values in at least two key ways. First, misinformation can cause people to hold false or inaccurate beliefs about *how* to achieve their values, therefore thwarting their understanding of the best means to their end. Subsequently, the actions they carry out may at best result in limited progress towards achieving some end, and at worst lead them further away from realising that end. Again, consider Tennille, who upon being informed, rather than misinformed, about the health risks of smoking tobacco, has decided to quit smoking.

> **Tennille 2**
>
> To move from situation A (smoking, on average, a packet of cigarettes per day), to situation B (complete cessation of smoking), Tennille requires information about the range of means available to help her quit smoking. Tennille's primary information source is her general medical doctor. Tennille's doctor provides her with information about the range of options available for smoking cessation (i.e. options M/N/O/P) together with advice about which strategy the current evidence base suggests is most efficacious, which he tells her is option M. Tennille selects M.
>
> However, it turns out that Tennille's doctor has in fact *mis*informed her about which strategy is most efficacious. The current evidence base actually suggests that the *only* strategies that have performed better than mere chance are O and P, and P has been the most efficacious strategy for the highest proportion of people like Tennille. In fact, there is no evidence that option M works. M is actually a substance that Tennille's doctor owns the patent for, and it generally enhances, rather than dulls, nicotine cravings. As a result, Tennille's addiction becomes entrenched, and she finds herself further from realising her values. Had Tennille's doctor provided her with information, rather than misinformation, she would have chosen option O or P and successfully quit smoking.

There are variations of this example that illustrate how someone's knowledge of the best means to their end can be jeopardised in similar ways. For instance, Tennille's doctor could correctly inform her that P is known to be the most efficacious strategy, but misinform her about which strategy she was in fact pursuing once she chose to pursue P; perhaps, for example, he tells her that the medication in the bottle is P, when it is actually M.

The second key way that misinformation can frustrate a person's ability to realise their values is when misinformation about the foreseeable costs, risks, and benefits associated with each of the options leads people to hold different all-things-considered preferences for strategies than they otherwise would have absent this misinformation. For instance, misinformation about potential side-effects may result in someone selecting an option that conflicts with and threatens other values they hold (i.e. values that are not at that

moment the primary object of reflection and action). Thus, they may end up not actually realising a number of their values, and this failure would not be appropriately attributable to their own agency. For example:

---

**Tennille 3**

Imagine that although Tennille is provided with information about her complete option set, and despite being correctly informed about which option is most efficacious, Tennille's doctor misinforms her about P's side-effects. Unbeknownst to him, initial clinical trials actually found P's side-effects to include significant neuropsychiatric risks, including the development of psychosis, especially in people who have previously had psychotic episodes. The pharmaceutical company that developed P actively worked to conceal this information by withholding these results and designing them away through methodology alterations in subsequent trials. The result is that Tennille's doctor, himself misinformed, gives advice to Tennille that is in fact misinformation, and which causes her to hold false beliefs about P.

Importantly, as Tennille has previously suffered from psychosis and places a very high value on being psychosis-free, she would be unwilling to accept certain levels of risk to her mental health. If she was properly apprised of P's neuropsychiatric risks, these risks would have acted as a constraint on Tennille's choice; she would have ruled out P as a possible means to a smoke-free life and selected O instead. But as this information was not transmitted to her, she bases her decision on misinformation and chooses to take P, which triggers psychosis.

---

Therefore, the chain of misinformation, including omissions of information, from the pharmaceutical company to Tennille's doctor to Tennille — *omissions and misinformation about empirical facts that knowledge did in fact exist about, or could be legitimately expected to exist* — frustrated Tennille's ability to live her life in accordance with what is valuable to her. Mele (1995/2001: 180) would describe Tennille as having been *informationally cut off from governing her life autonomously*. Her flawed epistemic state meant that she had no control over whether, given that she acted with the goal of improving her health and avoiding psychosis, she would succeed. False beliefs compelled Tennille to select an option that, unbeknownst to her, harmed one of her most cherished values (i.e. her mental health), even though it helped her to achieve another (quitting smoking). Under these

circumstances, Tennille's autonomy was diminished in comparison to a similar, but misinformation-free, counter-factual world.

The depth and breadth of reflection and deliberation Tennille could engage in about which action/s to take was undermined and distorted by external forces, i.e. another agent's actions or omissions. In each of the examples outlined so far, rather than having a fair chance to succeed, owing to misinformation and omissions in information, Tennille could not make progress towards her values; in some cases, misinformation led Tennille further away from them. The transmission of misinformation resulted in a situation in which certain preferences (values) became more likely to be arrived at than others in a way that wasn't respectful of Tennille's autonomy. To the extent that such misinformation causally affected Tennille's decisions and subsequent actions, we can say that there was something about these choices that were not 'Tennille's', and thus that misinformation — and the false beliefs it induced — undermined Tennille's personal autonomy.


## 2.1.4. The Internalist Objection (& Reply)

Some philosophers, such as Michael McKenna (2005), would object to my conclusions about Tennille's autonomy in each of the examples presented so far. McKenna holds that if a person attempts one act but, by virtue of a false or inaccurate belief, mistakenly does another, their autonomy should not be considered to have been affected. To support his position, McKenna (2005: 208–209) provides the following example:

> **Tal:**
>
> '[T]he perfectly competent morally responsible agent Tal arrives at Daphne's house and discovers her unconscious and in immediate need of the drug known as "The Good Stuff". Urgently searching through Daphne's cupboard, Tal finds a bottle marked "The Good Stuff". He takes from it the prescribed dose and gives it to Daphne. Unfortunately Daphne's fumbling pharmacist has accidentally given Daphne the wrong drug, "The Bad Stuff", a drug that will kill people with Daphne's condition. Daphne dies. Tal had no reason to be suspicious of Daphne's pharmacist, and had very good reason to believe that the bottle in her medicine cupboard marked "The Good Stuff" actually contained "The Good Stuff" and not "The Bad Stuff".'

McKenna concludes that Tal acted autonomously in this scenario: 'However we unpack the notion of "rule" in "self-rule"... there is no reason to think that Tal was not ruling himself in acting as he did' (2005: 209). But he claims that Tal was not *morally responsible* for his actions, as he could not have reasonably foreseen the consequences of the action he carried out, and his action was based on a self-endorsed intention. For McKenna, being the recipient of misinformation about what substance was in fact in the bottle, and coming to holding a false belief about its contents, affects determinations of moral responsibility, but not of autonomy.

As Killmister (2013a: 517) points out, McKenna's conclusion turns on his premise that autonomy concerns acting on the basis of an intention that one endorses. In giving the drug to Daphne, Tal was following his self-endorsed intention e.g. 'we should do what we can to rescue those who are suffering.' For McKenna, an autonomous decision-making process only has *internal referents*: an action is autonomous if and only if a person endorses the intention that prompted her action. As such, it is clear that McKenna would also take Tennille to be autonomous in all of the examples presented so far, as Tennille endorsed all of the intentions that moved her to act as she did.

However, against McKenna's general position, the first example demonstrated how false beliefs facilitated by misinformation can affect the very content of a person's values, desires, principles, motivations and intentions 'a long way down'. Recall that in this example, Tennille only preferred (valued) smoking over not smoking by virtue of her

misinformation-induced false beliefs. So if the presence of false or inaccurate beliefs determines, or significantly causally contributes to, the content of a person's values, principles, motivations or intentions, it seems that the extent to which these attitudes can be said to have been autonomously endorsed is reduced. This is why other scholars, such as Mele, require agents to have control over both their choices of values, as well their decisions about how to pursue those ends; this way, an agent's autonomy 'up-stream' from any decision point or action is also secured.[13] On Mele's account, if a person's upstream values and other pro-attitudes have not been formed or adopted autonomously, the autonomy status of linked downstream values, choices, and actions is also affected. In discussing whether a person (Beth) who has been brainwashed into having particular ends can deliberate autonomously about how best to achieve those ends, Mele says: 'In virtue of her being compelled to have these ends or goals, Beth's deliberation with a view to their achievement is not autonomously conducted' (1995/2001: 178). When compelled pro-attitudes are input into a reflective process, their lack of autonomy taints the autonomy status of related downstream deliberation and outputs.

Furthermore, as Killmister (2013a) argues, McKenna's position only makes sense if we accept the background assumption that autonomy does not require there to be any connection between our intentions and our actions. For McKenna, the content of a person's actions is irrelevant, because autonomy is determined with reference only to the mental command an agent gives themselves to act: '[autonomy's] scope comes to an end at the moment in which an agent performs simple mental actions, such as deciding or choosing' (McKenna 2005: 229, fn 14). On McKenna's account, if a person's command to themselves aligns with a self-endorsed principle, their following action would be considered autonomous, regardless of any features of the act itself. Killmister (2013a: 523) objects that internalists about autonomy like McKenna are therefore committed to the 'counter-intuitive notion that the success conditions for autonomous action bear no

---

[13] See Mele (1995/2001: Chapter Nine) where he argues that what he calls 'psychological autonomy' (roughly what I call 'reflective autonomy') requires that agents have no compelled values or other pro-attitudes.

relation to the action performed.' I agree that this is both counter-intuitive and an inadequate account of autonomy. Because, as Killmister (2013a: 520) emphasises, 'the reason why a person acts is *in order to* fulfil their intentions,' to not require some sort of relationship between one's intentions and the actual action seems to neglect something critical about autonomy.

Killmister's own position is that to count as having acted on the basis of an intention, there needs to be 'some' connection between the action and the intention (2013a: 520). As such, she requires there to be *some degree* of coherence between an intention and an action if the action itself is to be properly considered self-governed.[21] The aim of Killmister's article is to draw attention to the ability of false beliefs to disrupt this coherence. A person's intention and action can become separated if they are misinformed, she says, with their subsequent actions leading them away from the realisation of their intention, i.e. 'the reason [they had] for acting in the first place' (2013a: 521). Killmister concludes that whilst autonomy requires that the actions we perform follow from motivations that are in some sense our own, autonomy *also* requires that the actions we perform are those we intended to perform.

Using a weak reading of Killmister's account, one might then try to argue that there actually *was* sufficient coherence between Tennille's and Tal's motivations and their actions, and as such, that their actions ought to count as autonomous.[22] It might be argued that the actions performed by Tennille [and by Tal] sufficiently tracked their intentions and self-commands to act, and as such, that their actions were autonomous. For instance, Tennille did in fact perform actions she instructed herself to complete like 'open the bottle' and 'consume the medication'. So — the argument would run — there actually was the necessary correspondence between Tennille's intention and her action that would make it correct to say that she acted autonomously.

However, in reply to this kind of objection, Killmister (2013a: 523-524) suggests that this would only be plausible if we do not allow any act individuation and take all actions

---

[21] Note that Killmister doesn't specify just how much coherence there needs to be.

[22] I borrow from Killmister's (2013a:523) response to what she calls a 'weak externalist' reading of McKenna's argument here, and follow her argument's structure.

to collapse into one single event. For example, Killmister would say that we would need to see no separation between, e.g.:

(a) Tennille autonomously attempted to quit smoking;

(b) Tennille autonomously consumed the medication given to her by her doctor;

(c) Tennille autonomously consumed the nicotine craving enhancer.

If one was to claim that Tennille acted autonomously throughout the episode 'in general', such a broad-brush stroke would not allow for the possibility that there may have been aspects of the episode that were performed autonomously, and others that were not. As McKenna does not allow for act individuation, he would reject the suggestion that Tennille did not autonomously perform at least one act: consuming the nicotine craving-enhancer. Killmister would claim that the reason why taking the nicotine craving enhancer was not autonomous, while other actions (i.e. trying to quit smoking) likely were, rests on the connection between Tennille's intentions and actions outlined above. For an act to be autonomous, Killmister stipulates that it needs to be connected in the right way with someone's intention to act. She suggests that it needs to be the *same act* that the agent has reflected on and resolved to perform. On her account, we can therefore say that Tennille reflected on and resolved to perform actions (a) and (b), but we cannot conclude the same for (c).

Whilst I support Killmister's use of act individuation in this example and agree with her conclusions about Tennille's and Tal's autonomy, I disagree with Killmister's reasoning as to why Tennille's and Tal's false beliefs compromised their autonomy in these examples. I explain why next.

## 2.2. Which False Beliefs Undermine Autonomy?

Having defended the view that false beliefs are among those things that *can* undermine a person's autonomy, I now turn to the questions: do *all* false beliefs compromise autonomy? If not, *which* false beliefs compromise autonomy? Or in other words: under what conditions do false beliefs interfere with personal autonomy?

*2.2.1. Killmister: False Beliefs That Break Relations Between Intentions and Actions*

As discussed, for Killmister (2013a), the sorts of false beliefs that compromise autonomy are those that break the relationship between a person's intention and their action. This is because she takes autonomous action to minimally require acting on the basis of an intention, and this in turn requires there to be 'some degree' of coherence between one's intention and one's action. She argues that if the agent is mistaken about the object of her reflection – i.e. the action she is contemplating – she will fail to adequately comprehend the action, and her intention and action will become separated, with the act that she ends up performing failing to be directed towards her intention. Killmister uses the following example to illustrate her argument:

> **Cake/Dog food:**
>
> 'Let's say I form the intention to eat a piece of cake. To fulfil this intention, I go to what I take to be the fridge and retrieve what I take to be a delicious cream-filled Swiss roll. In fact, what I have done is gone to the cupboard and retrieved a roll of dog food, which I am eagerly bearing towards my mouth'…[This action] '*cannot* satisfy my intention, and so the action is divorced from my self-governance. The order I give myself is to satisfy my desire for cake, yet that is not possible given the action I perform. As such, the action was not under the direction of my will' (2013: 521).

Of course, the false belief in question needs to be *relevant,* and Killmister is quick to point this out. Thus if the agent had the self-governed intention to simply 'eat cake', and thought that what she was about to eat was a cream-filled Swiss roll when it was actually a chocolate roulade, then I suspect Killmister would conclude that her autonomy wasn't diminished.[14] This, as Killmister (2013a: 528) explains with reference to a similar scenario, is because her reasons for her action under the false-belief scenario

---

[14] Killmister raises a similar example and concludes there that her autonomy 'most probably would not be undermined' (2013a: 528).

would have broadly matched those she would have for her action under the scenario in which she was not mistaken about the precise type of cake she was about to eat.

However, there is something about Killmister's account that doesn't seem quite right to me. This, I think, is because Killmister rejects the relevance of the *etiology* of the false belief/s in question, and whether such facts influence determinations of autonomy. Killmister makes it clear that her theory 'does not make space for the source of the misinformation to affect the degree to which the agent's autonomy is reduced' (2013: 517, fn 4). However, not only does this position have an undesirable consequence — namely, an inability to be autonomously irrational or to autonomously fail, as it seems that failures and instances of irrationality will often be the result of some false belief/s disrupting the coherence between our intentions and actions — I also don't think it accords with our intuitions. Consider the following two scenarios:

---

**Peter**

Peter's higher-level intention is to pocket the black billiard ball to win the game. His sub-intention is to hit the black ball at an ~30-degree cut angle at a speed of 6-8km/h, as this is what he believes is needed for the ball to roll into the pocket. However, he has miscalculated what type of shot (action) is actually needed to sink the black ball; he in fact needs to hit the ball at a ~25-degree cut angle at a speed of 10-12km/h in order for the ball to go into the pocket. He therefore has false beliefs about the characteristics of the action that is needed to fulfil his (higher-level) intention. He perfectly carries out the action he planned, but the ball does not roll into the pocket.

---

> **Tennille 4**
>
> Tennille's doctor provides her with an information sheet about the known side-effects associated with different smoking cessation options. The neuropsychiatric risks associated with P are highlighted in bold and her doctor encourages her to consider those risks in particular. She has ample time to carefully read through this information, but Tennille only quickly scans it. Tennille overlooks key pieces of information and so comes to hold a false belief about the side-effects of P. She chooses P. The medication's box also clearly highlights the risks of psychosis. She doesn't read this either. She takes a course of P, and at odds with her deeply held values, ends up back in a bout of psychosis, despite successfully giving up smoking.

Even though Peter and Tennille (4) hold false and inaccurate beliefs about the actions needed to fulfil their intentions, the etiologies of these false and inaccurate beliefs are tied to Peter's and Tennille's own agency: Peter and Tennille both maintained adequate control, and the right kinds of control, throughout their belief forming processes. We might say they had a *fair opportunity to form accurate beliefs and to avoid forming inaccurate beliefs*, and it was with full autonomous agency (Peter's miscalculations, Tennille's inattentiveness) that they did not form accurate beliefs. They failed to do those things that were reasonably within their control to avoid forming inaccurate and false beliefs, and to form accurate or true beliefs. Thus, despite both Peter's and Tennille's intentions and actions coming to be being misaligned in the way that Killmister claims necessarily diminishes autonomy, these breaks were due to Peter's and Tennille's self-governed (autonomous) actions upstream of the point of action. When it's the case that someone can be said to have autonomously formed a false or inaccurate belief, or to have autonomously not formed an accurate or true belief, it's not clear that subsequent downstream reflections and actions tied to that mental state should count as non-autonomous merely in virtue of that false belief.

I anticipate that Killmister would resist this suggestion. She specifically states that on her theory 'the *culpably misinformed* individual is no more autonomous *at the point of action* than the manipulated individual, since *all that matters* is the way in which a false belief affects the decision-making process' (2013a: 517, fn 4 - my emphasis). However,

there are good theoretical and practical reasons why we should resist this conclusion. Firstly, my proposal for thinking about the difference the etiology of false beliefs makes to a person's downstream autonomy is simply the inverse of Mele's suggestion that when the autonomy-status of upstream choices is compromised, the autonomy-status of downstream deliberations and actions will be tainted; on my suggestion, if the autonomy status of an upstream event (such as the formation of a belief) is *not* compromised, then neither are the downstream deliberations and actions linked to that false or inaccurate belief, at least, as long as there are no other autonomy-compromising forces at play. Our autonomy has the ability to 'carry through' in the chain of events, so to speak.

Killmister anticipates and replies to this kind of idea by suggesting that 'the source of this intuition may have less to do with autonomy, and more to do with *moral responsibility*' (2013a: 517, my emphasis), stating that a culpably misinformed individual can retain moral responsibility for her non-autonomous actions. Whilst Killmister (ibid: fn 4) notes that she does not have space to defend the position that autonomy and moral responsibility can come apart in this way, it seems she accepts that they can. There is at least one way we can interpret an existing legal practice to support Killmister's view: the 'principle of prior fault' found in criminal law rules in several jurisdictions holds that, for example, voluntarily intoxicated persons are liable (i.e. morally responsible) for crimes they commit, even if at the time they commit the crime (i.e. what Killmister calls the 'point of action'), they are inebriated and their actions are non-voluntary (i.e. they literally lack the relevant criminal capacity at the time the crime is committed). Under this principle, conditions that would normally shield someone from some form of liability (e.g. not having voluntary control over one's actions) fail to do so thanks to the fact that they are to blame for those conditions obtaining (Yaffe 2011). So on one reading of the prior fault principle that accords with Killmister's view, people whose actions are non-autonomous at the point of (criminal) action can still be held morally responsible for a crime they non-autonomously committed.

However, I suggest that there is another way the prior fault principle can be interpreted which makes use of an idea that is the opposite of the one Killmister used earlier: in some cases, instead of act individuation being appropriate, prior fault may licence a type

of *act aggregation*.[15] Rather than carving up the event by deconstructing it (as under act individuation), a person's prior fault might provide good reasons for carving up an event more coarsely such that it integrates and unifies the person's governance state over the event more broadly. It is because someone was culpable for their upstream decision to become intoxicated, and they had the capacity at that time to anticipate the potential consequences of this choice, that we are justified in treating the episode more coarsely and holding them responsible in some way for the general event. They are morally responsible *because* they were — in some meaningful sense — autonomous over the event generally.

The following scenario illustrates this suggestion:

---

**Judy 1**

It's Monday and Judy's father's birthday is on Thursday. She has promised him that he will receive a card from her on or before his birthday. Judy lives in Newcastle and her father lives in Eastbourne. She knows that Royal Mail aim to deliver first class mail between Newcastle and Eastbourne the next working day. She puts the correct stamp on the card and posts it. But Judy holds the false belief that every red postbox in Newcastle is emptied daily. This particular postbox is only emptied on Fridays, and this is clearly stated on the outside of the postbox. However, Judy did not check the sign due to her assumption (a false belief) as well as the fact that she was in a hurry. Her father's card arrives two days late; he is hurt and disappointed in her for breaking her promise.

---

Judy formed her false belief about how frequently postboxes are emptied autonomously. Nobody, and importantly not Royal Mail, communicated misinformation to her. It's just that because it appeared that *most* of the postboxes are emptied daily, Judy overgeneralised and formed the (false) belief that they are *all* emptied daily. This was without

---

[15] There is also a slightly weaker interpretation of the prior fault principle which says that someone can be treated *as if* they were autonomous even if they actually were not. Some scholars take this to be equivalent to the principle licensing the use of a 'legal fiction'; e.g. in some jurisdictions the Intoxication Recklessness Principle justifies the replacing of one mental state (negligence) for another (recklessness) in cases of voluntary intoxication (see Yaffe 2011).

warrant. Moreover, Judy had a good opportunity to correct her false belief, because Royal Mail had clearly stated on the outside of the postbox when the mail would next be collected from that particular postbox. Judy's intentions (ensuring her father's birthday card is delivered on time/posting the card into a postbox that will be emptied that day) and action (posting the card in such a way that it will be late/posting the card into a postbox that would not be emptied that day) came to be misaligned in the way that Killmister claims necessarily diminishes autonomy, and did so by virtue of Judy's false belief. However, it would seem strange to say that due to her false belief Judy was in some way 'not autonomous' when she posted the letter, but that her father was still justified in holding her morally responsible for her failure to ensure his card arrives on time. Her father's reactive attitude was justified *because* Judy was autonomous over the event; she never lost the degree or kind of agency required for autonomy. Killmister might try to individuate the actions so that there is one that Judy doesn't autonomously carry out (e.g. perhaps: 'Judy posted the card so that it would arrive late'). But I want to say that due to her prior fault (i.e. the false belief being the product of her own autonomous agency), her autonomy status ought to carry through, as her prior fault for her false belief provides a reason to aggregate, rather than individuate, actions.

To provide a contrast to this scenario, let's now pretend instead that before Judy posted the card someone had mischievously swapped the signs around on the postboxes in Newcastle.

---

**Judy 2**

Before Judy posts the card, she checks the sign on the postbox — it informs her that it will be emptied this evening at 6:30pm. But because it *in fact* won't be emptied until Friday (someone has swapped this postbox's sign with that of another postbox), she holds a false belief about when the postbox will be emptied when she posts the card, and therefore about the chance her card will reach her grandpa on time by posting it in this particular postbox.

---

The key difference between the two examples is that Judy (2) did not autonomously form the false beliefs she held: she was misinformed by the signs being swapped, and it

would have been unreasonable to expect her to form any other beliefs under these circumstances; there was no evidence to suggest that the signs had been switched. *Because* the false beliefs she held were not formed autonomously, these non-autonomous false beliefs tainted the autonomy status of related downstream actions. The presence of a non-autonomously formed false belief should, I submit, license act individuation. This would result in the conclusion that whilst Judy (a) autonomously attempted to post her grandpa's birthday card to him to arrive on time and (b) autonomously put the card into a postbox, she *did not* (c) autonomously post the card in such a way that it would not arrive on time. In this rendition of the episode, the switched signs and the misinformation they conveyed interfered with and eroded Judy's autonomy. As such, it would have been inappropriate for her father to hold her morally responsible for the card's lateness (and for broken promises) under these circumstances.

To sum up: whilst Tennille (4) and Judy (1) hold false beliefs that cause them to make mistakes that lead them further away from achieving their ends and values, they did not lose the requisite forms of agential control for autonomous conduct throughout the episode. Their choices and actions should still be considered *theirs,* and their autonomy should not be considered to have been undermined. They might be irrational and/or unsuccessful, but they were *not* non-autonomous or heterotonomous. This is in stark contrast to Judy (2). Here, Judy held false beliefs that caused her to make mistakes that led her further away from achieving her ends and values. But she formed these false beliefs non-autonomously; another person's actions switched the signs around, and by doing so they misinformed Judy. Judy could not have legitimately been expected to resist forming such a false belief, because there was no evidence to indicate that the signs had been switched. She therefore lost a requisite form of control over her subsequent conduct for it to rightly count as autonomous. Certain actions should not be considered to be *her* actions — i.e. the card being posted in a way that meant it would be late.

Those false or inaccurate beliefs that a person has, and that they can legitimately be expected to have resisted forming or to have corrected, should not, I contend, be considered to be the sorts of false beliefs that undermine their autonomy. Therefore, *contra* Killmister, false beliefs that break the relationship between intentions and actions are neither necessary nor sufficient for non-autonomy.

### 2.2.2. Taylor's and Mele's Answer: False Beliefs That Are the Product of Deception

For both Mele (1995) and Taylor (2009), the variety of false beliefs that compromise autonomy are those that are a product of *deception*. Gerald Dworkin (1988: 14) also asserts that it is deception that differentiates autonomy from liberty, with deception undermining autonomy, but not liberty. Following a definition laid out by Mahon (2015), I understand deception as:

> *Intentionally* causing another person to acquire a false belief, or to continue to have a false belief, or to cease to have a true belief, or to be prevented from acquiring a true belief, or to *intentionally* allow another person to acquire a false belief, or to continue to have a false belief, or to cease to have a true belief, or to be prevented from acquiring a true belief.[23]

If we look at all the examples presented so far in which we concluded that Tennille's autonomy had been undermined, Tennille had been deceived, whether by the tobacco industry, her doctor, or indirectly by the pharmaceutical company. So too for Judy (2): whoever switched the signs on the post boxes deceived Judy. Deception therefore looks like a promising candidate for distinguishing between false beliefs that are autonomy undermining and those that are not.

Mele (1995/2001: 178-182) refers to deception as a type of 'covert and non-constraining control'. In discussing how deception affects autonomy, he asks us to imagine the case of King George, whose only access to the state of his kingdom is via his advisors.[17]

---

[23] Thus, the important thing is that I accept that deception must be *intentional.* See arguments made by Barnes (1997), Carson (2010), and Saul (2012) (among others).

[17] Mele doesn't make it clear why King George's only access to the state of his kingdom is through his advisors; for now, let's just assume that George is bedridden, and his advisors are the only company he has.

> **King George**
>
> King George's fundamental value is to do what's best for his kingdom and its population. His advisors provide him with information, and George makes various governance decisions — e.g. laws, regulations, tax rules, and budget allocations — using the information from his advisors as an input into his deliberations. But the advisors have their own agenda: to become wealthy at the expense of ordinary citizens. Because they know the King's preferences, they know how to 'game' his decisions by systematically providing him with misinformation so that he will be led to make decisions that further their own ends rather than George's. His advisors deceive him, and their acts of deception undermine George's autonomy.

Mele concludes that:

> George plainly is not ruling autonomously. Hence, he is not autonomous in the sphere of his life that is most important to him. The staff are in control of his activities in that sphere. They control George by controlling what information he receives. One can say, if one likes, that the staff *compel* George to have certain pro-attitudes: for example, a desire to enact a certain law. ... George still intrinsically values the welfare of his kingdom as a whole, but he is systematically *deceived* about the means to this end. (1995/2001: 180, my emphasis)

Taylor (2009) would say of this case that King George's own will has been subjugated to the wills (or the collective will) of his advisors (his deceivers); they control George in a way that is incompatible with his self-government, and this explains why deception is inimical to autonomy. The nature of the deceiver's controlling influence means that the deceived agent is no longer able to direct their life in the light of their own decisions about what is valuable. Here is Taylor's summary explanation:

> If the information upon which the agent bases her decision has been affected by another agent with the end of leading her to make a particular decision... and if she is not aware of the way in which this information has been affected, *then she* did not make the decision that (her deceiver) intended her to make. (2009: 7).

However, against Taylor's conclusion, it also seems possible for false beliefs to undermine a person's autonomy without those false beliefs being *deliberately* or *intentionally*

brought about by another agent (i.e. another person's 'will' is not the driving force). We can modify earlier examples in accordance with this point to demonstrate.

---

**Tennille 5** (modified Tennille 2)

Tennille's doctor provides her with information about the range of options available for smoking cessation (i.e. options N/O/P), and the strategy the current evidence base suggests is most efficacious (e.g. option P). However, distracted by thoughts of his long list of patients, he mistakenly gives Tennille the wrong bottle. Whilst he tells her that the medication in the bottle labelled P is P, it is actually filled with a substance (M), which has the capacity to enhance, rather than reduce, her nicotine cravings.

---

**Tennille 6** (modified Tennille 3)

Tennille's doctor misinforms her about P's side-effects. Unbeknownst to him, initial clinical trials actually found P's side-effects to include significant neuropsychiatric risks. However, the pharmaceutical company that developed P was reckless with the results of these studies; the trial data was lost and forgotten about and therefore not communicated. This misinformation led Tennille to select an option that, unbeknownst to her, harmed one of her most cherished values (i.e. her mental health), even though it helped her to achieve another (quitting smoking).

---

Neither of these examples feature agents that *intentionally* set out to mislead another person for their own ends; at most they feature agents who were negligent (example 5) or reckless (example 6). So even though they have not been deceived — as deception must be intentional — I think intuitively we would still conclude in these examples that Tennille's autonomy had been undermined (and I will explain why momentarily). Therefore, an agent's actions or omissions *need not be deliberate* to be a potentially autonomy-undermining force.

Perhaps Taylor might say that even though deception is not necessary it is still sufficient; if a false belief was brought about by deception then it has the capacity to be autonomy undermining. Against this, consider one last example:

---

**Tennille 7**

On the way to the doctor, Tennille speaks to a stranger at the bus stop about her goal to quit smoking. The stranger says that she is a pharmacist (even though she is not) and tells Tennille about P. Tennille asks her whether P has any known side effects and the stranger tells her no, even though she is well aware that there are a range of side effects, including the possibility of psychosis. The stranger happens to have some P and is willing to sell it to Tennille on the cheap. 'Brilliant,' Tennille thinks, and decides to go home instead of going to the doctor. She takes P and develops psychosis.

---

In this example Tennille was deceived by the stranger. But Tennille had a range of opportunities open to her — opportunities that were not onerous, costly, or beyond her capabilities — to check that the assertions of the stranger were correct before she formed beliefs about P and acted on those beliefs: e.g., she could have continued to her appointment with her doctor. Given that she only met the stranger momentarily, and at a bus stop, it would have been unreasonable for Tennille to have taken the stranger to be an authoritative source about P, or to have been trustworthy, as she did not have adequate access to higher-order evidence to make such evaluations. Moreover, other people would have been *entitled* to expect that Tennille *should have known* that the stranger's assertions might have been false. Goldberg (2017) defends the idea that we are entitled to expect that others have certain knowledge, and/or epistemic standards and practices, by virtue of the general requirements of morality. In this case, others' legitimate expectations of what Tennille should have known stem from at least two particular moral considerations. The first is that others sometimes need to make assessments about the culpability of Tennille's ignorance to understand the degree to which Tennille's downstream actions are the result of her autonomous agency. This is important for ascriptions of moral responsibility, and determinations of what Tennille is entitled to. Second, as an epistemic and moral agent herself, Tennille also, in turn, has epistemic responsibilities to others in terms of not further propagating misinformation.

Since Tennille, in these circumstances, 'should have known' that P carried a risk of psychosis, the autonomy status of her downstream deliberations, intentions, and actions cannot be affected by this false belief (similar to Judy (1)). In this instance, we should not think that Tennille's autonomy was undermined by the stranger's misinformation because the circumstances were such that Tennille herself was responsible for navigating the information and misinformation provided to her, and sifting truth from falsity. Tennille retained her autonomy in forming these beliefs, and thus over her subsequent actions.

Therefore, the distinction between the intentional and non-intentional inducing of false beliefs in another through deception does not — contra Mele and Taylor — always help us identify when false or inaccurate beliefs compromise autonomy.

## 2.3. An Alternative Account: Voluntary and Non-Voluntary False Beliefs

One might now be tempted to reassess the earlier examples and conclude that Tennille's reflective and/or practical autonomy was not actually diminished in at least some of those cases too; it might be thought that in those examples, Tennille's false beliefs were, to a sufficient degree, a product of her own agency. Perhaps one might think that in the first example Tennille could have taken a more critical stance towards the material propagated by the tobacco industry; had she simply gone to her library and read the latest medical journals, Tennille would have discovered that she had been misled. Hence, some might be inclined to say that Tennille's preference for smoking *should* be considered to have been formed autonomously in this example, as she was free to conduct these investigations, and it was Tennille who chose not to; she formed a false belief about tobacco's health risks as a consequence of her autonomous choice not to seek out more information. Likewise, it might be thought that because Tennille could have uncovered her doctor's deception in the second example simply by booking an appointment with another doctor for a second opinion, we should not view her autonomy as having been compromised in this example either.

I think such conclusions would be mistaken. Although Tennille may have been *negatively free* to form accurate beliefs and to avoid forming inaccurate and false beliefs, I

submit that Tennille's false beliefs in examples 1-3, and 5/6, were not formed *voluntarily*.[18] It is because Tennille formed false beliefs non-voluntarily in these examples that Tennille's autonomy was compromised. By contrast, it was because her false beliefs in examples 4 and 7 were formed voluntarily that her autonomy was thereby unaffected. I suggest that only those false beliefs that are formed non-voluntarily can compromise a person's autonomy. But before I present my arguments for this claim, it will be helpful to provide a brief overview of the theory that defends the very possibility that at least some of our beliefs can be voluntarily formed.

### 2.3.1. Doxastic Voluntarism

Doxastic voluntarism holds that people can have voluntary control over their beliefs in much the same way that they can have voluntary control over their choices and actions. Under doxastic voluntarism, beliefs are the sorts of things that can at least sometimes be formed under the control of a person's will. Although the position that people have *direct* voluntary control over their beliefs is widely rejected, there is a very strong contemporary consensus among philosophers that *indirect* doxastic voluntarism is true (Vitz 2018). Indirect doxastic voluntarism claims that people have indirect voluntary control over at least some of their beliefs through, for example, carrying out research and evaluating evidence and testimony. I assume that this theory is correct, and for the rest of this chapter refer only to *indirect* doxastic voluntarism.

But whilst indirect doxastic voluntarism establishes that we have the *capacity* to form at least some sub-set of our beliefs voluntarily, the crucial question is: under what conditions do we *in fact* form beliefs voluntarily within these domains, and what conditions have the potential to erode the voluntariness of these beliefs? So far, the literature on indirect doxastic voluntarism has had little to say about what these conditions are. Moreover, those conditions that have been put forward have often been cashed out in terms of having 'control over' X, Y, Z, for example, which of course raises the question of what, exactly, having 'control over' entails. We need to avoid this kind of vacuity and

---

[18] See Olsaretti (2004: 137-141) for further discussion of the relation between freedom and voluntariness. Olsaretti defines freedom as pure negative liberty, and I follow her for my purposes here.

circularity if we want an account that is meaningful and useful. That will be my aim here.

### 2.3.2. Can We Apply Pre-Existing Criteria for (Non)Voluntariness to Beliefs?

Because beliefs and choices/actions are deeply analogous — i.e. beliefs and choices/actions are all generally grounded by reasons; all can appropriately be described as objects of deliberation and decision; all are open to moral evaluation; and beliefs are action-like when understood as being assented to or accepted (Audi 2001) — one might wonder whether we can simply apply pre-existing criteria for (non)voluntariness to the case of belief formation in order to understand when beliefs formed in different contexts are (non-)voluntary. For instance, on Serena Olsaretti's (1998, 2004, 2008, 2018) account of voluntariness, a choice or action is voluntary if and only if it is not forced, and it is forced if and only if it is made only, or primarily, because the alternative to it is unacceptable, where the standard for the (un)acceptability of options is an objective standard of well-being. In short: a person chooses or acts non-voluntarily if they choose or act as they do because all other alternatives are unacceptable. Does it make sense then, to apply Olsaretti's voluntariness criteria to beliefs, such that we would consider a belief to have formed non-voluntarily if a person forms it only or primarily because all other alternatives to forming that belief are unacceptable?

I suggest it does not for two key reasons. Firstly, although there are important similarities between beliefs and choices/actions, there are also significant differences that make pre-existing accounts of voluntariness (like Olsaretti's) a poor fit for evaluating the voluntariness of beliefs. For one thing, when we're forming our empirical beliefs, we're making a decision about what we should believe about how the world *is*, descriptively speaking. On the other hand, when we're at a decision-point that requires us to make decisions about an action to carry out, we're generally making decisions that relate to how we would *prefer the world to be*, and we act with the aim of *bringing that state of the world into existence*.[24] Therefore, in some cases there might only be one real option in forming

---

[24] This sentence, and the one previous to it, has two sources of inspiration. Firstly, Anscombe's (1963) distinction between, on the one hand, our (epistemic) beliefs, and on the other, our desires or moral

a belief about the empirical world, because the evidence might be such that there is no other answer to a particular descriptive question that is more correct or accurate (e.g. 'Is there a red bus currently in this room?', 'Has the Earth's climate system warmed since 1950?'). If we applied Olsaretti's criteria to those situations in which we answered these kinds of questions, they would yield a strange conclusion: that the formation of the belief was non-voluntary, because all other alternatives to forming it would have been unacceptable given that the evidence suggested that the proposition was beyond reasonable doubt. This can't be right. For some empirical questions there will only be one correct or most accurate answer, and it can't be that if we accept that answer we have formed that belief non-voluntarily.

Second, Colburn (2008), building on Olsaretti's account, draws attention to the fact that whether a choice or act is voluntary or not depends upon the person's motivation for their action, and hence upon their *beliefs* about their options. He goes on to say that if those beliefs are false or inaccurate, then an agent can in fact have several acceptable alternatives available to her, but the action she ultimately chooses might nevertheless be non-voluntary due to her ill-informedness about the existence of those acceptable alternatives. He concludes: 'Hence, if we want to promote an agent's opportunities to make voluntary choices, we cannot simply attend to the existence and quality of her options, but must also ensure that she is well-informed about them' (2008: 111).[19] I interpret Colburn's stipulation here as a requirement that if we are to respect persons' autonomous agency, we must provide them with a fair opportunity to form accurate empirical beliefs relevant to their actions.[20]

---

beliefs. In the former case we want the 'direction of fit' to run from the world to the mind. We're aiming for our beliefs to fit the world—to be accurate—and we will change our beliefs if they do not. In the latter case, we want the direction of fit to run the other way, i.e. from the mind to the world. We want the world to fit our desires or our sense of morality and justice. When it doesn't, we want the world to change and might work to bring that about. The second is Arendt's (1967: 78) assertion in her essay *Truth and Politics* that 'all factual truth, of course, concerns the past' and that only the future is open to action.

[19] Colburn gestures towards similar requirements in *Autonomy and Liberalism* (2010).

[20] In the case of evaluative, rather than factual, beliefs, this requirement might translate to an agent needing to be provided with fair opportunity to adopt one or more of a variety of evaluative beliefs. This might then mean that they need to be exposed to some variety of ideologies or ways of life to have formed

So, if we were to apply Olsaretti's voluntariness criteria to beliefs, we would then need to incorporate Colburn's amendments in that context too. We would end up having to say something like whether or not our beliefs are voluntary depends on our motivations for forming those beliefs, and hence upon our beliefs about forming our beliefs; we would need to be well-informed (i.e. have accurate beliefs) about forming our beliefs. We would then need to posit substantive epistemic criteria at this point (i.e. what counts as 'well-informed'), otherwise we would end up in an infinite regress. But if we inevitably need to outline substantive epistemic conditions for the voluntariness beliefs at some stage, then it seems better to outline what these criteria are in the first instance. I see no advantage in having a two-step account that uses Olsaretti's account at the first step, and then undertakes an epistemic evaluation against some substantive criteria at a subsequent step or at a meta level. It seems better to develop separate epistemic criteria to determine what ensuring that someone is 'well-informed' amounts to (i.e. what is required for a person's beliefs to count as voluntary) which could be used by those using pre-existing accounts of voluntariness, but as a standalone account it could also be used for other purposes too.

Therefore, the task of understanding the degree to which people form beliefs (non)voluntarily requires distinct substantive epistemic criteria.


### 2.3.3. Epistemic Criteria for the (Non)Voluntariness of Beliefs

In this section I develop a set of jointly sufficient criteria for assessing the voluntariness of a person's factual beliefs about the world in a given circumstance. The application of these criteria, in turn, will aid evaluations of whether the transmission of misinformation in a given circumstance has the potential to undermine autonomy.

Legal philosopher Seana Shiffrin gives us a clue as to what these conditions might be in her book *Speech Matters* (2014). Here she says we must guard against 'distortive influences' that frustrate a person's reflection as they critically appraise another agent's

---

these evaluative beliefs voluntarily, and hence for further choices and actions to count as autonomous. I put evaluative beliefs and requirements for the voluntariness of these kinds of beliefs to one side and focus only on factual empirical beliefs here.

speech. So, my provisional high-level answer to the question of when we should judge communicative practices to compromise the voluntariness of a person's beliefs, and by extension, their autonomy, is: when these practices inhibit a person in exercising their reflective capacities such that they can no longer be said to have fair opportunities to form accurate beliefs and/or avoid forming inaccurate beliefs.

To give a more detailed specification of the conditions that constitute a 'fair opportunity', we can refer back to the examples presented earlier in which our general intuitive appraisal suggested that the individual wasn't adequately placed to reflectively criticise the influence that was exerted on her from outside. Which specific features accounted for our judgements that Tennille, Judy, and Peter in those examples were all unfairly compromised in their abilities to reflectively navigate the influences to which they were subjected?

It seems that *insulation* was one important feature. Insulation can occur if important evidence has been omitted from, or is being suppressed within, a person's socio-epistemic environment. Important evidence includes things like experts' testimonies about P, experts' reasons for P and for not-P, as well as higher-order evidence about these putative experts and the processes they used to arrive at their factual claims about P.[21] If adequate evidence is not available within a person's socio-epistemic environment, their ability to put their reflective capacities to work will be impaired, and it will be more difficult for them to evaluate other's speech well. For instance, in Tennille (2), critical higher-order evidence was omitted from the conversation she had with her doctor: the doctor did not reveal that he owned the patent for 'M'. Had this evidence been available to Tennille, she would have had a resource that she needed to better reflect on the trustworthiness of her doctor's assertions and advice. Similarly, in Tennille (3) and Ten-

---

[21] High-order evidence includes (but is not limited to): the degree and/or spread of expert (dis)agreement; who agrees/disagrees, and why; indicators of contenders' competence, vis-à-vis the specific hypothesis; relevant potential interests and biases of contenders; indicators of honesty, and track records of integrity; how and why any consensus formed; the degree of meta-expert agreement, and other indicators of meta-expert and interactive expert appraisal (Goldman 2001; Coady 2006; Anderson 2011; Coady and Corry 2013, pp. 22–34; Lane 2014).

nille (6), critical information about the drug's side-effects was omitted because the pharmaceutical company had either deliberately withheld this evidence, or because it had recklessly or negligently failed to gather it.

It might be objected that there were still ways that Tennille could have discovered at least some of this information for herself. Perhaps with enough effort it would have been possible for Tennille to track down the patent records. Whilst this might have been in some sense *possible* (i.e. Tennille was *negatively free* to do so) it seems unreasonable to *expect* Tennille to go to such a length to gather this kind of evidence herself. Rosen (2003, 2008) makes a similar point in his work on culpable ignorance: just because it is possible *in principle* for an agent to obtain certain information, it can still be unreasonable to expect them to have done so. In this case, this is because Tennille's doctor *qua* doctor (and the health system more broadly) had professional obligations to Tennille to provide her with information and to avoid misinforming her, as well as to make relevant higher-order evidence available to her. Her doctor's role within the public knowledge system *entitles* Tennille to expect this of him, and to trust that he is providing her with the relevant epistemic goods and meeting institutional epistemic standards — what Goldberg (2017) calls 'role specific epistemic obligations' — as well as all other relevant professional and basic moral obligations he has towards Tennille. Furthermore, Tennille's own position vis-a-vis others within the public knowledge system — namely, her reasonable epistemic dependence on her doctor — means that we are *not entitled* to expect Tennille to do more than defer to his testimony, at least in this case. It would be an illegitimate expectation of Tennille if we did expect her to do such independent research (Goldberg 2017).

*Contortion* was another important feature. By contortion I mean a bending or twisting of evidence, including other's speech (e.g. experts' testimony), through practices that misrepresent and obfuscate and render the epistemic landscape opaque. We saw this in many of the examples. Tennille's socio-epistemic environment in the first example, for instance, would have featured widely disseminated misinformation and misleading statements about the health risks posed by cigarettes. Not only did tobacco companies explicitly deny the causal link between smoking and cancer and its addictiveness in their

public communications (Barnes et al. 1995), they also created various positive associations between physicians and cigarettes, as well as between smoking and good (rather than poor) health (Gardner and Brandt 2006, Proctor 2011). Moreover, analyses of tobacco industry documents have revealed that tobacco companies were aware of the inherent deceptiveness of the statements and associations they made, and that such marketing tactics were intended to promote misperceptions (Pollay & Dewhirst 2002). Similarly, Tennille's doctor in the second example misrepresented which strategies the evidence base suggested were effective; in the examples featuring Tal and Tennille (5), pharmacists' labels had been placed on the wrong bottles; and in Judy (2), the official signs on the postboxes had been switched. All of these acts of commission or omission inhibited agents' capacities to reflect in ways conducive to the formation of accurate empirical beliefs.

It might be pointed out that in Tennille (7), the stranger also engages in contorting the evidence by lying to Tennille. However, in the context of this example, it seems reasonable that Tennille is expected to undertake further evidence-gathering activities to verify the stranger's assertions. Whilst it might still be morally wrong for the stranger to lie to Tennille (as Shiffrin 2014 argues), in this case this wrong doesn't extinguish or override the legitimate expectations we have of Tennille in this context to resist forming inaccurate empirical beliefs, and to improve her own epistemic situation. Moreover, Tennille had a good opportunity to do so by keeping her appointment with her doctor, rather than going home.

Finally, that a person's epistemic environment was *overwhelmed* in some way also seemed to contribute to the question of whether they formed beliefs voluntarily. The sheer volume and frequency of messages can sometimes compromise a person's ability to exercise their reflective capacities and update their beliefs and credences appropriately. There is some evidence, for example, that we tend to believe that those propositions that are more familiar to us are more likely to be true (Schwarz et al. 2007). In example 1, Tennille's socio-epistemic environment was likely inundated with frequent misleading messages from the tobacco industry; the 1950s-1970s was the 'golden age' of tobacco advertising before states started to regulate the industry's speech in any serious way. In 1964, the same year that the U.S. Surgeon General published a report directly linking smoking

to cancer, tobacco companies were spending $250 million per year on advertising (Whiteside 1971). It's also possible for information to be presented in a way that triggers the activation of motivated reasoning and cognitive shortcuts, which although can sometimes be an adaptive feature of our cognition, can in other circumstances be maladaptive and lead people to under- or over-estimate pertinent facts. Such mental phenomena are vulnerable to exploitation.

The following conditions are therefore, I suggest, jointly sufficient criteria for a person having a 'fair opportunity' to form accurate empirical beliefs, and hence for their empirical beliefs to be considered to have been formed voluntarily:

---

**Jointly sufficient conditions for the formation of voluntary factual beliefs:**

(1) In light of what is known at the time, or what one legitimately expects to be known, adequate evidence is available within the public knowledge system, and a person has effective access to this evidence.

(2) The body of putative evidence a person has access to is generally presented with integrity; there are no significant misrepresentations or misleading uses of evidence within important parts of the public knowledge system.

(3) A person is able to deploy their reflective capacities.

---

If, and to the degree that, one or more of these criteria are not met, a false or inaccurate factual belief formed under such conditions can be considered to have been formed non-voluntarily, as the person did not have an acceptable alternative to forming an inaccurate or false empirical belief. And as mentioned, non-voluntarily formed (forced) false or inaccurate empirical beliefs are the sorts of things that can affect a person's downstream autonomy, where and to the extent that those subsequent deliberations, choices, and actions are causally connected to the false or inaccurate belief.

These conditions obviously have a significant degree of ambiguity and are open to interpretation. This is intentional; whether or not the criteria are judged to be fulfilled in a given circumstance is an evaluation that is context dependent and that will need to be

made on a case-by-case basis. What counts as a person having 'effective access' to the evidence, for example, will depend on how technical the question at hand is, as well as various facts about that particular 'potential knower', such as the professional or social capacity in which they are seeking that information, as well as perhaps their educational background and other demographic factors. What we're entitled to legitimately expect another person to know/have known should vary with variations in such factors. It often won't be fair to apply the same standards of 'effective access' to a medical doctor operating in their capacity as a doctor with that of a patient: we would generally have higher expectations of the lengths we expect doctors *qua* doctors to go to know about medical topics compared with their patients. Similarly, it's plausible that it would be unfair, in many circumstances, to hold a single mother whose highest qualification is a high-school certificate to the same expectations of independent research as, for example, a highly educated professional who also has much more time available to do conduct the sort of research needed to form true/accurate beliefs, and resist forming false/inaccurate beliefs.

The second criterion requires us to make a judgement about what counts as the 'important' part/s of the public knowledge system for any given question. This is to protect against the possibility of *any* instance of misinformation from *any* source coming to be necessarily autonomy-compromising. It can't be the case that anyone's omissions or misinforming speech necessarily compromises another person's autonomy, as there are domains over which others have legitimate expectations of us to resist forming false and inaccurate beliefs despite being the recipient of false, inaccurate or misleading information from others. This is particularly so when our interlocutors or the persons we are listening to are our epistemic peers, when they are speaking *qua* citizen rather than in a professional capacity, and when speech occurs within more private spaces.[25] The bounds of this domain are also context sensitive. This explains why we can consider Tennille (7) to have still had a fair opportunity to form accurate empirical beliefs despite the stranger lying to her. The 'important parts' of the public knowledge system in this example were essentially Tennille's doctor's consultation room, and the parts of the public

---

[25] Importantly, this also provides for free speech protections.

knowledge system Tennille's doctor relied on to provide Tennille with accurate information (e.g. information from the pharmaceutical company, professional guidelines, peer reviewed articles, etc.). As there was not, in this case, any significant misrepresentations or misleading uses of evidence within *these* parts of the public knowledge system, this criterion was met on this occasion.

At this stage, the third criterion is a bit of a 'catch all' and admittedly doesn't seem satisfactory, even for a condition at this level of abstraction. It aims to capture the fact that rather than the evidential *inputs* into a person's reflective processes being inadequate or distorted, external forces can work to disrupt or hijack a person's *actual cognitive capacities*. For instance, in the example where Peter is playing pool, conditions one and two were fulfilled because there was enough evidence available for Peter to make the right calculations to pocket the ball, and this evidence has not been contorted by anything that would misrepresent this evidence (e.g. mirrors had not been sneakily placed on the table in such a way as to optically misrepresent the angle between the ball and the pocket). However, if, as Peter was reflectively making calculations and lining up his shot, his opponent was firing darts past his head to distract him, this would have unfairly compromised Peter's very capacity to reflectively make accurate calculations. Under such conditions Peter would not have had a fair opportunity to form accurate empirical beliefs.

## 2.4. Application of Criteria

Does the application of my framework to the examples presented within this chapter provide a good explanation for the conclusions we reached about the agent's autonomy status in those examples? In what follows I explain the results of applying my framework to the key examples I drew on from other authors within this chapter.

### 2.4.1. Cake/Dogfood (in Killmister 2013)

In Killmister's example (p. 34), the first criterion would have been met because there was adequate evidence available to the agent that would have confirmed, beyond reasonable doubt, that what she was about to eat was dog food rather than cake. The second criterion would have been met because the evidence had not been distorted; e.g. the dog food was not misleading labelled 'cake'. And unless she was suffering from something like somnambulism[24] at the time, the third criterion would have held too. Thus, contra Killmister, the person in her example, at least as the example is presented, should be considered to have autonomously, albeit irrationally, consumed dog food. They had adequate opportunity to form true/accurate beliefs, and as such they formed their false beliefs voluntarily. Because their false beliefs were formed voluntarily, they were autonomous over the episode.

### 2.4.2. Tal (in McKenna 2005)

In McKenna's example (p. 30), the first criterion would not have been fulfilled because Tal did not have access to adequate evidence within an important part of his information system. It was reasonable for Tal to believe, given the existence of an authoritative pharmacist's label on the outside of the bottle the medication was in, that the bottle contained the substance it reported to. In addition, there was no higher-order evidence that indicated that the pharmacy label was erroneous or untrustworthy. It would have been illegitimate for us to expect Tal to have gone to any further lengths to form true beliefs about what was in bottle under the circumstances. Therefore, because Tal did not form his false belief about the bottle's contents voluntarily, it's appropriate for us to individuate the relevant actions act individuation was appropriate, and contra McKenna, we can say that Tal did not autonomously give the 'bad stuff' to Daphne.

Our conclusions, however, might be different if the bottle *was not* actually labelled with an official label from a pharmacist. If Tal had found Daphne unconscious and he had formed false beliefs about what was in an *unmarked* bottle (as well as what Daphne was in need of), our conclusion about the voluntariness of his beliefs – and consequently,

---

[24] This is the formal name for a behaviour disorder that originates during deep sleep and results in people performing other complex behaviours unconsciously while asleep.

his autonomy over the episode — *may* have been different. It would depend on the specifics of the evidence available to Tal at the time, and what beliefs would have been reasonable for Tal to have formed in response to the evidence base in that particular context.

### 2.4.3. King George (in Mele 1995/2001)

Was there adequate evidence available to King George within his knowledge system about the best means to his ends, and if so, did he have reasonable access to this evidence? Without more information, it's hard to make a definitive judgment. Even if there was adequate evidence available, whether or not George in fact had 'effective access' to this evidence depends on why it was that George, according to Mele, *only* had access to the state of his kingdom was through his advisors. For example, why couldn't George have visited villages in his kingdom and seen for himself that various policies were not working, and that poverty was getting worse? It's likely that touring his community and talking to others outside his circle of advisors probably would have quickly revealed his advisor's deception too.

But for the sake of the argument, let's assume that George was bedridden with an infectious disease that meant that only a few people (his advisors) could visit him. Furthermore, let's also assume that the advisors were very clever at covering their tracks so that there was no higher-order evidence available to George that would have cast doubt on their trustworthiness. If these were the circumstances, then we probably would conclude that his false beliefs were non-voluntary, and that his autonomy was compromised by the misinformation his advisors transmitted to him. However, as George is a state leader, we would have legitimately had very high expectations of him to know various things and to resist forming false/inaccurate beliefs. Moreover, someone in his position generally has access to many resources to fulfil these expectations. Therefore, we would probably only conclude that George was non-autonomous if his circumstances were as extreme as those I've painted here.

## 2.5. Implications for Voluntariness and Responsibility Debates

Before I conclude, I also want to briefly comment on the implications of my positive account for debates within political philosophy related to voluntariness and responsibility.

As I mentioned, an objection raised in response to Olsaretti's account of voluntariness is that it exempts from responsibility those who act on the basis of false beliefs they hold as a result of negligence or obtuseness. Here's Colburn's description of this objection:

> It might be objected that if (unlike Hilda) one is responsible for being ignorant of acceptable alternatives, one is responsible for actions performed in that ignorance. I disagree: if an agent acts on the sincere (but mistaken) belief that she has no acceptable alternatives, then at the time she makes that choice it makes no difference to either her voluntariness or responsibility whether she could at an earlier time have become better informed. We might hold her culpable for that earlier omission, but that is a separate appraisal. (2008: 103, fn 6).

In response to this objection, Olsaretti, like Colburn, opts to bite the bullet. She says:

> Like Colburn, I think that we should bite the bullet here. If someone *genuinely* holds false beliefs about what options are available, and acts as she does in order to escape those options, then this has *some* normative significance, and she is not responsible for all the consequences she may be held responsible for had her choice been voluntary. (2008: 113, fn 3).

I think biting this particular bullet is an unattractive move. An account that has implications for ascriptions of voluntariness and responsibility that does not take into account the etiology of our beliefs about our actions is inadequately sensitive to the complex division of epistemic labour in our societies — and the interlinked epistemic standards, norms, and duties associated with that labour — that are crucial to peoples' very abilities to identify, understand and discharge their responsibilities as moral agents. Biting the bullet is therefore likely to result in erroneous ascriptions of substantive responsibility. Such errors can in turn lead to the wrong people within a network being blamed, punished, forced to compensate others or change their behaviour, etc. This would be unjust. We cannot justly hold someone substantively responsible if, and to the degree

that, their actions were due to false/inaccurate beliefs formed because someone else failed to discharge their epistemic duties to that person. Scanlon (1999: 258) argues that we cannot justly hold someone substantively responsible until we have 'done enough' by placing her in a situation in which she can make a choice that is not forced. Similarly, we cannot justly hold someone substantively responsible until we have 'done enough' by ensuring her socio-epistemic environment is such that her empirical beliefs are voluntary as I have characterised it. Moreover, leaving the underlying issues unaddressed by targeting the wrong actors may even encourage the flouting of epistemic responsibilities within our information systems and facilitate their corruption and decay.

As an illustration of how Olsaretti's and Colburn's account would lead to the wrong ascriptions of substantive responsibility, and how my framework would improve these assessments and practices, I'll use an example that Colburn (2008) and Olsaretti (2008) both refer to. In this example, an armed robber holds up a bank demanding all the money in the safe. Despite appearances, the robber's gun is actually not loaded, and so no one is actually at risk of being shot. But the bank staff have no way of knowing this. Fearing that she or others will be harmed if she does not follow his demands, the manager opens the safe and hands over the money. Both Colburn and Olsaretti conclude that she did so non-voluntarily as even though the gun was not actually loaded, she acted as she did because she believed it to be, and hence because she believed she had no other acceptable alternative other than the choices and actions she carried out. However, what's actually critical here is that, given the circumstances, it was *reasonable* for her to have formed this false belief. In this scenario, she did not have a fair opportunity to form accurate/true beliefs: she was confronted by an aggressive man holding what any sane person would have perceived to be an authentic, loaded weapon.

However, if instead she was confronted by a child demanding that she open the safe with a gun that was obviously a water pistol, we would have concluded that it would have been *unreasonable* for her to form the false belief that the gun he held was loaded with lead bullets, and that lives were at risk. If she had opened the safe and offered up the money to the child on the basis of these false beliefs, we would have concluded that

she opened the safe and gave away the money *voluntarily* (and by extension, autonomously), as it would not have been reasonable for her to have formed this false belief under these circumstances.

Therefore, what I have shown is why Olsaretti and Colburn need not, and should not, bite this bullet. My account of when empirical beliefs count as having been formed (non)voluntarily complements Olsaretti's account of voluntariness for choices and actions. Using my criteria to determine the voluntariness of false beliefs, and of the subsequent choices and actions related to these beliefs, closes a key gap in Olsaretti's framework. It will enable those users of it to more reliably determine when the agent's responsibility for their choices and actions is and is not preserved. This is important because it is not appropriate to attribute moral responsibility to someone for their choices and actions where those choices and actions are the product of false/inaccurate beliefs they hold by virtue of reasonable deference to others. It will also help us to identify where responsibility ought to be placed, which will in turn improve the operation of our public knowledge systems.

## Conclusion

In this chapter I defended the position that false beliefs are among those things that *can* undermine a person's autonomy. I provided an overview of two broad ways false beliefs can do so: by undermining a person's control over their choice of values, and by undermining their control over whether or not their actions realise their values. But because not all false beliefs are autonomy compromising, I then provided a framework to explain *which* false beliefs compromise personal autonomy. I argued that false beliefs are autonomy compromising when they are formed non-voluntarily. The (non)voluntary status of the etiology of false beliefs depends on whether a person possessed a fair opportunity to form accurate/true empirical beliefs, or to resist forming inaccurate/false empirical beliefs. I proposed a set of jointly sufficient conditions to identify these circumstances. A person has a fair opportunity to form true/accurate empirical beliefs (i.e. their empirical beliefs can be considered to have been voluntarily formed) if the following conditions all hold: (1) Given what is known at the time, or what one can reasonably expect to be

known, adequate evidence is available within the public knowledge system, and a person has effective access to this evidence; (2) The body of putative evidence a person has access to is generally presented with integrity; there are no significant misrepresentations or misleading uses of evidence within important parts of the public knowledge system; (3) A person's reflective capacities are not otherwise unreasonably interfered with from without.

Applying my criteria to the examples authors have used within the literature to date led to some interesting results: contra Killmister (2013) my account suggests that the person who eats dogfood rather than cake in her example was in fact autonomous; against McKenna (2005), Tal did not autonomously give 'the bad stuff' to Daphne; and the autonomy status of Mele's (1995) King George is unclear. My framework also demonstrates that Colburn and Olsaretti need not bite the bullet — false beliefs that are the product of carelessness and obtuseness can, in some circumstances, be voluntarily formed; hence, we are able to hold those people responsible for their choices and actions that are tied to such false beliefs.

As the integrity of our personal autonomy is not solely an individual enterprise, and significantly turns on those relationships one is legitimately dependent on to acquire certain types of information, respect for the autonomy of others demands that certain actors within knowledge and information systems meet their practice-based, and general moral, obligations to provide others with certain types of information and to refrain from transmitting misinformation. When these actors fail in their epistemic responsibilities, one important way people are wronged is in virtue of the potential of such acts and omissions to undermine their personal autonomy.

# CHAPTER THREE:
# PRE-EMPTIVE SILENCING

## *Introduction*

Chapter three argued that, under some circumstances, an individual's personal autonomy can be compromised when they are misinformed by others. In this chapter I explore what implications (if any) misinformation-induced losses of autonomy can have, in turn, for these misinformed persons' freedoms of speech.

It's often said that dissent over empirical claims, even that which misrepresents others' testimonies or that is unsubstantiated, deserves special protection simply because it is speech. Therefore, no matter what we think of its content, we must afford it the protections that we extend to almost all speech. In their book *The Fight Against Doubt*, de Melo-Martin and Intemann (2018: 75) say that:

> Prohibiting normatively inappropriate dissent could prevent or assuage public confusion, lack of support for needed policies, wasted resources, and unwarranted conservatism in scientific results. Nonetheless, in democratic societies this strategy is implausible. It clearly conflicts with democratic principles of free speech and expression.

My aim in this chapter is to examine the prevailing assumption embedded in this and other assertions like it: that the value of freedom of speech itself is necessarily only, or best, served via a thoroughgoing toleration of all forms of dissent about science in all parts of the public knowledge system.

Since the late 1980s/early 1990s, feminist free speech theorists have argued that some people's speech can silence others in a way that violates their rights to freedom of speech.[26] If these arguments are sound, then the very same free speech reasons for protecting some category of speech also provide a reason to restrict it, and hence we are left with a conflict of the *same rights*. Decisions must then be made about whose rights are

---

[26] I take the 'silencing argument' to have been first articulated by Catherine MacKinnon (1987, 1993), and since then to have been translated and extended by the likes of Rae Langton, Jennifer Hornsby, Caroline West, Mary-Kate McGowan, and Ishani Maitra (among others).

weightier. The free speech arguments of feminist philosophers and legal scholars have thus met traditional liberals and libertarians on their own sacred ground. 'Silencing' arguments have therefore been controversial and generated considerable attention within contemporary free speech scholarship over the past three decades.[27]

I endorse Lorna Finlayson's (2014) assertion that the silencing argument contains an indispensable insight, and more profound potential, than its defenders or its challengers have so far appreciated. Whilst most of the scholarship on silencing has been written in the context of issues like pornography and hate speech, I will here develop a new kind of silencing argument and apply it — perhaps somewhat surprisingly — to social relationships that have been outside its purview to date: the relationships between scientific experts, intermediaries who communicate science, and the public. Silencing, I contend, is a much more widespread phenomenon than is usually acknowledged, and one that does not only affect those who are members of marginalised social groups. In particular, I think that a type of silencing can occur when the transmission of knowledge, information and evidence between scientific experts and non-experts goes awry.

My key contribution here is to offer an account of a new type of silencing that captures and explains the second of the three inter-related non-epistemic wrongs and harms associated with some dissent about science that I identify in this thesis. I first survey five existing accounts of silencing in the philosophical literature (4.1.). I then argue in 4.2. that it is unlikely that these five species represent a complete taxonomy of silencing phenomena, as all of these accounts share two common background assumptions that often do not hold. From there I present my novel conception of silencing, which I call 'pre-emptive' silencing (4.3.). I argue that because pre-emptive silencing is a fairly widespread phenomenon, so are the wrongs and harms it constitutes. As I contend in section 4.4., there are compelling reasons for viewing existing accounts of silencing as free speech violations. Building off these arguments I will argue that certain instances of pre-emptive silencing should also be regarded as being in violation of peoples' free speech rights. Although this will not be not sufficient to show that *any particular instance* of pre-emptive silencing is a free speech violation, I show that this type of silencing is both

---

[27] This prominence makes its current neglect within debates in philosophy of science over dissent all the more surprising.

philosophically interesting and warrants concern and further attention. In section 4.5. I address two key objections: that pre-emptive silencing is ubiquitous, and therefore an empty concept, and that pre-emptive silencing represents a 'slippery slope'. My response to these objections clarifies the bounds of morally and politically problematic pre-emptive silencing.

## 3.1. Two Fundamental Types of Silencing

Although several different species of silencing have been identified within the literature to date, they can be organised into two fundamental kinds of silencing: direct silencing, and indirect silencing.

### 3.1.1. Direct Silencing

A locutionary act is the production of words, phrases, or sentences. Cases of direct silencing are those in which no locutionary acts are performed at all. Common cases of direct silencing are often associated with speech that orders or threatens (Langton 1993). Sometimes a person orders silence when performing a particular social or professional role, e.g. the judge who demands 'Silence in the court;' the tennis umpire who calls 'Quiet, please;' the exam moderator who requests that people in the corridor 'Stop talking;' and so on. Such speech acts bring about silence in virtue of the authority and/or power of the speaker within the relevant domain.

There can also be speech, or expression, that silences because it has been vested with authority or power to elicit a silent response. When I lived in Bangkok, if I happened to be in a public place at 8 a.m. or 6 p.m. I would have to stand still and remain silent while the Thai national anthem *Phleng Chat* (in Thai: เพลงชาติ) was broadcast over loud speakers. I also had to stand and remain silent whilst the royal anthem *Phleng Sanrasoen Phra Barami* (เพลงสรรเสริญพระบารมี) played in cinemas before films. Those who do not remain silent in such circumstances face rebuke and social penalties. But they can also face more severe penalties too. In 2008, Chotisak Onsoong was formally charged with

lèse majesté for failing to stand during the royal anthem at a Bangkok cinema.[28] Thailand's *lèse-majesté* law criminalises speech that defames or insults Thailand's King, Queen, Heir-Apparent or Regent. These laws silence speech by threatening potential offenders with imprisonment; people (generally) self-censor and remain silent because they (very reasonably) fear the severe consequences.

### 3.1.2. Indirect Silencing

In order to 'do things with words', certain things must be true of the context in which speech acts are uttered. These conditions, or the criteria that must be satisfied for a speech act to achieve its purpose, are called *felicity conditions* (Austin 1955/1962). Indirect silencing occurs when there are objectionable failures of the felicity conditions necessary for certain utterances to constitute speech acts of the kind the speaker desires. The distinctions among different types of indirect silencing relate to the different ways that felicity conditions can fail.

**Illocutionary Silencing[29]**

Illocutionary silencing[30] was first identified explicitly by Langton (1993) and further developed by Hornsby and Langton (1998). A speaker who is illocutionarily silenced is silenced because, despite her audience hearing every word she utters, she is unable to make herself understood. This occurs when the audience fails to correctly recognise the speaker's illocutionary *intention*, i.e. the speaker's intention in relation to what sort of

---

[28] See 'Thailand: Moviegoer Faces Prison for Sitting During Anthem', New York Times, 24 April 2008. Charges were not dropped until 2012. Prison sentences are three to fifteen years per count. Thailand's lèse majesté law has been described by some commentators as the 'world's harshest' and 'possibly the strictest criminal-defamation law anywhere'. Also see: 'How powerful people use criminal-defamation laws to silence their critics', *The Economist*, 13 July 2017.

[29] Maitra (2009) engages with Langton's and Hornsby's accounts of illocutionary silencing and argues that their account ought to be replaced with her own account of silencing as 'communicative disablement'. I don't summarise Maitra's account here due to space.

[30] According to Austin (1955/1962), illocutionary acts are those that a speaker performs merely in virtue of uttering the right words in the right contexts. In such contexts, the utterance constitutes the act in question e.g. marrying.

speech act they are performing. This *uptake failure* occurs, for example, if Georgina says 'No,' intending to refuse a sexual advance, and Michael fails to recognise her intention to refuse. Under such circumstances Georgina fails to communicate her refusal to Michael. Langton and Hornsby hold that pornography can silence women in this way because it can work to prevent women from achieving the effects they want to achieve in speaking (i.e. to successfully refuse a sexual advance); they contend that because pornography *is* subordination, it can cause men to hear a woman's 'no' as 'yes'.[31]

It's important to note that Hornsby and Langton do not think that *every* occurrence of an uptake failure constitutes a type of silencing; only those that are brought about *systematically*. A systematic disablement is one that occurs robustly across a diverse range of contexts, and that is difficult to overcome when it does occur (Maitra 2009).[32] For example, imagine that Fede is driving, and Sally is navigating. Sally tries to direct Fede to turn left, but he fails to recognise her intention to direct him because he is distracted by a person on a bike ahead who he thinks is his daughter. In this case, the addressee (Fede) fails to recognise the speaker's (Sally's) illocutionary intention (to direct), so there is a failure of uptake. But despite this, it is not a case of illocutionary silencing because the uptake failure in question does not occur as a result of systemic social dynamics. In this case, Fede's failure to recognise Sally's intention was an irregular fact about the situation. Under similar circumstances (i.e. when other men are driving and Sally is a passenger), it's very unlikely that others (and in particular, other men) would fail to recognise Sally's intentions. Thus, systemic features of society did not influence Fede's recognition failure.

However, in societies where men regard women as being incapable of navigating, and where this belief is very widespread, it's likely that a man in such a society would not regard a woman's utterance as a directive, just like they would not regard a cat's meow as a directive, either. Because such a failure of uptake in the case of a woman making a directive would have been caused by systemic features, such a failure would constitute

---

[31] Note that this is not to say that the act they're trying to perform (refusal) is causally downstream from the utterance. The utterance is constitutive of illocutionary acts e.g. a band leader giving an order to march is constituted by their speech act, not some effect they bring about after speaking.

[32] See Maitra (2004) for more on this notion of systematicity.

illocutionary silencing. Langton, Hornsby and others take the illocutionary speech act of pornography (subordination) to do just this, and so in doing, to be able to illocutionarily silence women.

**Sincerity Silencing**

McGowan (2014) outlines an account of a distinct silencing phenomenon that she calls 'sincerity silencing'. This type of silencing also involves a systemic interference with one of the necessary felicity conditions of A successfully communicating to B, but this time the inference is with the *sincerity condition* rather than with recognising the speaker's illocutionary intention. McGowan explains that this form of silencing occurs when a woman says 'No,' sincerely intending to refuse sex but when – although the man recognises her intention to refuse – he mistakenly believes that she is refusing insincerely. In this case, the addressee recognises the speaker's illocutionary intention (to refuse) but fails to recognise that the speaker is doing so sincerely. Where the hearer's failure to recognise that the speaker sincerity condition is met is brought about in a systematic manner, it constitutes a case of sincerity silencing. It is not a case of illocutionary silencing, however, since the man recognises the woman's intention to refuse his advances; thus, there is no uptake failure in cases of sincerity silencing.

**Speaker Authority Silencing**

In an earlier article, McGowan (2009) articulates another type of silencing, this time one constituted by a systemic interference with the recognition of the speaker authority felicity condition. Here, when a woman refuses sex by saying 'No', the addressee recognises her sincere intention to refuse the sexual advance, but does not believe she has the authority to do so, and hence falsely believes that the attempted refusal fails. When one person falsely believes that another person does not have authority over who has sexual access to their body, and this false belief has a systemic cause, the communication failure counts as a type of silencing. One can imagine that this type of silencing commonly occurred in historic societies, and frequently still occurs today in contemporary societies that regard sexual intercourse within marriage as a spouse's right.

**True Feelings Silencing**

More recently, McGowan (2017: 49-50) has outlined a fifth type of silencing that she calls 'true feelings silencing'. This variety of silencing occurs when the addressee fails to

recognise that the refusal reflects the speaker's true feelings. For instance, imagine a case in which a woman says 'No' and where the addressee recognises this as an authoritative and sincere refusal of his sexual overtures. However, in this case he also falsely believes that the refusal is not what the woman's 'deeper self' really wants; he mistakenly takes her innermost or 'truest' desire to be to have sex with him, and proceeds to rape her. McGowan says that if, like the three immediately prior types of silencing, this false belief is brought about systemically, then another type of silencing has been committed.

## 3.2. A Complete Taxonomy of Silencing?

Rather than critiquing or defending any or all of these accounts of silencing, my interest is instead whether these types of silencing – assuming they are all sound – together represent an exhaustive account of the ways we can be silenced.

One reason to suspect that they do not constitute a complete taxonomy of silencing is that the two fundamental types of silencing both share two background assumptions that don't actually hold in many circumstances. The first of these assumptions is that people already have the communicative resources they need to outwardly express the content of their mind. All of the accounts of silencing we've encountered so far assume that silencing occurs either when, and because, people are blocked from being able to put these resources to work (as in cases of 'direct silencing'), or that after deploying these resources, communication fails due to a mistake on behalf of the listener (as in all four varieties of indirect silencing). However, people can also lack the resources they need to communicate their thoughts and feelings to others.

In Kristen Roupenian's viral short story, *Cat Person,* we meet a 20-year-old college student, Margot, who has just started dating a much older Robert.[33] At the end of a date, Margot suggests that they drive to Robert's house. As she watches him undress, she realises that she is not interested in sleeping with him. However, she doesn't believe she possesses the communicative resources needed to halt the unfolding chain of events:

---

[33] Published in *The New Yorker*, 11 December 2017.

> [T]he thought of what it would take to stop what she had set in motion was overwhelming; it would require an amount of tact and gentleness that she felt was impossible to summon. It wasn't that she was scared he would try to force her to do something against her will but that insisting that they stop now, after everything she'd done to push this forward, would make her seem spoiled and capricious, as if she'd ordered something at a restaurant and then, once the food arrived, had changed her mind and sent it back.

However, moments later, Margot seems to change her mind: the thought passes through her head that she may be able to derive some enjoyment from the experience after all. However, her train of thought reverses again after Robert's hastiness is physically painful for her. She flinches, and Robert apologises and promises to slow down. Margot doubts the promise is sincere, but, once again, she feels a lack of communicative resources to convey the content of her mind:

> Yeah, right, she thought, and then he was on top of her again, kissing her and weighing her down, and she knew that her last chance of enjoying this encounter had disappeared, but that she would carry through with it until it was over.

Margot ends up tolerating Robert as he single-mindedly throws 'her around as if they were in a porno' and makes her feel like a rubber doll that was a 'prop for a movie that was playing inside his head'.

Margot and Robert's sexual encounter seems to be technically consensual.[34] But Margot still considers it to be 'the worst life decision' she's ever made. However, if it was consensual, what (if anything) can we say was problematic and objectionable about it? In a recent article in *Ethics*, Rebecca Kukla says that 'Consent, including completely autonomous, unmanipulated consent, is never going to be sufficient to make sex go well — we can consent to all sorts of lousy sex, including demeaning, boring, alienated, unpleasantly painful or otherwise harmful sex' (2018: 3). The interesting question is *why* do

---

[34] Whether or not it in fact was even consensual may also be debated. It might be possible that even if Margot took herself to have consented, she had not. The reason for this is that consent presupposes authority over the thing we're consenting to, and plausibly there are kinds of treatment that we just can't give effective consent to (e.g. seriously debasing treatment). The debate would then shift to whether this sexual encounter in fact met the criteria for that sort of treatment. I put such issues to the side here and assume that this encounter did not raise any issues for consent. Thank you to Lachlan Umbers for raising this issue.

people consent to be treated that way? Of course, there are all sorts of explanations. However, given that much of what went wrong in *Cat Person* involved communication issues, it's possible that silencing phenomena are at least sometimes part of the explanation. But if we reflect on *Cat Person*, it doesn't seem like any of the five types of silencing on the table so far occurred. Margot did speak, and Robert did seem to understand what Margot outwardly communicated. What Margot seemed to lack were the sorts of communicative resources that would have helped her to outwardly express what was *actually* running through her mind and what were her preferences *in fact* were. And it's possible that the reason why Margot lacked these communicative capabilities has something to do with systemic patriarchal forces acting upon her, and whose power Robert did little to dissolve. So a new account of silencing might be in order if we want to capture these features and dynamics.

Although I think that this would be a worthwhile project, I won't be pursuing it here. I am more interested in probing the more radical implications of a second background assumption in existing accounts of silencing (and that an account of silencing based on Margot's lack of communicative resources to speak her mind, as sketched above, would also likely share). That assumption is that we already have thoughts in our heads that are sufficiently well-formed and meaningfully our own that we wish to communicate. But it doesn't seem like this is always or even generally the case: another prominent, but distinct, feature of *Cat Person* was the way that at various points in the story Margot seemed unable to identify what her thoughts, feelings, and desires actually *were*, or even whether she had any that could be counted as such at particular points. For much of the story she was confused as to what she thought, felt or wanted.

Each type of silencing outlined so far takes the person to be silenced either at the point at which their expression, or the effect of their expression, is frustrated in some way; thus, each account assumes that the thoughts a person wants to express, or does express, are already present in their minds; that they are sufficiently complete; and that they have been formed 'freely' or voluntarily and can therefore be considered their own, in a meaningful sense. For instance, in the direct silencing example that references Thailand's *lèse-majesté* laws, it's assumed that individuals in Thai society generally actually have thoughts like 'I think the King acted immorally' or 'I think Thailand should be a

Republic,' and that the main issue is that they cannot outwardly express these beliefs due to their reasonable fear of imprisonment. However, it seems likely, given the severe criminal penalties, together with the presence of a well-oiled royalist propaganda machine which ensures that every nook and cranny of Thai society is saturated with royalist imagery,[35] that such thoughts never enter into the minds of many Thai people in the first place. At least in one sense, I think these cases are even more concerning than those in which people do in fact entertain and hold such thoughts, but do not possess the freedom of speech required to outwardly express them, i.e. situations in which people retain their freedom of thought.

Similarly, the focus of the other four types of silencing is on people being prevented from doing what they intended to do with their words. It's not that they don't outwardly express pre-existing thoughts that are their own, because they of course do in all of these examples. All of the women actually had the thought 'I do not want to have sex with this man,' and did in fact express this thought by saying 'No' or 'Stop.' Margot also had these thoughts, but was unable to express them. Yet within patriarchal societies, is it reasonable to assume that women generally already have such thoughts? Surely a defining feature of patriarchy's success — or that of any other oppressive ideology — is that there will be (perhaps many) women who *never even have* such thoughts, possibly because they've been conditioned to think that making men comfortable is the most important thing, and who subsequently always say 'Yes.'[36] When these women consent to sex,

---

[35] After the military takeover of May 2014 the propaganda-peddling apparently worsened. As *The New York Times* reported in 2015, coup leader General Prayuth Chan spent $US 540 million on a promotional campaign with the moniker 'Worship, protect and uphold the monarchy': 'The campaign includes television commercials, seminars in schools and prisons, singing contests and competitions to write novels and make short films praising the king.'

[36] The phenomenon I'm trying to highlight here has an affinity to the literature in analytic philosophy on adaptive preferences (e.g. see Nussbaum 2001, and improvements to Nussbaum's account by Begon 2015). But even more importantly, the concept of pre-emptive silencing that I develop in the next section has obvious links to theorising within both feminist and black radical thought about the mechanisms of oppression, the place of propaganda in these mechanisms, and the way that oppressed persons can come to internalise and perform their oppressors' messages about them. In feminist thought, see e.g. legal scholar MacKinnon (1989, 1993) and philosopher Clare Chambers (2007). Prominent examples from black radical thought are Frantz Fanon (1952) and the South African apartheid activist Fanon's writings inspired, Steve Biko (2002 [1978]). Notably, Crasnow and Waugh argue that feminist philosophers have

should we be unconcerned? Should we simply conceive of their speech as an accurate reflection the contents of a freely operating mind of her own?

What seems to have been neglected in extant accounts of silencing are precisely these kinds of cases: those involving people who *lack the sorts of thoughts* we might expect it would be reasonable to have under the circumstances, and who by virtue of this say 'Yes' when it would be reasonable for them to say 'No.' What seems to have been neglected to date in the silencing literature — which takes for granted the prior existence of voluntary, sufficiently formed thoughts, and assumes that agents' freedom of thought is intact — is the *very relationship* between our freedom of thought and our freedom of speech. In what remains of this chapter I argue that when certain conditions hold within our external socio-epistemic environments, we can be 'pre-emptively silenced', and silenced in a way that impinges on our free speech rights.

### 3.3. Pre-Emptive Silencing

As I shall now argue, another type of silencing occurs when people's freedom of thought has *already been* or is, in the course of a social interaction, compromised by the speech of other agents. As I argued in chapter three, when we examine the etiology of mental content, it's sometimes the case that the features of a person's prevailing socio-epistemic environment mean that some of their beliefs cannot be considered to have been formed voluntarily. If we were able to compare the worlds that manifests under these conditions with a world in which people's mental content was formed voluntarily, it's plausible that we would then observe different states of affairs regarding: (i) *who does and does not speak* in relation to some topic; (ii) what people *actually say* when they do speak; and (iii) what

---

paid insufficient attention to popular culture. The say popular culture is the 'primary vehicle for presenting and reinforcing social roles and stereotypes of women that follow from, lead to, or perpetuate discrimination on the basis of sex or gender' (Crasnow and Waugh 2013: 1). Similar arguments have long been made by black feminist scholars. E.g. bell hooks (1992: 5), argues that the way blackness itself is seen, and how black women are treated, are determined by people's reaction to popular representations of black women in the media. In a similar vein, Shawna Hudson argues that black stereotypes 'reflect and distort both the ways in which black women view themselves' and 'the ways in which they are viewed by others' (1998: 249).

people *actually do* with their words. Of course, the degree of (dis)similarity between these different worlds in relation to each of these features will vary depending on which beliefs within a person's web of beliefs have been formed non-voluntarily. But I want to say that under certain circumstances, this can produce another kind of silencing that I call pre-emptive silencing because — whether intended or not — the conditions that result in a person forming beliefs non-voluntarily simultaneously serve to curtail the potential scope of a person's future speech, and subsequently, what they are able to do with their words.

It will help to present some examples of the phenomenon I have in mind.

### 3.3.1. Intimate Trust Relations

I'll start with another fictional example, this time from the British film *Sliding Doors* (1998). The film begins with Helen, the protagonist, being fired from her job in London. She leaves the office and goes to take the Tube home. At the station, her life — and the film — splits into two parallel worlds. In one of these worlds she manages to catch the Tube and she arrives home early. In the other, she narrowly misses the Tube, the next train is then delayed, and a subsequent chain of events means that she arrives home much later by cab. The following are the parallel conversations we observe within the two worlds when she first arrives home (note that it's probably best to read the first column, and then the second):[2]

---

[2] The text that follows has been paraphrased from the original screenplay by Peter Howitt (1998).

|                                          |                                          |
| ---------------------------------------- | ---------------------------------------- |
| SCENARIO ONE                             | SCENARIO TWO                             |
| HELEN CATCHES THE TUBE                   | HELEN MISSES THE TUBE                    |

*Setting: Helen enters the apartment. She is about to call out Gerry, her boyfriend, when she notices the towel on the floor and a discarded coat. She can hear Elton John playing in the bedroom. She walks towards the bedroom and opens the door. In the bed is her boyfriend with another woman.*

*Setting: Helen enters the apartment. Gerry is in the shower. His ex-girlfriend left moments ago.*

**Helen:** 'I didn't think you liked Elton John.'

**Helen:** 'Hello... Gerry?'

**Gerry:** 'I do sometimes.'

**Gerry:** 'Helen, is that you?'

**Helen:** 'Yeah? Just shows, you never really know people, do you? Well, I've had a dreadful day. I got sacked. And so did you, it would seem. Cup of tea?'

**Helen**: 'Have you just got up, you lazy git?'

*The other woman (Gerry's ex-girlfriend) leaves. Helen turns and walks out of the bedroom leaving Gerry staring, <u>speechless</u>. Suddenly she turns and comes back through the door and launches herself on him screaming and punching.*

**Gerry:** 'Yeah... I er... I didn't sleep that well, you know er... then I went into a deep... really, er, deep, deep sleep when you left for some er... you know...I think I might be coming down with er...anyway what are you doing home at... Oh my God, what happened to you?'

*Gerry turns radio off.*

'Can't stand Elton John, you know that. What happened? What happened to your head?'

**Helen:** 'You bastard. You useless shagging bastard... I am working all the hours under the sun to support you while you supposedly write your first novel and I come home and catch you in Lady Shagging Godiva.'

**Helen:** 'Well, I got sacked first, and mugged later.' [*Helen begins to crumble with the shock of it all*].

**Gerry:** 'Helen, try not to over-react.'

**Gerry:** 'Wait wait — whoah, whoah, slow down — stop, wait a sec, wait a sec. OK. Come here. Just... [*He puts his arm around her and takes her to the bed and sits her down, comforting her*]. 'That's it. Now then, just relax - take a deep breath. OK, just... Take it easy, you've obviously had a bit of a shock... You're in shock, sweetheart. Now then... there we are.'

**Helen:** 'I should try-not-to-over-react?! Like I don't have a point? Is that what you're saying? Is it? I have a point, Gerry! Believe me — I have a very big point. How long, huh? Do you love her? No, don't tell me. I'm not interested — No, do tell me. I am interested. Hmm? Question's too difficult? I'm only asking because I need to know exactly how big a mug I am. Hmm?'

*But Gerry <u>simply can't speak</u>. So Helen announces 'It's over', and picks up her coat and walks out, slamming the door.*

**Helen:** 'I waited for hours in casualty and I had to give a statement and all that and I couldn't help thinking if I'd have just caught that bloody train it would never have happened. I'd have been home ages ago and...'

**Gerry:** 'Well look you mustn't think like that. You can't go around wondering "what if this", you know, "if only that". It's done now. Come on. I'm going to take you out.'

**Helen:** 'Oh, I love you Gerry'.

Clearly, the conversations within the two possible worlds are *radically* different. Most obviously, the words uttered and propositions expressed differ markedly, as well as the balance of who speaks and who listens. But most importantly, not only does Helen speak her mind and tell Gerry precisely what she's sincerely thinking and feeling in the first scenario, she also *does* several things with her words that she is not able to do in the second scenario — i.e. she performs radically different *speech acts* (Austin 1955/1975:

151-164). Her locutions perform at least three illocutionary acts. First, she gives a *verdictive*[37]: she characterises Gerry as a 'bastard'. Second, she issues a *behabitive*[38]: she protests his behaviour and treatment of her. Third, she performs an *exercitive*[39]: she announces that their relationship is over. Furthermore, communicating this makes it so: the *perlocutionary effect* of this exercitive — the act performed by her communication — ends their relationship.

In stark contrast, in the second scenario, through her locution 'I love you Gerry,' Helen's verdictive is that Gerry is worthy of her love; the behabitive she issues is that she's grateful for his concern and care for her welfare; and the exercitive she tacitly performs through her locutions is to endorse the continuation of their relationship. The speech acts between the two scenarios are therefore diametrically opposed to one another, and the two worlds are very far apart.

The claim I want to argue for is that Helen was being *silenced* in the second scenario. In that world she was being prevented from speaking *her* mind. Whilst Helen does of course literally speak and perform successful locutions and speech acts, the words she speaks and the speech acts she performs are *not hers*, in a meaningful sense. This is because if Helen were adequately apprised of pertinent facts — that Gerry had been sleeping with his former partner for many months — we know that she would not choose to utter what she does in scenario two; we know that it is likely that she would utter words and perform speech acts that say and do *precisely the opposite*, as scenario one makes clear. In the manifest world of scenario two, Helen was prevented from speaking these words, words that represented and reflected *her* mind; she was therefore being *silenced*.

Helen was silenced by Gerry's own speech, together with his omissions of speech. Through the speech Gerry expressed or withheld he conveyed a picture of the world to

---

[37] A verdictive is, according to Austin (1955/1975: 153) 'the delivering of a finding, official or unofficial, upon evidence or reasons as to value or fact.'

[38] Behabitives include 'the notion of reaction to other peoples' behaviours and fortunes and of attitudes and expressions of attitudes to someone else's past conduct or imminent conduct' (Austin 1955/1975: 160)

[39] An exercitive is 'the giving of a decision in favour of or against a certain course of action, or advocacy of it' (Austin 1955/1975: 155).

Helen that was *not real* — i.e. he *misrepresented* the world — which led Helen to form false and inaccurate beliefs about the world she actually inhabited (e.g. that her relationship with Gerry was sexually exclusive). By controlling what Helen believed, Gerry gained control over Helen's internal reflections and the deliberative conclusions she reached. When Helen says 'I love you' in the second parallel world, the image of Gerry's guilt-stricken face over Helen's shoulder indicates that Gerry is fully aware that Helen does not *voluntarily* believe and thus *autonomously* say these words. He knows that by not providing Helen with an adequate opportunity to form true/accurate beliefs to input into her deliberative processes — in Mele's (1995) words: by ensuring Helen was 'informationally cut off' — that he *manufactured* Helen's doxastic states, her profession of love for him (i.e. her speech), as well as the acts Helen performed with her words. We can therefore call the type of silencing Helen experienced *pre-emptive silencing* because it prevented other speech and acts from happening and taking place.

Helen's socio-epistemic environment in scenario two would have failed each of the epistemic criteria I posited for voluntarily formed beliefs in chapter three:

(i) Adequate evidence was not available to Helen;

(ii) Relevant evidence, where it was present or referenced, was significantly misrepresented; Gerry misleadingly presented pertinent evidence by lying, misleading or otherwise deceiving Helen;

(iii) Gerry also interfered with Helen's reflective capacities via 'gaslighting', which made Helen doubt her own reflective abilities (e.g. when later in the film he denied there was a second glass on the table and told Helen that she was paranoid and 'seeing things').

If instead Gerry had confessed to Helen that he was having an affair, Helen would have had the information (testimonial evidence) she needed to reflect freely and to voluntarily form thoughts and preferences that were meaningfully *hers*. When she subsequently spoke, she would have been able to properly speak *her* mind and thus do things *she* wanted to do with *her* words. She would not have been silenced.

In sum: where, and to the degree that, a person holds false or inaccurate beliefs, and these beliefs have been formed non-voluntarily, the autonomy status of their down-stream actions — including their locutions and associated speech acts — suffers. In the domain of speech, this can produce a form of silencing: pre-emptive silencing.

### 3.3.2. Scientific Experts, Intermediaries & the Public:

It seems that the phenomenon of pre-emptive silencing is likely to be more common within some social relationships and circumstances than others. It will be a particular risk, I submit, in relationships characterised by significant asymmetries of knowledge, information and evidence between parties. One of these domains, as illustrated by the example from *Sliding Doors*, is intimate relationships. We don't have direct access to the true thoughts and feelings of a partner; only they can know what they are truly thinking and feeling, and we rely on their testimony to gain access to this knowledge. In addition, short of hiring a private investigator or using surveillance equipment, we also don't have direct access to another person's actual behaviour and actions when we're not present to observe it. Of course, it's possible for there to be other evidence, including other people's testimonies and higher-order evidence, that allow us to make inferences about another person's thoughts, feelings and actions. For example, in the film, that Helen saw two brandy glasses beside the bed was one the clues that led her to eventually ask Gerry outright whether he was having an affair. But unless there is enough good evidence to piece together a sound picture of what's really going on — and evidently, cheats are often good at covering their tracks — these inferences will be made under conditions of significant uncertainty, and the picture we build in our minds could easily turn out to be radically inaccurate.

Relationships between technical experts and non-experts are also often characterised by stark asymmetries of knowledge.[40] There are at least two key ways in which the potential asymmetry in knowledge between experts and non-experts can be reduced or augmented. The first relates to the actions of the experts themselves. If experts don't share

---

[40] It's possible that these two relationships are not actually different relationships, if one thinks that one is an 'expert' about one's own thoughts, feelings, and actions, and that other people are, in comparison, not experts about these things, or at least that they generally have less expertise (there may be exceptions).

their knowledge, or they share their knowledge in a way that is inaccessible to those who are not experts, the asymmetry in the knowledge of experts and non-experts will grow. This is not always problematic, but it can be when the experts' knowledge is morally or politically relevant. Likewise, if experts share *mis*information with non-experts, the gap will also increase. For instance, let's assume that in the first example featuring Tennille in chapter three (p. 25), there was a handful of scientists within the broader scientific community who were working for the tobacco industry and who asserted, when they believed otherwise, that smoking and cancer are unrelated. By propagating this misinformation, they likely contributed to widening the knowledge gap between the expert community and non-experts by creating the sort of confusion that can lead non-experts to assign lower credences to propositions that were accepted as being beyond reasonable doubt by a strong consensus of experts not employed by the industry. These experts may also have created or entrenched knowledge asymmetries by contributing to distorting higher-order evidence; they may have misrepresented the degree of (dis)agreement within the scientific community more broadly; exaggerated their own qualifications; omitted important caveats about the methodologies they used in their research; not been transparent about their own funding sources, etc. On the other hand, experts can contribute to reducing knowledge asymmetries by communicating scientific knowledge to non-experts in accessible ways, and by providing higher-order evidence alongside their first-order claims, as well as adequate accounts of the first-order evidence and reasoning that supports their claims.

The second key type of actor within the public knowledge system that can augment or reduce knowledge asymmetries between experts and non-experts are intermediaries. Intermediaries, like online and offline media outlets, are the primary conduits of scientific information to non-experts (Scheufele 2014). At their best, intermediaries can help to reduce knowledge asymmetries between experts and non-experts and provide non-experts with the higher-order evidence they need to assess the trustworthiness of putative experts and their claims. On the other hand, intermediaries can contribute to augmenting knowledge asymmetries between experts and non-experts by not reporting scientists' first-order claims or giving these issues minimal coverage; misrepresenting experts' testimonies; failing to provide adequate higher-order evidence for non-experts to make good judgements about the trustworthiness of putative experts; failing to present higher-order

evidence about experts with integrity; or otherwise interfering with non-experts' reflective capacities, perhaps by exploiting their cognitive biases by intentionally framing information in a particular way. These problems are all the worse when the sort of higher-order evidence non-experts need to assess the trustworthiness of the intermediaries themselves is also not made available.

Therefore, when non-experts lack adequate access to experts' testimonies, the science-based information that should be made available to them, and pertinent higher-order evidence, they can be pre-emptively silenced. To demonstrate, let's return to the case of Tennille from chapter three, who has decided to give up smoking and has asked her doctor for advice on the range of options available to help her quit smoking (see Tennille 2, p. 27).

| SCENARIO ONE<br>TENNILLE'S DOCTOR MISIN-<br>FORMS HER | SCENARIO TWO<br>TENNILLE'S DOCTOR <u>DOES NOT</u><br>MISINFORM HER |
|---|---|
| **Tennille:** 'I've decided to quit smoking and I'd like some advice on the best way to go about doing so. Is there anything available to help me?' | **Tennille:** 'I've decided to quit smoking and I'd like some advice on the best way to go about doing so. Is there anything available to help me?' |
| **Doctor:** 'There are four main treatments available on the national health service. First, there is "M", which we've had stunning success with. Then there is therapy "N", medication "O", and medication "P". The evidence suggests that "M" is — by a long shot — the most effective product available to help people stop smoking.' | **Doctor:** 'There are three main treatments available on the national health service: therapy "N"; medication "O"; and medication "P". The current evidence suggests that the only strategies that have performed better than chance are O and P. We've found that P has been the most efficacious strategy for women of your age.' |
| **Tennille:** 'Ok, great. Are there any side-effects or risks associated with taking "M"?' | **Tennille:** 'Are there any side-effects associated with taking "P"?' |

Doctor: 'No... apart from a dry mouth.'

Doctor: 'Yes... there are several, such as nausea, insomnia, and drowsiness. There have also been a couple of recent observational studies in Sweden and the U.S. that have found P to be associated with a small increase in the risk of psychosis in people with pre-existing psychiatric disorders. However, at this stage, the empirical evidence is still very limited.'

Tennille: 'Could you please prescribe me a course of "M" then?'

Tennille: 'I have actually previously been diagnosed with a psychiatric illness. I value my mental health above all else so I'm unwilling to risk it in any way. I think I'll go with medication O then.'

Doctor: 'Certainly. In fact, I happen to have some here you can have.'

Doctor: 'Certainly. I'll write you out a prescription.'

Just like in the earlier *Sliding Doors* example, the conversations within the two possible worlds are very different. Again, the words uttered, and the contents of the propositions expressed, are noticeably dissimilar. Likewise, Tennille also performed very different speech acts in the two scenarios in response to her doctor's own speech and speech acts. In the second scenario, Tennille issues at least one verdictive that she's unable to in the first: that, all-things-considered, she prefers option O. Therefore, in the first scenario, Tennille was pre-emptively silenced by her doctor's speech. His speech in this scenario conveyed a picture of the world to Tennille that was a *misrepresentation* of how the world in fact is, and this misrepresentation led Tennille to form false and inaccurate beliefs about the world she inhabited (i.e. that M was the most efficacious option, that M had no side effects, etc.). Thus, by controlling what Tennille believed, her doctor gained a significant degree of control over Tennille's internal reflections, the deliberative conclusions she reached, *and her future speech*. It should be concerning to us when someone gains control over another person's future speech, because by doing so, as I have explained here, they simultaneously gain control over some of her future actions, since we do things with words (Austin 1955/1962). I contend that pre-emptive silencing is, therefore, a key mechanism used to manipulate people's behaviour, and one that does not seem to have been identified and appreciated as such to date.

Tennille's socio-epistemic environment in relation to this matter would fail at least the first two epistemic criteria I posited for voluntarily formed beliefs in chapter three:

(i) Adequate information or evidence was not available to Tennille;

(ii) Relevant information was significantly misrepresented, and Tennille's doctor misleadingly presented pertinent evidence (e.g. by lying, misleading or otherwise deceiving Tennille);

So, when Tennille asks her doctor to prescribe her M, she does not *voluntarily* say these words. Her doctor knows that by not providing Tennille with an adequate opportunity to form true/accurate beliefs to input into her deliberative processes — again, by ensuring Tennille was 'informationally cut off' from the necessary and sufficient information and evidence — that he can *manufacture* Tennille's doxastic states, her choice of M as

her preferred smoking cessation strategy, and her speech requesting that it be pre-scribed.

If instead, like in scenario two, Tennille's doctor was honest and diligent and thus pro-vided Tennille with the pertinent information and evidence she needed, Tennille would have had what she needed to reflect freely and to voluntarily form thoughts and doxastic states that were meaningfully *hers*. When she subsequently spoke, as she did in scenario two, she was able to speak *her* mind because she was not pre-emptively silenced by a deprived and distorted epistemic environment constructed by others' speech.

Since misinformation about science, as well as omissions of pertinent information and evidence, is currently a common feature of important parts of information systems (e.g. Oreskes and Conway 2010; Proctor and Schiebinger 2008), it seems that pre-emptive silencing is likely to be a pervasive phenomenon in many contemporary societies. As just one example, Furman (2016, forthcoming) describes the situation internally within South Africa's ruling party during the AIDS crisis in the early 2000's. She argues that the dissent of then President, Thabo Mbeki, about the cause of AIDS and the efficacy of anti-retroviral (ARV) treatment, suppressed other voices (other dissent) within his party over policy. One way to understand Furman's claim is that these members were silenced in the direct silencing way of being silenced: they feared that disagreeing with Mbeki might ruin their political careers, for example. But another way to understand Furman's claim was that Mbeki *pre-emptively silenced* his colleagues by acting in ways that distorted the epistemic landscape, such as by loading his Presidential Advisory Panel[41] on AIDS with an unrepresentative array of putative scientific experts on HIV-AIDS.

This in turn means it's plausible that there would be different actions — and, ultimately, outcomes — in a counter-factual world in which misinformation and misrepresentations were absent or reduced. Of course, this all depends on the degree of robustness between worlds, as some people's beliefs, preferences, and actions — including speech-related

---

[41] Approximately 50% of members of this panel rejected or did not accept that (i) HIV caused AIDS and/or that (ii) ARVs could effectively suppress the HIV virus and check the progression of the disease into full blown AIDS.

actions — will remain the same regardless of any changes within their epistemic environ-ments. Similarly, even if conversations differed between possible worlds, it could often be the case that the outcome would remain unchanged. I will talk more about these issues, as well as how we can determine under what circumstances the risk of pre-emptive silencing should concern us, in the final section. However, before we do so, I'd like to explore the implications of pre-emptive silencing for people's free speech rights.

## 3.4. On Violating the Right to Free Speech

Philosophers have argued that, under certain circumstances, silencing amounts to a vi-olation of a person's free speech rights. Feminist philosophers have argued that this includes those cases that involve mistakes on the part of the hearer. As Finlayson (2014) underscores, if these arguments are successful, they manage to turn the tables on tradi-tional liberal defences of certain speech; if some speech (i.e. speech that silences) violates others' rights to freedom of speech, that speech can no longer be defended simply on the grounds of the alleged fundamental or overriding importance of the right to free-dom of speech. As we are left with a conflict of rights, and moreover, a conflict of the *same* rights between different parties, the question then becomes one of whose right to freedom of speech ought to take precedence.

But one might wonder how this could be so if silencing phenomena do not involve limitations or restrictions on speakers as one might think free speech violations do. With regard to pre-emptive silencing, it doesn't look like the sort of barriers to speech that are normally associated with free speech violations, such as threats of imprisonment or physical harm, are present. We therefore need a better explanation of how pre-emp-tive silencing might plausibly be regarded as a free speech violation. But first, we must clarify what the right to freedom of speech amounts to.

### 3.4.1. Extant Conceptions of the Right to Freedom of Speech

To clarify the scope, extent, and structure of what a right to free speech is or should be, it's helpful to first get a better understanding of what a right to free speech is not, or

should not be. Obviously, the right to free speech does not mean that a person is free to say absolutely anything, anywhere. Lots of speech is regulated (e.g. libellous statements, child pornography, some advertising content, state secrets, insider trading, defamation, contracts, certain forms of hate speech in some countries, etc.), and regulating speech in these ways is widely regarded as being consistent with a commitment to free speech. Schauer (2014) calls these instances of speech *patently uncovered speech*: communication that warrants no special protection against government regulation. Moreover, even speech that is not regulated directly can be regulated indirectly through time, place, and manner restrictions. Every society places some limits on speech, both formally through legislation and informally through social and professional norms, because speech always takes place within a social and political context of competing interests, values, and rights. If two instances of speech clash (e.g. when a protest disrupts and prevents a lecture) a decision must be made about whose speech ought to be prioritised in that context. This means that there can be no *unlimited* right to free speech, and it is for this reason that Stanley Fish (1994) asserted that there is *no such thing* as free speech.

West (2003) suggests that the right to freedom of speech is therefore best thought of as a spectrum of different possible degrees and types of restrictions on speech. At one extreme, West says, is a conception of the right to freedom of speech that merely protects a speaker's negative freedom in expressing words in private. It doesn't protect speakers from threats or severe penalties for expressing these same sentiments in public or quasi-public settings, nor does it provide them with a right to have those words heard by an audience, nor for their uptake to be secured. On this understanding of the right to freedom of speech, government censors and totalitarian thought police would not violate a speaker's right to freedom of speech. I doubt that anyone outside of an authoritarian regime would conclude that this narrow conception of the right to freedom of speech was an adequate account of such a right.

West notes that at the other end of the spectrum lies a very broad conception of the right to free speech which seems equally implausible, at least as a general account of the right to free speech. On this account, the public distribution of one's speech is guaranteed rather than merely permitted. A speaker's right to free speech on this account is violated when their words fail to receive wide public distribution and uptake, and for

whatever reason, whether that be because the speaker is afraid of public speaking or illiterate, or because potential hearers choose not to read their ideas or walk out of lectures.

Given this continuum of possible understandings of the right to free speech, the question is where on that continuum we should draw the line (or where we should draw various lines if we think this line might vary with context).[42] Minimally, freedom of speech should allow speakers to express themselves as they please in at least some forums (e.g. their own homes, public places), and ensure that they are not prevented from speaking by severe state threats (such as imprisonment). And perhaps there will be some limited circumstances in which something like the maximal conception of the right to freedom of speech will be guaranteed—e.g. for a defendant at a criminal trial—but this generally won't be the case. In most circumstances, we will not take the right to free speech to be equivalent to the broad public distribution and widespread comprehension of a person's speech.

This is partly because it's not feasible for people to be given such a right; there are scarce resources (e.g. only so many minutes in public meetings, column inches in newspapers, slots at conferences, etc.), many of which the state does not control, and that it has no place in dictating how these resources should be allocated. It's also wouldn't be appropriate to give people such a general right, because doing so would give rise to extremely burdensome duties that other agents would be obligated to fulfil, but which would inevitably unreasonably interfere with others' freedoms of speech, especially those of the press, as well as other important rights and freedoms, such as freedom of association, the freedom to spend our time as we please, etc.

We therefore need to draw a line (or lines) somewhere between these two extremes. In doing so we need to protect individuals' freedom to communicate with others, but not at the cost of imposing unacceptable burdens on others, or by unjustifiably infringing other rights and interests. But exactly where should we draw the line(s)? What barriers

---

[42] This is the approach that I ultimately prefer. To my knowledge such an approach has only been outlined to date in a non-peer reviewed article by philosopher Robert Simpson in *Aeon* (2017), "Free Speech' is a Blunt Instrument. Let's Break It Up': https://aeon.co/ideas/free-speech-is-a-blunt-instrument-lets-break-it-up

on speech are permitted, and who is allowed to erect them? Which barriers are justifiably considered impermissible? Two general lines have been proposed within the literature to date: the traditional liberal and the feminist conceptions of the right to freedom of speech.

**The Traditional Liberal Conception of Freedom of Speech**

Scholars such as Thomas Scanlon (1972), Robert Post (1990), Ronald Dworkin (1991), Daniel Jacobson (1993), and more recently Eric Heinz (2016) defend what I call here the traditional liberal conception of free speech. This understanding of the right to freedom of speech assumes that people are free to speak just so long as they are not prevented from producing sounds and scrawls that others are not prevented from hearing or seeing. For these scholars, if someone is able to complete a locutionary act — that is, they are free of 'literal obstacles like gagging, drowning out, or arrest by the police' (Hornsby 1994: 140) — freedom of speech is satisfied. Many liberals also recognise actual or threatened interferences by non-governmental agents as actions that may also violate a speaker's right to freedom of speech.

The traditional liberal position is also that the right to free speech in itself does not require a receptive audience. Although an unreceptive audience does in some way 'prevent' a speaker from communicating, this kind of prevention does not amount to a violation of the speaker's right to free speech for those who endorse the traditional liberal conception of free speech. Those who subscribe to the traditional liberal conception hold that it should not amount to a rights violation because to accept it as one would permit the imposition of a range of demanding duties on individuals to attend to others' speech, as mentioned above: duties to not to avoid demonstrations, not to tune out when someone is speaking to you, not to walk out of events, and so on (Dworkin 1993, Green 1998).

**The Feminist Conception of Freedom of Speech**

However, feminist philosophers and legal scholars have argued that these conditions are insufficient for freedom of speech, even by libertarians' and liberals' own lights (Brison 1998, 2000; Hornsby and Langton 1998; Langton 1999; West 2003). As a challenge to

the traditional liberal conception of free speech, Caroline West (2003) asks us to imagine a government that allowed people the freedom to produce whatever sounds and scrawls they like, but that implanted meaning-scramblers in the heads of hearers that systematically prevented would-be hearers from comprehending the intended meaning of those locutions. For example, every time a citizen heard an utterance that was critical of government, the device would cause any listener to hear it as praise of the government instead. West argues that such a scenario would be at least equally as bad as that in which a government prevented speakers from producing sounds and scrawls altogether. Under both scenarios, speakers are prevented from communicating with others, which defeats what liberals take to be among the values of free speech in the first place: the right of speakers not to be prevented by the actions of other agents from communicating their ideas or opinions to those who might wish to hear them. West concludes that the speaker's free speech rights should count as being violated under both scenarios.

However, despite this, a right to free speech cannot simply be a right to communicate and to have one's words listened to and fully understood by others. And the feminist conception of free speech definitely doesn't guarantee communicative success in all cases. If I consistently fail to communicate instructions to my cat, my right to free speech has not been violated. My freedom of speech rights, even under the feminist conception, do not guarantee communicative success under all the circumstances I desire it inSuch rights do not permit me to speak anywhere, or at any time that I wish to. They also don't impose a general duty on others to provide me with a communicative outlet for my speech. Rather, a feminist conception of the right to free speech stipulates that I have right to be free from certain sorts of *systematic* or *systemic* barriers to, and interferences with, successful communication. Here 'successful communication' can be understood as something like a hearer being able to accurately interpret the propositions I express. By systematic, I mean barriers that have been purposefully and methodologically erected according to a concealed plan (e.g. in the case of Sliding Doors), or that have been brought about as a result of a widescale failing of epistemic responsibilities due to negligence or recklessness. By systemic I mean barriers that arise merely because one is a member of a particular demographic group. This is the conception of the right to free speech endorsed by feminist free speech theorists like MacKinnon, Langton, West, Hornsby, Maitra and McGowan.

### 3.4.2. Communication-Based Conception of the Right to Free Speech

Although I think the feminist conception of the right to free speech is a great improvement on that of traditional liberals, I argue here that it still fails to track, account for, and protect an important value that ought to ground an understanding and justification of what a right to freedom of speech entails. This, I think, is because the feminist conception of freedom of speech, like that of traditional liberals, is exclusively a *speaker-based account* of the right to free speech. It therefore neglects our *listener-based interests* in free speech, i.e. our interests in having ready access to the speech of others in our society. What West's thought experiment brings into sharp relief is, I think, that speech is valued *because* of its role in *communication*, where communication refers to sending *and receiving* information. Moreover, West's thought experiment also highlights the fact that the traditional liberal conception of free speech, and our right to it, fails to pay sufficient attention to the way that language actually works. In particular, it fails to attend to the way in which what words mean — and so what it is possible for speakers to say or communicate in turn — depends on the social context, and that this context is significantly shaped by speech.

Simpson (2015) suggests that we can describe the 'social context' of language by referring to the community or communities inhabited by speakers and listeners. As he points out, the words 'communication' and 'community' are etymologically descended from Latin terms that refer to the notion of things being shared or held in common. Simpson (2015) thus urges us to view these two concepts in ways that emphasise their connections to one another:

> Community might be conceived of as a state of human affairs defined by the establishment or maintenance of reliable channels of interpersonal communication. Communication might be understood, in turn, as encompassing forms of idea-transmission that express or conduce to communal relations with others. One consequence of endorsing these kinds of conceptual links is that we'll be unable to say much about the ethics or politics of communication without first thinking hard about how our speech affects (and effects) our community. Claims about what we should and shouldn't say, and about what are appropriate reactions to communication (one's own, or others), will to

at least some degree hinge on our views about what constitutes a good and/or just 'common life'.[43]

Legal philosopher Seana Shiffrin (2014) has perhaps done the most to bring the importance of this relationship into sharper relief. In developing a 'thinker-based approach' to free speech that encompasses both our listener- and speaker-based interests, she stipulates that our listener-based interests relate to what conditions are needed to accommodate the meaningful exercise of our capacities *qua* thinking beings. Among these conditions is the integrity of the processes and channels by which we reliably receive information from others. Should these processes be manipulated, hijacked or distorted, or transmission channels come to be blocked, Shiffrin thinks our free speech rights ought to be considered to have been violated (Shiffrin 2014, ch. 4).

What Shiffrin overlooks, but is important here, is that our listener-based interests are not entirely separate from our speaker-based interests, but rather entangled with them. Not only do we need access to other peoples' speech literally to be able to learn to speak, more importantly, we need access to other peoples' speech in order to speak our *own minds*. We use others' ideas and statements as essential building blocks and scaffolding for our own. When we receive others' ideas or statements, we evaluate them and either endorse, reject, or modify them before incorporating them into our doxastic web and, in turn, externalising them by expressing them out into the world. And speaking out loud is, in turn, also part of the process by we form mental content of our own. A well-grounded and properly justified conception of the right to free speech must therefore account for these kinds of interests too.

However, just as a right to free speech can't be a right to be free from interferences in speaking under all circumstances, a right to free speech also can't be a right to be free from interferences in properly *receiving* others' communications in all circumstances either. Therefore, I propose that we conceive of this aspect of our free speech right as the

---

[43] Interestingly, when Isaiah Berlin talks of negative liberty, he gestures to a 'basic nucleus of needs and goals' and suggests that the boundaries of this nucleus are 'determined by the basic need to communicate with other similar beings' (1969: lxii).

right to be free from certain sorts of systematic or systemic barriers to successfully receiving others' communications. This, you will notice, is the mirror image of feminist philosophers' existing conception of the right to free speech, which I noted was based on the speaker's interests. I suggest that the speaker- and listener-based aspects of the right to freedom of speech can be unified simply by understanding communication here more broadly, as an iterative and two-way, rather than a once off, one-way, process. The right to freedom of speech then simply becomes: *the right to be free from systematic and systemic interferences in successful communication.*[44]

### 3.4.3. Pre-Emptive Silencing as a Free Speech Violation

With this understanding of the right to free speech established, i.e., *the right to be free from systematic and systemic interferences in successful communication,* I now turn to considering whether pre-emptive silencing constitutes a free speech violation, and if so, under what circumstances. I'll first briefly review the reasons why direct silencing and indirect (illocutionary) silencing have been argued to be violations of people's rights to freedom of speech before presenting my argument for why we ought, in addition, to consider pre-emptive silencing to sometimes be a free speech violation too.

Recall that direct silencing occurs when no locutionary acts are performed. These circumstances occur when, in response to the threatened consequences of expressing oneself, people self-censor and refrain from speaking. I suggested that a paradigm example of this type of silencing was Thailand's *lèse-majesté* law. Libertarians and liberals would classify this this law as an infringement of freedom of speech because it interferes with a person's ability to make locutions. It interferes with a person's negative liberty because the risks and costs associated with speaking — i.e. imprisonment — are much too high for someone to count as 'free'.

---

[44] A similar understanding of this right has been put forward by McGowan (2014). My conception differs from hers slightly by encompassing what I have defined as systematic, as well as systemic, interferences. This enables me to account for the way that certain types of deception impact on others' free speech rights, whereas McGowan is focused on the consequences of socially patterned power dynamics.

However, as previously mentioned, many feminist philosophers take this way of conceiving of the right to free speech to be much too narrow; they claim that the right to free speech, properly understood, also extends to illocutionary acts. As I explained, illocutionary silencing involves the addressee's failure to recognise the speaker's illocutionary intention, where this failure has systemic causes. This type of silencing involves communicative failure precisely because it involves an uptake failure on behalf of the listener. Because illocutionary silencing systematically blocks people's ability to successfully communicate with others, it has been argued that the speech that causes systematic illocutionary silencing (e.g. certain types of pornography or certain types of hate speech) denies people's rights to freedom of speech (Hornsby and Langton 1998, Matsuda et al. 2018, McGowan 2014, West 2003).

I also want to claim that, at least under some conditions, speech that causes pre-emptive silencing can violate other people's free speech rights. However, it does so for different reasons than the other types of silencing surveyed. When speech about scientific matters pre-emptively silences, it can undermine the free speech rights of others in three key ways. First, if experts' testimony is misrepresented and distorted by an intermediary, it interferes with experts' free speech rights in a similar way to the way that women's free speech rights are interfered with by pornography. Under both cases, successful communication between relevant parties is compromised by the speaker's message to listeners coming to be scrambled in some way by a third party. In the case of experts transmitting testimony to non-experts, experts' speech can be scrambled by intermediaries' speech in a direct, literal, sense when an intermediary straightforwardly mis-reports experts' testimony. Alternatively, experts' testimony can be scrambled indirectly when intermediaries cast unsubstantiated and unjustified doubt and confusion over the epistemic landscape, e.g. when they misrepresent higher-order evidence.

Such interferences in the channels of communication between experts and non-experts frustrate non-experts' listener-based interests in receiving experts' testimonies unadulterated. There are two key aspects of this listener-based interest. First, non-experts have an interest in receiving experts' claims undistorted so that *they themselves* can make a decision as to the experts' trustworthiness and decide whether to defer to their claims. Second, as previously mentioned, this is important because we use others' testimony as

building blocks and scaffolding for our own thoughts. This represents the second way peoples' free speech rights — understood as the broader, communication-based conception I outlined in §4.4.2 — can be interfered with.

The third way free speech rights can be undermined falls out of the consequences of these kind of interferences, as such interferences simultaneously frustrate a non-expert's own speaker-based interest in successfully communicating their *own* minds to others. When matters are scientific or technical, or when they significantly turn on specialist information, a necessary condition for speaking our own minds in relation to these matters is that our access to experts' testimonies are not disrupted or distorted. If this channel is interfered with in a way that cuts off or scrambles experts' testimonies, then I will be unable to speak my *own* mind on this matter.

At this point, I might be accused of confusing negative and positive liberty. Ronald Dworkin's (1993) objection to Langton's account of silencing was that it rested on a 'dangerous confusion' between negative and positive liberty. In Isaiah Berlin's famous essay on this distinction, he initially defined negative liberty as *freedom from*, that is, the absence of constraints on the agent imposed by other people (Berlin 1958/1969). He uses a metaphor to capture what he means by this:

'The extent of a man's negative liberty, is, as it were, a function of what doors, and how many, are open to him; upon what prospects they open; and how open they are.' (1969: lviii)

He defined positive liberty both as *freedom to*, that is, the ability (and not just the opportunity) to pursue and achieve willed goals; and also as *autonomy or self-rule*, as opposed to dependence on others.

Dworkin's position was that *if* pornography silences (which he doubted), it does so by depriving women of positive, rather than negative liberties: it 'denies them the right to be their own masters' (Ibid: 106). This argument might be levelled at my account of pre-emptive silencing too. Specifically, it might be argued that *if* certain instances of dissent about science — e.g. misinformation, or misrepresentations of scientists' testimonies or higher-order evidence — silences, it does so by depriving non-experts of positive rather than negative liberties: it denies them the right to have the ability to think certain

thoughts and speak certain things (rather than just the opportunity, which they retain), and/or to be their own masters.[45]

Langton, and other feminist scholars since, have strongly resisted this objection. In their replies, they have tried to find new ways of explaining how silencing effects women's negative speech-based liberties, hoping to finally get those who disagree with them to see what they see. But rather than resist this objection, I accept the observation at its core: that some speech curtails non-experts' positive speech-based liberties is precisely *part* of what I am arguing, as I will explain momentarily. But before I do, I want to briefly deviate from the main discussion to offer reasons for why we should not shy away from acknowledging the impacts upon positive liberties; the importance of positive freedoms, especially in the domain of speech, has been neglected in this debate for too long.

Firstly, the view that we should prioritise negative over positive liberties today lacks good justification. Many of those who hold such a view cite Berlin's 'warning' about the to-talitarian tendencies of positive liberty. But in the introduction to *Four Essays on Liberty*, Berlin concedes that he didn't initially pay sufficient attention to, or emphasise the pernicious tendencies of negative liberty enough, in his original essay. He reflects that he:

> should perhaps have stressed (save that I thought this too obvious to need saying) the failure of such systems [i.e. laissez fair capitalism] to provide the minimum conditions in which alone any degree of negative liberty can be exercised by individuals or groups, and without which there is little or no value to those who theoretically possess it. For what are rights without the power to implement them? (1969: xlv)

He goes on to call such situations an 'odious mockery'. So here it seems that Berlin acknowledges the importance of both forms of liberty, and their interrelationship. He agrees that both negative and positive liberty are ends in themselves (Ibid: xlix), and that 'The case for intervention, by the state or other effective agencies, to secure conditions for both positive, and at least a minimum degree of negative, liberty for individuals, is overwhelmingly strong' (Ibid: xx).

---

[45] Note that it is the former I argue for in this chapter. I argued for the latter in chapter three on false beliefs and personal autonomy.

Berlin goes on to note that:

> [T]he rise and fall of the two concepts can largely be traced to the specific dangers which, at a given moment, threatened a group or society the most: on the one hand, excessive control and interference, or, on the other, an uncontrolled 'market' economy. Each concept seems liable to perversion into the very vice which it was created to resist. (Ibid: xx)

At the time Berlin initially penned *Two Concepts of Liberty* in the mid-20th century, the general threats were to negative liberties – hence his focus on protecting it and exposing aberrations of positive rather than negative liberty. But arguably, the tables have now turned; the notion of negative liberty has been perverted into its opposite, and it is our positive liberties that demand attention and protection.

So, when I say that some dissenting speech can pre-emptively silence non-experts, I accept that one aspect of this claim is that a positive, rather than negative liberty, is being curtailed. There is a sense in which, when someone is 'prevented' from speaking certain things as I am claiming they are, that they technically retain their negative liberty to say or write these words. Using Berlin's metaphor, these doors remained open to non-experts. What misinformation and misrepresentations did was simply change the apparent *rationality* of going through one door rather than another, i.e. saying 'X' rather than 'Y'. But it did not close either door. Absent dissent that misinforms and misrepresents, non-experts would have better capacity to say what they want to say, but that is a positive, rather than negative liberty.

But there is also an alternative reading of the situation. Under this interpretation, misinformation and misrepresentations narrow a non-expert's field of view in a way that means they cannot actually see that there are many more (open) doors to them. Perhaps misinformation and misrepresentations had the effect of throwing invisibility cloaks over these doors and concealing them, leaving non-experts to perceive there to be fewer doors in the room that there in fact are. A non-expert in this situation might even 'feel' very free because they're able to walk through several doors and make a number of speech acts. However, what they don't realise is that they're simply walking through doors all within the one large room; little do they know that the doors they currently can't perceive actually lead to several more rooms with lots more doors, some of which

lead to an outside world with yet more doors, and so on. If this is the effect some dissent about science has on non-experts, then it does in fact infringe upon non-experts' negative speech-based liberties, because these opportunities have been blocked. The barrier isn't a physical one, or as literal as those barriers accepted by libertarians and traditional liberals to block free speech, but it is a barrier all the same. Berlin himself even spoke of the 'rigging of the means of influencing opinion'. He goes on to say that this often has had the effect of blocking and diminishing human freedom as effectively as the more overt and brutal methods of direct oppression — slavery and imprisonment — against which the original defenders of liberty lifted their voices (ibid: xlviii).

While I have predominantly focused on non-experts' liberties, it's also important not to forget that some dissent can compromise experts' negative liberties too, if it has the effect of systematically blocking experts' ability to carry out particular (illocutionary) speech acts, such as informing non-experts.

A similar objection to the one I have just explored would be to agree that the people who are pre-emptively silenced have been wronged, but to disagree that they have been wronged because their free speech rights have been violated. Daniel Jacobson (1995) raises this objection to feminist philosophers' arguments about pornography and illocutionary silencing. Although he agrees that a woman who is unable to refuse sex does suffer a moral wrong, he does not think there is any need to appeal to her right to free speech to explain what that wrong is. Instead, he says:

> What is so terrible about a woman's being unable to refuse sex is the disablement of her autonomy, the resulting violation of her body, and assault on her well-being (Jacobson 1995: 76).

One might object to my argument along similar lines and say that there isn't any need for me to appeal to rights to freedoms of speech to explain what's wrong: non-experts are wronged by certain dissenting speech about science because it can disable their autonomy (as I argued in Chapter Three), or because their freedom of thought rather than speech is interfered with.

I think this objection is too quick to dismiss the possibility that the silenced person was wronged in *all* of these ways, including with respect to their right to free speech. Such

an objection also overlooks the possibility that our autonomy and freedom of thought can be intimately tied up with the integrity of our free speech rights, properly defined, and vice versa. What my argument ultimately highlights and endorses is J.S. Mill's thesis in *On Liberty* that freedoms of thought and speech are *practically inseparable* (1859; Nash and Simpson under review). We cannot, when it comes to speech *in practice*, distinguish between violations of freedoms of thought and speech even if we can in theory. In fact, Mill was interested in the inseparability thesis because he wants free speech to be upheld *for the sake of* freedom of thought; in Lucas Swaine's terms (2016), thought is the basic liberty in the relation. The two freedoms may be practically inseparable, but freedom of thought is analytically distinct from — and in justificatory terms, more fundamental than — free speech, as well as other core liberties (e.g. conscience, association). So in this sense, if we should prioritise anything, it should be our freedom of thought and its preconditions.

To summarise: I am indeed concerned that some speech silences others' speech in a way that means there is less speech, and less diversity of speech, overall, and that as a consequence, the Millian arena is deprived of ideas. And I hold that the mechanism for such silencing operates primarily via interferences in peoples' freedom of thought. In blocking or scrambling the channels of communication between experts and non-experts, non-experts are prevented from using experts' testimony as building blocks to construct other thoughts, and therefore to subsequently express other assertions and ideas. This can be construed as a type of positive freedom. But I am *also* arguing that certain speech pre-emptively silences others by depriving them of negative liberties to perform important speech and speech acts. Pre-emptive silencing goes some way to explaining why freedom of speech, as it has traditionally conceived by liberal orthodoxy, is a hollow privilege for victims of indoctrination, oppressive socialisation, or a poverty of epistemic resources — people who are only ever 'parroting the views of someone else hidden behind the scenes' (Gilmore 2011: 536). Whilst parroted communication would indeed represent 'successful communication' from the standpoint of those who indoctrinate and engage in pernicious forms of socialisation, it cannot be considered successful communication from the standpoint of those who speak. Nor will it be considered ethical communication by those who are affected, in turn, by such speech. Successful commu-

nication must refer to those locutions and other speech acts that are sufficiently a person's own. A person's free speech rights are thus compromised in this sense where such interferences occur.

## 3.5. Objections to Pre-emptive Silencing (& Replies)

In this final section, I consider and reply to two general objections to my account of pre-emptive silencing and of it being, under some circumstances, a violation of freedom of speech rights.

### 3.5.1. Objection one: pre-emptive silencing is an empty concept

One might be concerned that pre-emptive silencing is a ubiquitous and seemingly unavoidable phenomenon, and therefore, that it is ultimately an empty concept. What one person says is of course always going to be contingent on others' speech, and in that sense, there are always going to be other things a person could have said and done with their words. In replying to Langton, Dworkin (1993: 38) says: 'Creationists, flat-earthers, and bigots, for example, are ridiculed in many parts of America now; that ridicule undoubtedly dampens the enthusiasm that many of them have for speaking out and limits the attention others pay to what they have to say.' Yet surely, says Dworkin, we should not think that that this violates their right to freedom of speech.

In reply, I want to emphasise that not all communication is on a par, and therefore, neither are all communication successes, failures, alternatives, or omissions. I suggest that free speech concerns with regard to pre-emptive silencing should be limited to only those varieties of the phenomenon that can be reasonably anticipated to be morally and politically problematic. Given that adequate communication of expert testimony is an important aspect of a social division of epistemic labour that augments autonomous agency (as I argued in Chapter Three), misrepresentations of experts' testimonies by third parties seems to be an important type of communicative failure that can cause morally and politically problematic pre-emptive silencing; it is thus a plausible candidate for concern.

So too with speech and omissions of communication in intimate relations. It is the potential for information asymmetries, combined with the high-stakes of the situation, that makes pre-emptive silencing in these situations morally problematic. This explains why we already have strong norms of honesty in our intimate relations. By contrast, there is a sense in which my friend deprives me of an opportunity to say 'Could I please come too?' when she initially declares that she's not going on a holiday this year, but then later books a trip without telling me. But this isn't a morally or politically problematic type of pre-emptive silencing because my friend doesn't owe me that information, nor the opportunity to speak those words, morally speaking.

As well as limiting our concern to varieties of pre-emptive silencing that can be reasonably anticipated to be morally and politically relevant, we could also vary our concern about people's free speech rights by the degree to which communicative interferences are unlikely to result in different states of affairs regarding moral or political outcomes between possible worlds. I take it that one of the reasons why many people are so concerned about the widespread transmission of misinformation about HIV-AIDS, climate change, vaccinations, and the like, is because it seems plausible that, absent such misrepresentations and misinformation, a very different world would have emerged with potentially vastly different moral and political outcomes (I discuss this point in more detail in the next chapter). To the contrary, given that (1) a person can accept evolutionary science *and* be a theist (Weisberg et al. 2018) and that (2) many medical professionals who don't accept evolution can still perform their professional roles (Everhart and Hameed 2013), it's not clear that pre-emptive silencing associated with evolutionary science has the same sort of moral and political risks and harms attached to it.

### 3.5.2. Objection two: the 'slippery slope' objection

Finally, I anticipate that many will be worried that by recognising the phenomenon of pre-emptive silencing and its harm to freedom of speech, we will inevitably slide down a slippery slope to the sort of political correctness that leads to extreme self-silencing, and then to pernicious varieties of state-based speech regulations and censorship. Whilst these concerns are both reasonable and important, by way of reply it's worth drawing

attention to a response to slippery slope arguments made by Frederick Schauer (1985). Schauer suggests that slippery slope arguments are misleading because they implicitly suggest that we can be on or off the slope when no such choice in fact exists: we are *inescapably* on the slope whether we like it or not. The task, Schauer says, is always to decide how far up or down the slope we should choose to go, and both directions (i.e. towards more regulation or less regulation) can be slippery. Both directions can also lead to totalitarian-like outcomes, whether it be at the hands of the state (e.g. in terms of over-regulation) or non-state actors (e.g. in situations of under-regulation). It's false to think that certain places on the slope are completely riskless or harmless, either to freedom of speech or other values. Once we abandon the incoherent position that there should be no limits on speech, it's clear that we will always have to make controversial judgements about what should and should not be permitted to be expressed, where, and under what conditions; but there are no 'invisible hands' with the 'marketplace' of ideas.

Fears about 'slippery slopes' are also generally based on mistaken understandings of how the regulation of speech works in practice. By identifying the phenomenon of pre-emptive silencing, and by arguing that it is a violation of the right to freedom of speech, I have identified a kind of disabled speech (or potentially disabled speech) and located a new speech-related harm (or wrongful risk of harm). What I *have not* done is argued that in response to this phenomenon any speech should be regulated by formal or informal means. At most I have argued that there is a *pro-tanto* reason to address the *source* of the disabled speech. But much more needs to be done to argue that, all-things-considered, some speech ought to be regulated, by either legislation or other means such as professional or social norms.

## Conclusion

In a reply to Berlin's original essay and his almost exclusive focus on the virtues of negative liberty, McFarlane (1966) makes the point that to know one's chains for what they are, instead of decking them with flowers, is often the first step to freedom, which may never come about if one either ignores their chains or loves them. In this spirit, this

chapter has identified and analysed a new freedom-constraining silencing phenomenon, that of pre-emptive silencing. It has also shown that we should have just as much reason to regard it as a free speech violation as we do to regard other types of silencing as such. Moreover, we have reasons to believe that pre-emptive silencing is a widespread phenomenon, especially in the context of scientific expert-intermediary-non-expert relations. Since it constitutes a speech-related harm, or at best the wrongful imposition of risk of harm, we have excellent reasons to attend to pre-emptive silencing in both theory and practice.

# CHAPTER FOUR:
# DISSENT, PUBLIC POLICYMAKING &
# DEMOCRATIC LEGITIMACY

*Introduction*

Numerous scholars involved in debates about dissent about science have cited the negative impact of problematic dissent on *public policy* as a key cause for concern. For instance, in *Merchants of Doubt*, Oreskes and Conway (2010: 297) state that the failure of the U.S. (and, implicitly, other states) 'to act on global warming and the long delays between when the science was settled and when we acted on tobacco, acid rain, and the ozone hole are prima facie empirical evidence that the doubt mongering worked.' Likewise, in *Science in a Democratic Society,* Philip Kitcher (2011: 243) asserts that: 'Climate deniers have successfully blocked attempts to introduce policies for coping with potential environmental change.' Many other academics since, from a wide variety of disciplines, have expressed concerns about the impact of some forms of dissent within or about science on public policy (e.g. Biddle and Leuschner 2015; de Melo-Martin and Intemann 2013, 2014, 2018; Ding et al. 2011; Furman 2016; Nash 2018; van der Linden et al. 2015).

Many of these same scholars also cite concerns about the negative impacts of some dissent on *democracy*[46] or *democratic values*.[47] Kitcher (2011: 243) characterises our inaction on anthropogenic climate change as a 'huge failure of worldwide democracy.' He claims that: 'When [voters'] electoral choices ignore the threats of climate change, they freely

---

[46] By democracy I mean those societies that use a process of collective decision-making characterised by some type of equality amongst all members of that society. I take democracy as a fixed starting point in this chapter (i.e. I don't interrogate the merits of democracy vis-à-vis other types of governance systems). One of this chapter's aims is, however, to contribute to debates about *what sort of democracy* we should favour.

[47] Intemann (2017) and Pearce et al. (2017) are also concerned about the 'democratic deficits' in policymaking associated with quarrels over empirical facts. However, they're both of the opinion that the problem isn't with the dissenting speech acts as such.

vote against their own interests. They are victims of the decay of democracy into statistical plutocracy' (2016: 13). Similarly, in a reply to Miriam Solomon's book *Social Empiricism* (2001), Oreskes (2008: 253) faults Solomon for ignoring the way that certain dissent can 'serve *undemocratic* political ends'. She reiterates this claim later in the article in a section that provides a descriptive account of the tobacco industry's misuse of scientific dissent: 'In short, dissent can be a potent political tool, and nurturing minority opinions does not necessarily advance science, *democracy*, or rationality' (2008: 261, my emphasis). It has been a common feature of Oreskes' work since to point to the impact of mis- and disinformation on democracy (e.g. see Oreskes and Conway 2010; Baker and Oreskes 2017a: 9, 2017b: 65).

In this chapter, I suggest that reservations about the impacts of problematic dissent on public policymaking and democracy are related, and best understood as concerns about the ability of some dissent about science to undermine the *democratic legitimacy* of public policy. Democratic legitimacy is a quality of political institutions, and of the decisions made through these institutions. It is what establishes, and qualifies, a democratic assembly's right to rule by defining the terms on which the members of that society should settle their disagreements about how to live their lives together. Democratic legitimacy can thus be understood as what *creates* political authority. If a political decision is made 'democratically', then we have a strong *pro tanto* reason to comply with it, or to refrain from interfering with its implementation (Kolodny 2014a: 197).[48] A unique feature of democratic societies is that this right to rule *is created by those who are ruled*. In democracies there are no sharp divisions between the 'binders' and the 'bound' (Hershovitz 2003: 210fn).

To understand the legitimacy, and hence political authority, of public policy made within democratic societies, we therefore need an account of the conditions that ensure citizens exercise coercive power on equal terms with each another (Peter 2008; Kolodny

---

[48] It is a further question as to how the democratic legitimacy of public policy should be weighed against other considerations, such as whether or not the policy is morally permissible or just. For example, even if a policy is democratically legitimate, it might still be a gross violation of human rights. In such a case, citizens might have moral obligations of civil disobedience, or perhaps even stronger obligations. I put these interesting and complex 'all-things-considered' decisions about the limits to democracy to one side for this chapter. Of course, Oreskes, Kitcher, et al. might want to make a claim that some climate policy is unjust, and so grossly unjust that it over-rides any value of democratic legitimacy. But this is generally not the how they have discussed the issue of dissent for policymaking.

2014a, 2014b). Whilst some have argued that that these conditions concern only the *procedural features* that shape the decisions made, others argue that legitimacy depends — either exclusively or in part — on *substantive features* realised by the decision, such as whether or not a policy is collectively rational or in some other way the 'correct', 'right' or 'true' decision.

In this chapter I argue that, thus far, concerns about the impacts of dissent about science on policymaking in democratic societies have implicitly involved a commitment to one of the latter conceptions of democratic legitimacy, i.e. those that depend, at least in part, on substantive features of outcomes. However, there are several issues with accounts of democratic legitimacy that turn on substantive features of outcomes. In particular, the assumption that there is a 'correct' policy outcome is controversial and, moreover, not well-suited to liberal democratic societies that endorse value pluralism and recognise the potential for reasonable disagreement about 'the good'. Instead, I adopt a version of the former, procedural accounts of democratic legitimacy — Fabienne Peter's *Pure Epistemic Proceduralism* — and show that even on an account that does not require a commitment to there being a correct outcome, some dissent about science can undermine democratic legitimacy. In doing so, I offer an alternative diagnosis of what goes wrong for public policymaking in democratic societies when problematic dissent about science flourishes in our public knowledge systems.

This chapter proceeds as follows. First, in §4.1., I provide an overview of how a number of high-profile science studies scholars, like Naomi Oreskes and Eric Conway, and Philip Kitcher, have characterised the issue of problematic dissent for public policy in democratic societies. I argue that we can best understand these scholars' concerns with reference to a strong version of a specific conception of democratic legitimacy known as *rational epistemic proceduralism*. In §4.2. I highlight the short-comings of rational epistemic proceduralism when it is used to understand what's wrong with problematic dissent for policymaking in democratic societies. In §4.3 I then introduce Fabienne Peter's account of democratic legitimacy, *pure epistemic proceduralism*, and in §4.4 explain how, on this account, some dissent about science erodes the democratic legitimacy of public policy by undercutting political and epistemic fairness.

## 4.1. Rational Epistemic Proceduralism

Philip Kitcher has argued that when problematic dissent distorts the Millian arena of ideas, public policy can come to be disconnected from the 'will of the people' (Kitcher 2016). He writes:

> The picture I have painted is one in which the citizens of the affluent nations that might collectively solve large global problems find themselves in a state of *false consciousness*, partly induced by the breakdown of genuine democracy, in which the *preferences they express at the ballot box fail to align with their interests* (Kitcher 2016: 11, my emphases).

Kitcher's claim seems to be that there is something irrational about many people's reasoning, either individually or collectively, because many people are failing to vote in ways that would best further their own 'true' interests.

Some even hold this to be the case for those people or collectives that are the most resistant to action on anthropogenic climate change. For example, the argument embedded within Oreskes and Conway's (2014) fictional account of the future, *The Collapse of Western Civilization*, is that by resisting action on climate change, many conservatives and free market fundamentalists are in fact acting in ways that are increasing the likelihood of the very long-term outcomes they fear — namely, a vast expansion of government and market regulation. In other words, Oreskes and Conway's implicit charge is that such people are failing to be rational, because the best means to their end of limiting the interference of government in the market over the longer term would be to concede some state-based action to address climate change now.

We could view the issues these scholars are raising here as mere paternalistic worries. But to do so would miss a deeper concern that I think they have about the current state of science-informed policymaking in democratic societies. I suggest that these scholars' concerns can be best understood as claims that some problematic speech acts about science, including some scientific dissent, has the potential to, and sometimes does, erode the *legitimacy* of democratic decision-making processes and the resulting public policy, where democratic legitimacy is understood (more or less) as it is expressed by the

family of accounts of democratic legitimacy that Fabienne Peter (2008) calls *Rational Epistemic Proceduralism*.[49]

Rational Epistemic Proceduralist accounts of democratic legitimacy consider *both* procedural and substantive features of the decision-making process and its outcome to be relevant to evaluations of the legitimacy of democratic decisions (Peter 2008). Rational Epistemic Proceduralists, however, differ amongst themselves with respect to how much weight they think should be attributed to each of these criteria. But most importantly for my argument, all Rational Epistemic Proceduralists hold the background assumption that there exists, independent of the decision-making procedure, a *true* or *correct* decision, i.e. an outcome that is the 'right' one. Joshua Cohen (1986: 34), for instance, in his influential account of epistemic democracy in *Ethics*, includes 'an independent standard of correct decisions' as one of the three elements of an epistemic interpretation of voting. Cohen (1986: 34, my emphasis) clarifies that 'independent standard' refers to 'an account of *justice* or the *common good* that is *independent of* current consensus and outcomes of votes.'[50] Similarly, as List and Goodin (2001: 280, my emphasis) make clear in describing epistemic democracy:

> The hallmark of the epistemic approach, in all its forms, is its fundamental premise that there exists *some procedure-independent fact of the matter as to what the best or right outcome is*. A pure epistemic approach tells us that our social decision rules ought to be chosen to track that fact.

Since democracy's performance with respect to tracking the 'correct' answers is what *justifies* democracy as means of collective decision-making for Rational Epistemic Proceduralists, whether or not democracy is in fact working to track that fact has a bearing on the *legitimacy* of the decision made.

We can now see why some policy commentators' and scholars' concerns, like Kitcher's and Oreskes', about the impact of problematic dissent on public policy and democracy can be understood, more fundamentally, as concerns about the democratic legitimacy of public policy. In explaining their concerns, each of these authors invoke some notion

---

[49] This is Peter's (2007) terminology; I adopt her classification system throughout this chapter.
[50] Note that in this article Cohen makes it clear that he is offering an account of epistemic democracy (or what he calls 'epistemic populism' in some places in the article), rather than a defence of it.

of truth or correctness with respect to individuals' or a collective's decisions. Kitcher (2016: 5), for instance, says that he takes it for granted that 'we can legitimately speak of some events as objectively bad' and that as such that he assumes the possibility of an 'objective discussion of values.' For Kitcher, alternative substantiations of the democratic processes relating to climate policy are therefore viewed and assessed as differing in their truth- or correctness-tracking abilities; the better the truth- or correctness-tracking potential, the more legitimate its outcomes. Kitcher and others view collective policy choices (e.g. about (in)action on climate change) as objectively (in)correct, and their concern is that particular manifestations of democratic policymaking processes — i.e. those in which problematic dissent flourishes — reduce democracy's correctness-tracking potential, and thus its democratic legitimacy.

However, there are two key reasons to hesitate in accepting this explanation of what goes wrong for policymaking in democratic societies when problematic dissent flourishes. The first is that the background assumption that there is *generally* a 'correct' policy outcome is controversial (Peter 2008: 345).[51] Policy choices are, after all, fundamentally forward-looking value-laden decisions (i.e. about the kind of world one thinks ought to be brought into existence), rather than backward-looking or contemporary descriptive statements about factual matters (propositions about features of the world as has been or is). Even if there is, in at least *some* circumstances, a policy option which is in some objective sense more just or morally better than another, it's not clear that this is even generally or often the case. It's plausible that, by virtue of value incommensurability, there often won't be an objectively correct decision, and that political life is replete with situations that are characterised by reasonable disagreement (Raz 1986: 342). This is perhaps all the more likely in societies that embrace and want to make room for value pluralism.

---

[51] I take Fabienne Peter herself to be *agnostic* about whether there are 'correct' policies independent of the democratic procedure. This distinguishes her from both rational epistemic proceduralists (who do take this to be the case, even they do not think that the correct policies can be identified apart from via democratic procedures, or who only think 'correctness' applies in general, at a higher level of abstraction than that of individual policies), and non-epistemic pure proceduralists, who take any type of 'correctness' to refer only to what the 'correct' or the 'right" procedures yield.

A Rational Epistemic Proceduralist might reply that what they mean by the 'right' policy, or a policy that is in some objective sense collectively rational, is simply that the policy would be selected for by majoritarian aggregation if the votes that individuals cast were in fact based on their laundered preferences.[52] However, while this is a fine response in theory, it is difficult, in practice, to make the case that voters are failing to meet this standard, and hence to make the argument that dissent is undermining democratic legitimacy in various cases. This is because people generally vote on a package of policies, or range of probable outcomes, when they vote for an elected representative, or participate in referenda, and many of these policies and outcomes will stand in trade-off relations with one another. It's therefore extremely hard — if not impossible — to read off how much weight individuals place on different policy beliefs and preferences by examining their all-things-considered policy decisions, and thus to assess whether the votes they cast were or were not aligned with their interests.

As an illustration of this point, commentators often cite a probable reduction in gross domestic product (GDP) if some policy is implemented as sign that people who voted for this policy 'voted against their own interests'. But it's possible that oftentimes people view this as a price they are willing to pay for an increased chance of realising some other value. Claims that people voted irrationally or unsoundly are often guilty of overlooking such possibilities. Additionally, with respect to elected representatives, political and other strategic factors also come into play, which due to the backdoor nature of much of political manoeuvring, are factors that will not be able to be observed and accurately accounted for by an outsider attempting to judge the all-things-considered rationality or correctness of a policy move.

Furthermore, it can be disrespectful at best, and dangerous at worst, to make claims about an individual's rationality. An important part of Isaiah Berlin's famous essay *Two Concepts of Liberty* (1958) was a warning about making judgments from our own limited

---

[52] 'Laundered' preferences are those preferences people would have were they to engage in a process of critical reflection, with full information, to filter and/or refine their preferences. Laundered preferences will therefore not be misinformed, maladaptive, self-defeating, etc. For more on laundered preferences see Goodin (1995).

standpoints about what other people's 'real' interests are, and the grave injustices inflicted on people when they are forced to follow these interests rather than those that they themselves take to be their interests. When we are selecting an account of democratic legitimacy that we hope to operationalise in the 'real world', these considerations are important. If an account of democratic legitimacy can only be used to make assessments of actual political decisions and situations at great cost of imposing extreme risks — perhaps especially to democracy itself — this is a serious limitation.

Finally, as Peter notes (2007: 347), any assessment of a policy's substantive qualities will only be able to, at best, account for those attributes we already know to be flawed in some way (i.e. sexist, racist, classist, etc.), and not those we are yet to collectively appreciate as being wrong, and that are presently hidden within our background assumptions. It is also going to be the case that there will always be some degree of uncertainty as to whether *we really are correct* about which policies are unjust/irrational. Perhaps this uncertainty will be even higher when the policy is attempting to adjudicate the rights claims or interests of two or more minority or disadvantaged groups, or when a society is characterised by such stark inequalities that it becomes very difficult for people to appreciate what the lives of their fellow citizens are like. Whilst there will always be these epistemological barriers — which is what democracy is ultimately trying to overcome, resolve, and instrumentally benefit from — I take it that in these types of cases these barriers are even higher, and a sense of being 'sure' we have chosen the 'correct' outcome is even further from reach.

On a much weaker version of Rational Epistemic Proceduralism it is assumed that there are correct policies to be found, though the rational epistemic proceduralist does not presume to know what they are. Whilst some democratic theorists may indeed only defend this weaker notion of democratic legitimacy, rather than anything similar to Kitcher's implicit acceptance of a stronger version of democratic legitimacy, my arguments in what follows do not depend on whether the form of Rational Epistemic Proceduralism adopted (or implicitly accepted) is a weaker or stronger version. My concerns hold even when no claims are made about what the right policy is. The only condition necessary for my critique is that there is assumed to be at least one correct policy decision, independent of the decision-making procedure; and as I have explained above,

this assumption is shared by both the stronger and weaker varieties of Rational Epistemic Proceduralism.

In sum, there are worries that a procedure-independent standard for the legitimacy of policy made in democratic societies is in many or most cases either non-existent, elusive, or dangerous to make inferences about. In the remainder of the chapter, I offer a diagnosis of how some dissent about science can undermine democratic legitimacy without relying on the controversial assumption that there is such a standard.

## 4.2. Pure Epistemic Proceduralism

Fabienne Peter's account of democratic legitimacy — Pure Epistemic Proceduralism — defines democratic legitimacy *exclusively* with regard to the fairness of democratic *procedures* (hence why it is 'pure'). Unlike other conceptions of epistemic democracy, such as those of the Rational Epistemic Proceduralists we encountered in the previous section, it does not treat the *outcomes* of decision-making processes as a source of normativity; democratic legitimacy is in no way dependent on criteria that refer to the quality of the outcome that the procedure generates.

Peter's account starts with, and is built up from, her recognition that there are both intrinsic and instrumental, and epistemic and non-epistemic, reasons to value democracy. As these values are what ground Peter's criteria for democratic legitimacy, it's important we have a good understanding of these values before I outline her criteria at the end of this section.

The primary intrinsic and non-epistemic reason democracy is valuable, Peter argues, is that democracy is a particularly important constituent of a society in which people relate to one another as *equals*. In democratic societies, the aspiration is that, as Kolodny (2014a: 196) says, 'none rules over any other.' Since all members of society are fundamentally moral equals, democratic decision-making procedures should be concerned with providing all persons with equal opportunity to influence political decisions (Kolodny 2014a: 227). Call this the *social equality* aspect of *political fairness*.

The primary epistemic value associated with democracy that Peter (2007:344) emphasises is its *knowledge-generating potential*, or what she notes Amartya Sen (1999) calls the *constructive function* of democracy. In Sen's words:

> [T]he practice of democracy gives the citizens an opportunity to learn from each other. ... Even the idea of 'needs' (including the understanding of 'economic needs') requires public discussion and exchange of information, views and analyses. In this sense, democracy has constructive importance, in addition to the intrinsic value it has in the lives of the citizens and its instrumental role in political decisions. (1999: 3)

As I see it, this knowledge-generating capacity stems from at least two key sources.

Firstly, it is by exchanging information with our fellow citizens *qua* citizens, or *qua* intermediaries or advocates, that we come to obtain knowledge about what other people within our society, or other demographic groups: view as issues affecting their lives; take to be the best means to solve these issues; are and are not willing to trade-off to realise certain ends, etc. Moreover, this process itself can, in turn, generate new knowledge when and where it helps to unearth new truths, lead to improvements in the accuracy of people's beliefs, or produce new ideas.[53]

For example, there are many large asymmetries in socio-economic indicators between Indigenous and non-Indigenous Australians, and every election each of the major political parties proposes different policies to 'close the gaps'. In reflecting on and then coming to a decision about which of the policy packages appears to have the greatest merit (which will in turn partly determine my voting choices), assuming I want to support the best possible policy for Indigenous Australians, I, as a non-Indigenous Australian, might gain valuable information by seeking out and engaging with Aboriginal and Torres Straight Islander Australians. This is because, as a non-Indigenous Australian, and moreover one who does not live in close community with Indigenous Australians, I am likely

---

[53] This idea is connected to the concept of 'situated knowers', and thus to standpoint epistemology (see: e.g. Hartsock 1983; Collins 1986; Harding 2004; Wylie 2012). It has recently been the subject of a book-length defence by Hélène Landemore (2012) in *Democratic Reason*. Democratic reason refers to the collective political intelligence of the many, which in turn is a function of individual epistemic competence and the cognitive diversity of the group. Landemore argues that it is for this reason that rule by the many is superior to alternatives, such as technocracy or autocracy. Similar ideas about the value of epistemic diversity and the 'wisdom of crowds' can also be found in Solomon (2001, 2006, 2008).

to have many misperceptions about the lives of Indigenous Australians, and to lack a good understanding of Indigenous Australians' values and priorities, and what Indigenous Australians take to be the best and most effective means to various ends. Information and testimony from my fellow [Indigenous] citizens, can, for example: (i) help to improve the accuracy of my empirical beliefs e.g. about the rates at which Indigenous youth are incarcerated, which can in turn lead to a better appreciation of the magnitude or seriousness of an issue; (ii) help me obtain a more comprehensive picture of the range of views and opinions within the indigenous community as to how the situation for Indigenous youth might be improved, and (iii) help me to better appreciate *why* it is that an Indigenous person, or sub-set of the Indigenous community, view one of the policies proposed as being preferable or morally unacceptable.

The second key means of knowledge generation is that democracy provides citizens with an opportunity to receive the testimony of their fellow citizens *qua* experts, and moreover, to critically engage with and contest these experts' claims. Through such processes, both experts' and non-experts' factual and other beliefs may change and improve, and again, new shared knowledge can be realised and objectives better met. Consider the way that, just decades ago, homosexuality was classified by psychiatrists as a mental illness and included within medical diagnostic manuals. The overturning of this particular consensus within medicine was, in a large part, thanks to the persistence of gay and lesbian rights activists, who were for the most part not psychiatrists or scientific experts, but who levelled incisive criticisms of many of the background assumptions and extant empirical literature on which psychiatrists had based their conclusions (Bayer 1987a, 1987b; Drescher 2015). These contributions from 'lay experts' not only drove new knowledge generation and the correction of incorrect empirical beliefs amongst both experts and non-experts, but also helped overturn several unjust medical practices and broader public policies. The recognition that many of those members of society who aren't contributory experts to an academic field often have expertise of other kinds, or valuable knowledge to contribute, has been an important development in philosophy of science and science studies over the past several decades.[54]

---

[54] One of the seminal papers on this front being Wynne (1989).

With this in mind, let us return to Peter's Pure Epistemic Proceduralist account of democratic legitimacy. For Peter (2008a: 138), an outcome (e.g. public policy) is legitimate if and only if the democratic processes that led to it meet conditions of *political fairness*, where political fairness requires that public deliberation and associated voting is:

i) Carried out under conditions of *social equality*; and

ii) Characterised by *epistemic fairness.*

In the following two sections I further articulate what each of these conditions require, whilst also showing that some dissent about science can undermine the democratic legitimacy of public policy, because some dissent reduces the extent to which these conditions are met.

## 4.3. Problematic Dissent and Socio-Political Equality

Although Peter states that political fairness requires public deliberation under conditions of *social equality* (2007: 138), she doesn't provide more fine-grained detail on the precise nature of this condition. For our purposes here then, we can look to other democratic theorists to provide us with a better sense of what these conditions are. In particular, I will draw upon Niko Kolodny's (2014a; 2014b) recent scholarship on this matter.

### 4.3.1. Socio-Political Equality as Equal Opportunity to Knowingly Influence Political Decisions

Kolodny's argument starts from two basic premises: that all human beings are, on a fundamental level, moral equals, and that relations of social inferiority and superiority are inappropriate among moral equals (2014a: 226, 2014b: 300). Although certain hierarchical social relations might be objectionable in many domains of human life, Kolodny thinks that relating to one another as social equals is especially important in the political realm. This is because political decisions have a type of final authority over collective life and their implementation involves the use of coercion and force (2014a: 226).

This leads to the further question: what it is to relate to others as social equals in the political realm? Kolodny contends that what matters first and foremost is the distribution of *political power* and *de facto authority* over communal life (2014a: 227).[55] Thus he argues that for individuals to stand in relations of equality with each other, people need to have *equality of opportunity* to *knowingly influence* political decisions in line with their own judgments (2014b: 332, emphasis in original).[56] Any deprivation of this equal opportunity would count as a type of subordination.

Kolodny (2014b: 309-310) goes on to further specify what having equal opportunity to influence political decisions amounts to. He outlines five features.

1. Firstly, what matters is that someone has an *opportunity* for equal influence, not that they exercise this opportunity. Thus, people can choose not to participate in a decision-making process without their decision to not participate having any bearing on their standing as a social equal.

2. Second, our interests are in *relative*, rather than absolute, influence: each person simply needs to have as much opportunity for influence as anyone else has (Kolodny 2014a: 199). It is also an *ongoing* relational interest, in that these relations of equality need to be instantiated in each application of a collective decision-making procedure.

3. Third, an individual can be said to have an equal opportunity to influence a decision when, and to the extent that, the decision was reached by a process that was positively sensitive to the choice or judgment of every person (2014b: 309). This *contributory* type of influence occurs when an outcome — which is determined by a combination of vectors of force (e.g. votes) — is sensitive to each particular vector of force without any of the magnitudes or directions of individual vectors being distorted by the application of other vectors (e.g. your vote does not annul my own, but it can in combination with more votes in the same

---

[55] Whilst Kolodny acknowledges that asymmetries in 'the distribution of means' to chosen ends can be indicative of problematic varieties of socio-political inequality, he argues that this is not what principally matters because even if a benevolent dictator or a group of technocrats could reliably achieve distributional justice, we would still take that society to be marked by some form of objectionable inequality.

[56] Importantly, it is not required that decisions have equal opportunity of being made.

direction, out-compete the option I voted for) (Kolodny 2014a: 200; 2014b: 320). More precisely, Kolodny stipulates that equal opportunity for contributory influence occurs when the following holds:

By X-ing, I exercise equal contributory influence over a decision just when my X-ing has equal a priori chances of being decisive over the decision, that is, has equal chances of being decisive on the assumption that no pattern of X-ing by others is more likely than any other pattern. (2014b: 321)

4. Fourth, Kolodny stipulates that what matters is equality of opportunity for *informed* influence. He argues against those scholars who hold that it's enough for social equality that individuals have *formal* social equalities (e.g. universal suffrage of the one person, one vote variety, where each vote has equal weighting). As Kolodny points out, in the extreme, disparities of information and knowledge could make one person another person's slave (2014b: 310). Kolodny explains this with a metaphorical reference to two individuals — one who is in chains and whose attempts to unlock their chains to liberate themselves are misfiring, because they're currently only able to press numbers at random, and another person who knows the code to unlock their chains, but who doesn't share the code with the unfree person because keeping them in chains somehow allows them to exploit this knowledge asymmetry to manipulate that person in a slave-like fashion. Note that the same idea applies to disparities in knowledge between social groups within societies too, such that one group can exploit asymmetries in knowledge to co-opt other groups and individuals to further their own political projects, even if these activities are not in fact in the best interests of the individuals or groups they co-opt. Such power imbalances and risks of this kind of exploitation can be eased by ensuring that all persons have access to as much information as everyone else has.

5. Lastly, they must have *autonomous* influence. By this Kolodny means that people's judgments must themselves be reached via 'free reflection on what one takes to be relevant reasons' (2014b: 310). He seems to include this condition to safeguard the voluntariness of individuals' decisions about their final values

and ends. Kolodny says that: 'It scarcely defuses the threat to social equality if I can manipulate the judgments that underlie your vote' (2014b: 310).

It's not clear that these last two conditions are as distinct as Kolodny seems to think they are; being informed, or at least having sufficient opportunity to be informed, is related to personal autonomy, as I argued in chapter two. And the worry in both is about one person or group having the ability, by virtue of power differentials, to manipulate the others' political influence in order to bolster their own. We might keep these two conditions separate though by saying that in condition four, the worry is that even though two people or groups share the same ends, they differ in the means by which they'd be willing to achieve that end, and one person/group disrespects this and uses the other person's political influence to achieve one set of shared ends. By contrast, condition five guards against situations where one person or group is manipulated by another, more powerful group or person, in to thinking that they share a common interest in X, when in fact Y is the less powerful person or group's 'real' interest.

However, it might be objected that even within democratic societies, some individuals will inevitably and inescapably have more power and authority (and thus influence) over collective decision-making than others. An objector might point to the presence of 'technical experts' in society, who necessarily generally have more knowledge in some domain than others, i.e. 'non-experts'. They might argue that this asymmetry in knowledge gives experts greater relative power over decision-making vis-à-vis non-experts, and that this violates the equality constraint.[57] Kolodny (2014b: 333) even suggests that:

> [T]he very division of labour [in society] seems to militate against [equality of opportunity for influence]. In any moderately complex society, some will be paid to acquire relevant political information, whereas others will have to acquire it, if they acquire it at all, in their spare time.

I agree with Kolodny that the division of labour within society sets up inherent vulnerabilities within democratic governance systems. It's also true that these vulnerabilities can be exploited by those who wish to have greater relative (political) influence vis-à-vis others within society. But Kolodny himself provides us with resources that help us to

---

[57] Kolodny (2014b: 296) calls greater knowledge a 'raw' power.

appreciate, firstly, that it would be a mistake to view the mere existence of these asymmetries as *necessarily* problematic, and second, that there are things we can do to ameliorate potentially problematic asymmetries, or the problematic aspects of them.

With respect to the first point, Kolodny argues that what he calls a 'judgment-dependent' inequality in opportunity for influence does not compromise equality of opportunity (2014b: 334). This type of inequality arises merely as the result of a hearer excising their judgment and pursuing their own convictions as part of their own equal opportunity for political influence. He gives the example of an Expert and a Crank, who both make factual assertions pertinent to a policy issue. If a Hearer, on the basis of their autonomous judgment, chooses to disregard the Crank's testimony and defer to the Expert's, thus accepting the Expert's factual assertion as a premise in their own practical reasoning, the resulting inequality of influence that seems to arise between the Expert and the Crank is not a problematic one. This is because, on Kolodny's reasoning, this inequality was the result of a third party's own judgment.

To extend Kolodny's argument, the inequality in opportunity to influence others that arises between Experts and Non-Experts (and perhaps especially between 'Experts' and 'Cranks'), also seems to emerge as the result of prior democratically mandated and controlled decisions about how epistemic labour will be divided up and allocated to best benefit society. We might say that there is a tacit background social contract between scientific experts and non-experts where the public has provided scientific experts with resources and a social licence to carry out research in exchange for practice- and policy-relevant scientific information and knowledge. The Crank, on the other hand, has no such social contract with the public. As such a Crank may be ignored by a non-expert on the basis of not being bound by the same social contract, and thus may have less 'influence' than the expert in the process. On the other hand, the Expert is understood to have a responsibility to share this knowledge and information, and moreover, to share it in a way that does not unfairly bolster their own personal political projects. As such, the actions of a scientist who behaves as a 'stealth issues advocate' (Pielke 2007), for example, may interfere with democratic legitimacy.

Second, Kolodny explains that where potentially problematic inequalities in influence could arise, it's generally possible for these asymmetries to be regulated in ways that

neutralise their presence, or convert them into assets that all benefit from equally. For instance, Kolodny (2014b: 295) argues that it isn't necessary for social equality that there is a strict equality of 'raw' power or authority between people (e.g. all having the same knowledge). Rather, what social equality demands is that those who do have a latent potential for relative power or de facto authority over others (whether formal, legal, or otherwise), be resolutely disposed to refrain from exercising that greater power. They have to be willing to concede to others as much influence as they enjoy over common affairs, and to recognise that this is something to which others are entitled.

This can be achieved via both the external regulation of potentially powerful or authoritative people and sectors and their activities by democratic civilian control, as well as internal regulation (self-regulation or industry-regulation) and the development of appropriate moral dispositions amongst practitioners (2014b: 296). These dispositions have a bearing on how those who occupy various social and professional roles — e.g. scientific experts, journalists who communicate about science, and politicians who make claims about science and scientists — inhabit these roles and use their considerable margin of discretionary power.[58] When those with potential for greater power use that greater power over others and subordinate them, they exhibit a pathological disposition that undermines social equality and political fairness. By contrast, when they self-consciously check this power and ensure that they exercise no more influence over political decisions than any other, they respect and further social equality, and as such, strengthen democratic legitimacy.

The five aforementioned conditions therefore constrain the *type* of democracy we *should* have (Kolodny 2014b). Because speech makes up a large part of what politics is, these conditions might constrain the sorts of speech that we ought to welcome in a democracy. If particular types of speech interfere with these conditions being met, people will not be provided with an equal opportunity to influence political decision-making, and the democratic legitimacy of associated public policy will be undermined. In what follows next I argue that some dissent about science can do just this.

---

[58] Also see Zacka (2018), who makes a similar argument for the development of the right moral dispositions among 'street-level bureaucrats'.

*4.3.2. How some dissent about science interferes with realising equality of opportunity for political influence*

As I argued in Chapter Two, some dissent about science can lead to some people within a society non-voluntarily forming false or inaccurate empirical beliefs. That they have formed these false/inaccurate beliefs non-voluntarily in the first place is partly because there was not adequate information or higher-order evidence of sufficient integrity available within the public knowledge system. When public knowledge systems are characterised by such conditions, many people will not have equal opportunity with their fellow citizens to influence policy in an *informed* way, as Kolodny's fourth criterion requires.

Furthermore, in Chapter Two I also argued that the possession of non-voluntary false/inaccurate beliefs has the effect of compromising people's personal autonomy on two fronts: by undermining a person's control over their choice of values, and by undermining their control over whether or not their actions realise their values. Because Kolodny requires that individuals have equal opportunity for *autonomous* influence over political decisions (condition five), where dissent about science produces non-voluntary false/inaccurate beliefs amongst hearers, and where these false/inaccurate beliefs are pertinent to public policy, political fairness will also be undermined from this angle too.

In Chapter Three, I argued that when someone's personal autonomy is undermined by certain types of dissent about science, they are pre-emptively silenced, and their freedom of speech is curtailed. When someone is silenced in this way, and as a result are either side-lined from democratic discussions and deliberations, or are not making speech acts in these conversations that reflect their own mind, it's hard to see how someone can be understood to have a genuine *opportunity* for influence. Because someone who is pre-emptively silenced and marginalised from democratic debates cannot be understood to have chosen to not participate — their non-participation was forced or compelled — the absence of their voice in political discussions and decisions bears on their standing as a social equal.

When these dynamics play out, Kolodny's second criterion is violated because we are left with a situation in which some actors have greater relative opportunity for influence over political decisions vis-à-vis others within society. This is because problematic speech acts about science can work, on the one hand, to exclude some from having any influence on political decisions and, on the other hand, to co-opt others' political influence and make use of it in a way that serves their own purposes. The cumulative effect is that some within society end up with much greater influence over political decisions compared with others, and people's interests will therefore be unfairly unequally advanced or hindered.

With regards to climate policy, for example, what the domination of public debate with endless contestations of fundamental empirical claims has ultimately achieved is a constraining of both opportunities for political dissent within these arenas – due to opportunity costs as well as the effects of pre-emptive silencing (see Chapter Three) – and a covert narrowing of the political space over time, as certain policy options have become physically impossible or otherwise unviable (e.g. holding global warming to 1°C).[59] In contrast to this kind of tapering, efforts to clarify important features of the epistemic landscape don't close down democratic deliberation or the range of policy possibilities, or favour one particular interest, because the acceptance of factual propositions *cannot determine policy*. In stark contrast to the way that problematic dissent confuses, conceals, and constrains, this clarity *opens up* the range of feasible policy directions at a point in time, and helps to provide for an equality of opportunity to influence decision-making.

Moreover, the inequalities that do arise under such scenarios cannot be dismissed as they are not judgement-dependent inequalities; they are generally achieved through misrepresentations of science and expressions of misinformation that have the effect of bypassing people's critical faculties rather than engaging them. To be clear, this need not be a conscious, intentional process; some people might sincerely believe in the truth of the claims they express or transmit, and might not be intentionally taking away others'

---

[59] At the time of writing, Earth's climate system has warmed by about 0.9 °C from the mid-19th century to the present decade. It's therefore almost certain that we will see an additional 0.1°C of warming by 2100. And whilst keep warming to 1.5 °C is not yet physically infeasible, it will be extremely challenging to do so (Millar et al. 2017).

opportunities to voluntarily make up their own minds as they see fit. They might not be making calculated decisions to mislead people. But this does not mean their actions (their speech acts) are not negligent or reckless.

In sum: where and to the extent that dissent about science undermines social equality in these ways, the democratic legitimacy of the resulting public policy is correspondingly eroded.

## 4.4 Problematic Dissent and Epistemic Fairness

As part of what counts as 'political fairness', Pure Epistemic Proceduralism calls for *epistemic fairness*. Recall that this is motivated in a large part by democracy's 'constructive' function or knowledge-generating potential, as explained above. Peter makes clear (2007: 344) that the criteria for epistemic fairness are to extend to regulating public deliberation at the *fact-gathering and analysis stages* of policymaking processes, i.e. those stages many might have viewed as being only, or predominantly, within the purview of technical experts. Given the value-ladenness of science and science-informed policymaking (Douglas 2009), this is an important virtue of Peter's account in comparison to other accounts of epistemic democracy that overlook this quality.

Peter doesn't outline or definitively endorse one set of criteria for epistemic fairness. However, she does say that those epistemologies developed by feminist philosophers of science and epistemologists are among those that best fit her account (2007: 341). Because Peter seems to be particularly persuaded by Helen Longino's social epistemology of science,[60] I will use it here, first, to give more content to Peter's account of democratic legitimacy and, second, to show how some dissent within or about science can erode epistemic fairness and thus democratic legitimacy.

### 4.4.1. Epistemic Fairness Through Longino's Social Epistemology

Longino's account of the production of scientific knowledge, first developed in *Science as Social Knowledge* (1990) and then expanded upon within *The Fate of Knowledge* (2002),

---

[60] See Peter 2013, fn 11, p. 1264 for Peter's own statement to this effect.

is known as 'contextual empiricism' (1990: 215), or 'critical contextual empiricism' (2002: 208). For Longino, it is the social character of scientific inquiry that allows scientific knowledge to be produced. Individual observation and reason are necessary but not sufficient; they must be supplemented by collective, social processes.

The social process of paramount importance for Longino is critical discursive interaction. She says that: 'Once propositions, theses, and hypotheses are developed, what will become scientific knowledge is produced collectively through the *clashing and meshing of a variety of points of view*' (1990: 69, my emphasis). She says that scrutiny and criticism from a diversity of perspectives 'not only spurs evaluation and re-evaluation of hypotheses, but also leads to better appreciation of their grounds and of their consequences' (Longino 2002: 129).

She prescribes four criteria for scientific inquiry that she takes to be *necessary* to ensure the *effectiveness* of these discursive interactions (2002: 128–135). These four features are also known as 'criteria of knowledge-productive capacity', 'norms applying to the social practices and processes of cognition' and 'conditions of effective or transformative criticism' (2002: 134). She thinks that the satisfaction of these norms helps to provide assurance that the theories and hypotheses accepted by scientific communities have been divested of the sorts of biases of individuals and sub-groups that can mislead (2002: 134). She is also confident that her criteria will enable us to distinguish between social interactions that are productive of knowledge and those that are not.

Her first criterion is that there be *recognisable public venues* for critiquing research to provide for the criticism of evidence, methods, assumptions, reasoning and claims from a variety of viewpoints. In outlining this criterion, she reminds us that criticism not only spurs evaluation and re-evaluation of hypotheses, but also leads to better appreciation of their grounds and of their consequences (2002: 129). She also thinks that various practicalities (e.g. space limitations), contemporary values (e.g. for positive, rather than negative, research results), and structural developments in science (e.g. commercialisation and privatisation of research, information, and ideas), all work against critical discourse and the ideal fulfilment of this first criterion.

Second, she requires communities to *respond to and 'uptake' expressed criticism*. This is to ensure that criticisms are not ignored but, rather, given careful consideration and addressed. Communities satisfy the requirement of uptake if they consist primarily of individuals who are genuinely responsive to criticism, and individuals are genuinely responsive to criticism if they are willing to question *any* of their beliefs by considering any criticism from *any* source, which importantly includes 'non-experts' (2002: 129).

Longino's third criterion, that there are *publicly recognised shared standards* for the evaluation of theories, hypotheses and methodologies, is meant to ensure that criticism is *relevant* — that is, that it appeals to some standards held by the person(s) maintaining the position — as well as to make sure that what counts as relevant is not merely a function of the arbitrary whims of individuals but, rather, of a set of shared standards to which the community is bound (2002: 130). She also explicitly requires that these standards themselves be subject to occasional criticism (131). She again makes it clear that evaluations can be performed by *anyone*, rather than only those members of a community that share those standards.

Lastly, she stipulates that there must be a *tempered equality of intellectual authority*. Longino clarifies that intellectual authority refers to a person having the sorts of cognitive skills (e.g. of observation, synthesis, analysis, and critique) that enable them to participate in critical discussions and contribute to the understanding of matters they have less knowledge about than others (2002: 133, fn 19). Intellectual authority therefore differs from what Longino calls *cognitive authority*, which relates to the amount of domain-specific knowledge a person holds. Equality of intellectual rather that cognitive authority, she makes clear, means that scientific communities must include and attend to the criticisms of 'outsiders' (134). She requires that every participant of the discussion be respectfully regarded as equally capable of contributing persuasively and decisively to critical discursive interactions. She clarifies (132, fn 15) that equality here, as an epistemological rather than moral/political criterion, requires that *all* relevant perspectives be *included*, rather than only that different perspectives have an *equal chance* or *equal opportunity* of being included.

Therefore (and importantly), not only is her criterion different from equality as a moral/political criterion (as in section 4.3. above), it is also *stronger* and therefore *more*

*demanding.* She asserts that the point of the requirement of equality is to 'ensure the exposure of hypotheses to the *broadest* range of criticism' (132, my emphasis); to *maximise* the community's critical resources. This is required not only to protect against the sorts of injustices women and racial minorities have experienced in the past, but also help to ensure that there are no 'cognitive failings', and that any consensus that does form does so for legitimate reasons. For Longino, a legitimate scientific consensus is one that results from a 'critical dialogue in which *all relevant* perspectives are represented' (131, my emphasis), and where any assent is 'unforced' rather than merely the product of social and economic power dynamics (132).

### 4.4.2. How problematic dissent interferes with the fulfilment of Longino's criteria

On an account of democratic legitimacy that uses Longino's criteria, public policy in democratic societies, especially that informed by scientific information, will therefore be legitimate only if the science that underpins it was scrutinised within communities that meet Longino's four criteria. The question then is whether the presence and extent of some dissent about science within public knowledge systems contribute to a failure to meet these criteria, and therefore by extension, to preventing/undermining epistemic fairness and thus democratic legitimacy.

Firstly, it is plausible that some speech acts or omissions about science, including problematic dissent over scientific claims, can lead to situations in which non-experts are not able to recognise the public venues for critiquing scientific research. When this happens, the fulfilment of Longino's first criterion will be interfered with. For instance, some dissent can lead non-experts to perceive that certain groups of people are *in fact* the relevant scientific community, when they are not, or when they are at best only a small part of it. A non-expert who believes this group of people *are* the relevant scientific community, and who proceeds to engage with this particular community on the basis of such a belief, may be under the impression that their comments and criticisms have been heard by the relevant scientific community when they will not have been.

Perhaps something like this happens when members of the public and policymakers come to engage with the work of the Non-Intergovernmental Panel on Climate Change

(NIPCC) rather than the Intergovernmental Panel on Climate Change (IPCC), or simply with mainstream climate science. The NIPCC describes itself as:

> an international panel of scientists and scholars who came together to understand the causes and consequences of climate change. NIPCC has no formal attachment to or sponsorship from any government or governmental agency. It is wholly independent of political pressures and influences and therefore is not predisposed to produce politically motivated conclusions or policy recommendations.[61]

They even say that the NIPCC was created:

> to *take the IPCC's place* by giving honest scientists an alternative outlet for their work and citizens of the world a genuinely independent source of research on this issue (my emphasis).

The NIPCC operates by mimicking the structure and activities of the IPCC. It publishes reports with similar titles such as: *Climate Change Reconsidered: The 2009 Report of the Nongovernmental International Panel on Climate Change (NIPCC)*, *Climate Change Reconsidered: The 2011 Interim Report of the Nongovernmental International Panel on Climate Change (NIPCC)*, *Nature Not Human Activity Rules the Climate: Summary for Policymakers of the Report of the Nongovernmental International Panel on Climate Change*, and *Climate Change Reconsidered II: Physical Science.* The NIPCC says that 'In the tradition of the IPCC', it publishes its *Summary for Policymakers* in advance of completing the underlying report. These reports are written and reviewed by authors that claim expertise in climate science, but who generally don't have a track record of publishing articles in peer reviewed academic journals on the topics they write about for the NIPCC.

However, given that the NIPCC's operation and funding is tied to highly partisan think tanks — such as the free market Heartland Institute — and the fossil fuel industry, including from the Koch brothers, there is room for scepticism about their claim to be a source of 'genuinely independent research'. Additionally, at least two people involved in the NIPCC, Fred Seitz and S. Fred Singer, also have a long history as the U.S.'s chief merchants of doubt (Oreskes and Conway 2010).

---

[61] Taken from the NIPCC's website: http://climatechangereconsidered.org/about-the-nipcc/, September 2018.

The problem is that much of this higher-order evidence — information relevant to evaluations of the reliability of the source of the first-order evidence — is not relayed to the public. In a context in which this evidence is absent, and in which the NIPCC portrays itself misleadingly, it's perhaps understandable that many non-experts have come to believe that the NIPCC, rather than the IPCC, is the relevant scientific community to engage with about climate science. To date, the NIPCC has run twelve 'International Climate Change Conferences', which are attended by many policymakers and interested lay persons. Lamar Smith, chair of the United States House Committee on Science, Space and Technology, was a speaker at last year's conference. Perhaps the biggest worry though is not those very ideological people who have chosen to close themselves off within an echo chamber and not engage with the IPCC, but the people they in turn influence. The NIPCC is often uncritically quoted or referred to by journalists,[62] and this may have the effect of misleading readers about both the first-order claims of climate scientists, and the higher-order evidence that relates to their trustworthiness. These misrepresentations can also cut them off from these communities in other ways, as I explain below.

Some dissent also appears to frustrate the fulfilment of Longino's second criterion: the need for scientific communities to respond to and 'uptake' expressed criticism. The expression of certain types of dissent can lead scientists and particular scientific communities to — in many cases, very reasonably — avoid engaging with dissenters in general, due to the time wasting and attacks they've experienced in their previous interactions with other dissenters (Biddle and Leuschner 2015; Biddle, Kidd and Leuschner 2017). At one conference I attended last year, an astrophysicist admitted that he has a special email folder where he files away emails from those who claim they have found errors in Einstein's mathematical equations. He doesn't properly read these emails, or in many cases even open them. He said that he is 'inundated' with these kinds of emails, and that seriously reading and writing back to each one would have opportunity costs which

---

[62] As just one example, the *Daily Mail*, Britain's highest circulating newspaper and the world's most trafficked online news site, published a news article summarising the NIPCC's 'scientific report'. See Johnathan O'Callaghan, 'Don't fear climate change - rising CO2 levels and a warming planet could be beneficial, claims study', 9 April 2014: https://www.dailymail.co.uk/sciencetech/article-2600565/Dont-fear-climate-change-rising-CO2-levels-warming-planet-beneficial-claims-study.html

would mean he would fail to meet a range of other professional obligations that he perceived to be more important or worthwhile. It seems very plausible that many scientists, especially those whose science is embroiled in public controversy, deal with dissent in similar ways, for similar reasons. But it's of course possible that in amongst the deluge of dissent that is not epistemically beneficial, there is dissent that is. Thus, some acts of frivolous or problematic dissent, by making unreasonable demands on scientists, may be displacing the attention other dissent would receive from scientists in a counter-factual world in which scientists' inboxes were less inundated by dissent from non-experts.

Some dissent may also contribute to making it more difficult for other non-experts to criticise the background assumptions of experts (i.e. the 'shared standards' within the scientific community) — Longino's third criterion. As with public venues for critiquing research, these shared standards are meant to be able to be 'publicly recognised'. But as Oreskes and Conway (2010: 247) note, some dissent not only misinforms hearers about scientists' first-order claims, but also misrepresents the methodologies and processes scientists have followed that led to them making those claims. If non-experts cannot accurately identify the actual shared standards the scientific community is bound by, they will also be unable to recognise when the scientific community is and is not meeting these standards, or be able make good judgements or critiques about their appropriateness. Criticisms under such circumstances are likely to misfire, and criteria three will remain unfulfilled.

Lastly, some dissent can undermine the degree to which there will be a tempered equality of intellectual authority, which recall refers to equal abilities in observation, analysis, synthesis, etc. This is because in many cases, a pre-requisite for someone having a certain degree of intellectual skill is that they simultaneously have access to an undistorted picture of the relevant epistemic landscape. It seems unlikely that someone could count as a 'skilled observer', for example, if the evidence base upon which they are relying to make observations is systematically distorted in some way. One's 'skills' as an observer don't exist completely independently of the knowledge base upon which they are deployed, and which subsequently, through a looping effect, in turn condition what the observer 'sees'.

Perhaps someone might say that a skilled observer *just would know* that the evidence base is systematically distorted. But such person would then be some type of expert themselves, or at least that they would likely have a certain degree of interactive expertise as well as access to the sort of higher-order evidence that would indicate that the evidence is systematically distorted (rather than the sort of ordinary person Longino's criterion is meant to make room for, who e.g. possess neither contributory or interactive expertise, nor direct access to higher-order evidence). For instance, someone could have perfect observational skills in one sense; they might, like Neo in the film *The Matrix* (1999), see a woman in a red dress walking through the street. But like Neo, they also might not know, or even be able to know through their own senses alone, a deeper truth; in Neo's case, that the woman in the red dress, like the whole streetscape, is a computer simulated training program rather than the Matrix. Neo did not know he was in a simulation until someone with more expertise and knowledge than he had (i.e. Morpheus) testified to Neo that this was the case (i.e. shared his knowledge with Neo). Our cognitive skills are intrinsically entwined with the knowledge and skills of others, and perhaps especially with experts. To have equal intellectual authority with others, we depend on having undistorted access to others' claims, as well as relevant higher-order evidence.

Where and to the extent that dissenting speech acts about science, or omissions of information, mean that Longino's four criteria are not met, epistemic fairness will be undermined. This, in turn, means that some dissenting speech acts about science, and omissions, can check the democratic legitimacy of public policy.

## Conclusion

In this chapter I argued that concerns about the impact of problematic dissent about science on public policy and democratic values are best understood as concerns about the democratic legitimacy of public policy. Although much of the extant literature frames these concerns in ways that accord with the Rational Epistemic Proceduralist characterisation of democratic legitimacy, I argued that Fabienne Peter's account of Pure Epistemic Proceduralism yields an alternative, and more attractive, diagnosis of the how problematic dissent can interfere with the democratic legitimacy of public policy. Peter's account of democratic legitimacy avoids controversial assumptions about the correctness of the *content* of public policy, focusing instead only on the *procedure* it was made through. Peter's Pure Epistemic Proceduralist account of democratic legitimacy rests on a link between procedural fairness and epistemic value. By requiring that public deliberation tied to policymaking is carried out under conditions of *social equality* and characterised by *epistemic fairness*, Peter's account captures the important idea that inclusive public deliberation is valuable not just for the ideal of political fairness that it embodies, but also for its epistemic contribution to public deliberation and democratic decision-making. Using Kolodny's more specific conditions for the social equality aspect of political fairness, and Longino's criteria for epistemic fairness, I demonstrated that some dissent about science can interfere with the degree to which such conditions are met, and hence, can undermine the democratic legitimacy of public policy.

# CHAPTER FIVE:
## MAKING DECISIONS ABOUT DISSENT

*Introduction*

My analysis so far has drawn attention to the fact that what makes some dissent within or about science 'problematic' is not merely a function of its negative epistemic consequences, but also its undesirable non-epistemic risks and consequences. I have argued that under some circumstances, some dissent has the potential to compromise three interlinked values that are particularly important in liberal democracies: personal autonomy, freedom of speech, and the democratic legitimacy of public policy.

But what, if anything, should we do about problematic dissent? It's worth noting that many people *inevitably* and *inescapably* must make decisions about dissenting speech about science every day. Teachers, librarians, media editors, journalists, journal editors, peer reviewers, book shop managers, politicians, social media companies, and so on, are all forced, due to limited resources and professional obligations, to make numerous decisions about speech about science, including what speech merits and will receive attention, and what kind of attention, in different parts of the public knowledge system.

Whilst scholars seem to agree that some dissent can be problematic, there is currently deep disagreement about how we ought to respond to its expression and transmission, and whether certain types of responses help to facilitate progress or undermine it. For instance, in a recent article — 'Who's Afraid of Dissent?' — Inmaculada de Melo-Martín and Kristen Intemann (dMMI) (2014) agree that challenges to a scientific consensus by dissenting voices can have considerable negative consequences, including to public policy (p. 596, p. 599). However, the purpose of their article is to argue that it is both

misguided and dangerous to try to minimise or discourage the expression and transmission of dissent within the public knowledge system that is perceived to be problematic. They label such minimising or discouraging the 'targeting'[63] of dissent.

According to dMMI, targeting dissent is misguided because it fails to address what they diagnose as the 'real problem': the prevalence of false assumptions about the nature and role of scientific consensus, dissent and evidence in policymaking (pp. 599–607; p. 609). Targeting dissent is also dangerous, dMMI argue, because it likely increases instances of problematic dissent, deprives us of resources to criticise special interest science, and stifles legitimate dissent that is crucial to scientific progress and sound policy (pp. 607–609). Consequently, dMMI conclude that scholars should turn their attention elsewhere to make progress (p. 610) and propose correcting false beliefs about science-informed policy through education as an alternative intervention. Two more recent papers echo similar arguments. Pearce et al. (2017) and Intemann (2017) claim that efforts to communicate the expert consensus about empirical claims — among those strategies that dMMI say constitute targeting dissent — are at best ineffective and at worst have a deleterious impact on public policy. Whilst Pearce et al. characterise consensus messaging as a 'distraction' from more urgent matters, Intemann warns that it may simply reinforce what she argues are problematic assumptions about scientific consensus and its role in policymaking and that it may backfire, undermining trust in science.

On the other side of the debate, science historian Naomi Oreskes (2004; and Conway 2010; 2017) and cognitive scientists Jonathan Cook (et al. 2016; 2017) and Sander van der Linden (et al. 2014; 2016; et al. 2017), among others, hold that targeting problematic dissent in certain ways is sometimes an effective, reasonable, and legitimate response to the risks associated with problematic dissent, especially in contexts where powerful political actors have either deliberately manufactured unwarranted levels of doubt about scientists' empirical claims, or displayed a reckless disregard for the truth about empirical matters. To support their case, they point to research that suggests that misperceptions of, for example, an expert consensus about empirical claims can be consequential

---

[63] I would prefer to use 'checking' dissent instead, as I think 'targeting' is an unduly loaded term. However, I will use dMMI's language of targeting within this chapter for consistency. But I will return to the issue of terminology at the end of this chapter.

for policy, and that an improved understanding of the degree of consensus can lead to better engagement with empirical issues, and influence peoples' policy attitudes (e.g. Ding et al. 2011). Even more promising for their position, studies have also found that this outcome often holds across the political spectrum, and that information about the degree of consensus can sometimes neutralise the effect of a person's pre-existing worldview on the formation of their beliefs about empirical matters (Lewandowsky et al. 2013; van der Linden et al. 2014; et al. 2016).

My aim in this final chapter is to offer a less polarised perspective that integrates, rather than sets into competition, insights from both sides of the debate.[64] However, because I think some scholars have been too quick to dismiss the importance and legitimacy, in some contexts, of various responses that dMMI classify as targeting dissent, I will also be providing a qualified defence of certain targeting practices, including consensus messaging. I develop my case primarily via close critical engagement with dMMI's article, as I take it to be perhaps the key paper in this debate. Notably, it is unclear whether dMMI take targeting dissent to always be misguided and dangerous, or just generally so. Although they do not offer any such caveats, I will assume that they intend the latter, as it is more plausible.

In the first part of the chapter (§5.1–§5.2), after clarifying what it means to target dissent and providing a more detailed overview of dMMI's argument that targeting dissent is misguided, I present an additional set of assumptions that dMMI overlook, which make targeting certain problematic dissent as an obstacle to public policy sound in some circumstances. I call this the 'Probability Argument', as it focuses on the policy consequences of non-experts' having distorted perceptions of the probabilities that pertinent hypotheses are correct. Before moving on, I raise and respond to two objections to the Probability Argument: that it does not guarantee that any changes to policy as a result of targeting problematic dissent will make for 'sound' policy, and that the Probability Argument is simply the 'deficit model' in disguise. Whilst I have developed the Probability Argument in an attempt to go some way towards resolving disagreements over the

---

[64] Some of the scholars in this debate have already acknowledged their view that different approaches can coexist (e.g. Cook and van der Linden 2017).

wisdom and ethics associated with targeting problematic dissent, it seems a useful framework for reasoning about science communication more generally. It also further clarifies the critical role and significance of higher-order evidence, a concept that has often been neglected in scholarship on science communication. As such, it has significance beyond just this corner of the philosophical and science studies literature.

In the second part of the paper (§5.3–§5.4), I examine dMMI's argument that targeting dissent is dangerous. I show that the same reasons dMMI provide to ground their argument that targeting dissent is dangerous also apply to not targeting dissent; as such, it remains very unclear whether targeting dissent, or not targeting dissent, will be riskier. I then move to considering dMMI's education solution to problematic dissent. I argue that in many cases we have good reasons to doubt both that correcting false philosophies of science-informed policy will be effective, and that any approach heavily dependent on education will be sufficient and riskless in our non-ideal world. These arguments lead to further misgivings over the position that targeting problematic dissent is (generally) misguided and dangerous. I conclude, in §5.5, by showing that education activities are often not clearly distinct from strategies that target dissent, and suggest that this offers us some clues as to a way forward. I also provide guidance for reasoning about, identifying, and responding to problematic dissent, and raise what I see as important considerations for future scholarship.

I want to stress that I do not make a general argument that pushes toward it being generally misguided and dangerous not to target dissent. Nor do I argue that targeting dissent is always or generally sound, or that targeting dissent is never risky or harmful; under certain circumstances, it surely is. I am simply calling attention to the fact that there are cases where targeting dissent is not misguided and dangerous, and where doing so will be a sound, reasonable, and legitimate response. I hope to push the discussion toward the development of a more contextual and nuanced approach to evaluating and responding to dissent, one that attends to both epistemic and non-epistemic values and consequences and takes an all-things-considered perspective.

## 5.1. What is 'Targeting' Dissent?

According to dMMI, 'targeting dissent' is any attempt to minimise or discourage dissent that is perceived to be problematic (p. 593). They draw attention to three common targeting practices:

First, dMMI say that dissent can be masked, when, for example, the full range of opinions that exist on some matter are not reported, or when a consensus position is emphasised (pp. 596–597). As an illustration of the former, dMMI cite the example of scientists presenting only those general claims they all agree on to the public and policymakers, whilst simultaneously not conveying information about other claims they disagree about, or downplaying their significance. As an example of the latter, they point to efforts to quantify and highlight the scientific consensus on the existence of anthropogenic climate change.

Second, dMMI point out that dissent that is perceived to be problematic can be silenced (p. 597). According to dMMI, an example of silencing occurs when dissent is 'hindered through the peer-review process' (p. 597). A prominent example of an explicit silencing of dissent within an academic journal was the *South African Medical Journal*'s announcement that it would no longer accept articles containing certain dissident views on AIDS.[65]

Third, dMMI note that dissenters can be discredited, when, for example, their financial and political ties to private industry or partisan think tanks are publicised (p. 598). Oreskes and Conway's research on the 'manufacturers of doubt' is raised again as an example of this practice. According to dMMI, not only does discrediting dissenters hinder dissent directly, but it also does so indirectly by deterring funding agencies, publishers, and other scientists, from engaging with dissidents' work, or conducting, funding, or otherwise supporting, similar research.

These explanations and examples are not exhaustive of the kinds of activities that count as masking, silencing and discrediting dissent and dissenters. I understand these catego-

---

[65] See van Niekerk (2003). Thank you to Katherine Furman for providing this example.

ries to also include things like: the exclusion of some dissent from curricula, or emphasising a consensus position rather than providing 'balanced' lessons; non-academic media refusing to publish certain dissent, such as the decision of the Los Angeles Times to no longer publish letters claiming there is no evidence for anthropogenic climate change (Thornton 2013); and journalists discrediting scientists by revealing fraudulent scientific practices and undisclosed financial interests, as in Brian Deer's investigations of vaccine dissident Andrew Wakefield (Godlee, Smith and Marcovitch 2011). Nor are the three categories — masking, silencing and discrediting — a comprehensive taxonomy of what constitutes targeting dissent; according to dMMI they simply represent three particularly common ways dissent is targeted.

Now that we are clearer about what counts as targeting dissent, we can move to critically engaging with and evaluating dMMI's arguments.

## 5.2. Is it Misguided to Target Problematic Dissent?

The first of dMMI's two key arguments is that treating dissent about science as an obstacle to the development and enactment of public policy, and consequently targeting this dissent, is *misguided* (pp. 599–607). dMMI contend that those who target dissent err because the means they deploy will not achieve their end. This is so, dMMI reason, because the decision to target dissent rests on unjustified assumptions about the nature of scientific consensus and dissent, and the role of each in policymaking (p. 595). They focus on two assumptions — a scientific consensus is *necessary* (pp. 599–604), and a scientific consensus is *sufficient* (pp. 604–607), for sound policy — and claim that concerns about dissent are *only* plausible under *these* assumptions (p. 599). dMMI argue that both of these assumptions are false (p. 609), and therefore that targeting dissent is misguided because it will not actually facilitate sound policymaking. I will refer to these assumptions about the role of a scientific consensus in policymaking as *false philosophies of science-informed policy*.[66]

---

[66] dMMI note other false philosophies of science-informed policy e.g., that: science needs to be *certain* (p. 600); *unanimous* expert consensus is required (p. 601); scientific knowledge is the *only* factor that can

According to dMMI, it is the *prevalence* of *these particular* false beliefs amongst non-experts that is the most fundamental issue (p. 604), and therefore the 'real problem' (p. 609); the dissenting communicative acts themselves are, dMMI assert, the wrong targets of concern (p. 599). They claim:

> If our arguments are correct, the problem is *not* with the dissent as such, even dissent that is not epistemically valuable, but with an *incorrect* understanding of the nature of consensus and dissent and their role in sound public policymaking. Addressing the *real* problem should be our aim (p. 609, emphasis added).

Note that dMMI are making *causal claims* here, including a claim about which causal factor is the predominant or most fundamental driver of what they call unsound policy (i.e. an *attribution* claim). Based upon this understanding of the causal mechanism, dMMI propose an *alternative solution* to the concerns of those who currently target dissent: educating policymakers and the public about nature of scientific consensus and dissent, and the role each can and should play in policymaking (p. 595, p. 604; 2013). dMMI believe that this alternative will not only successfully facilitate sound policymaking, but that education is a 'better strategy for addressing the problem' because 'it avoids the additional dangerous consequences associated with targeting dissent' (p. 604).

Without doubt, dMMI are right to highlight the way that flawed conceptual schemes can distort decision–making. By pointing to the way that false philosophies of science-informed policy can contribute to generating the consequences of problematic dissent, dMMI have made an important contribution to identifying the sorts of misconceptions that should be corrected, and ideas that should be covered, within education on science and society. They have also identified *one high-level strategy* that could be implemented to enhance non-experts' resilience to problematic dissent, and that may be efficacious in dampening its harmful consequences.

---

legitimately influence policy (pp. 606–607). I don't focus on these false philosophies for two reasons: 1) they aren't dMMI's focus; 2) I take it that they are caricatures of science-informed policy that few, if any, contemporary scholars endorse.

However, dMMI's argument that targeting dissent is (generally) misguided is unsuccessful. This is because there is an alternative set of assumptions that capture several considerations that dMMI overlook and that together make it sound to target problematic dissent *in some circumstances*. I show this via the Probability Argument.

### 5.2.1. The Probability Argument

The Probability Argument can be summarised as follows:

P1.   Where a scientific hypothesis (H) is pertinent to policy (P), a person's level of support (or otherwise) for P can be influenced by their perception of the probability (Pr) that H is correct.

P2.   Non-experts rely on experts' testimony, as well as higher-order evidence — such as the *degree*[67] of expert agreement about Pr(H) — to form beliefs about Pr(H).

P3.   Problematic dissent can causally contribute to non-experts holding false or inaccurate perceptions of experts' first-order claims, as well as of higher-order evidence. This affects a non-expert's ability to form well-calibrated beliefs about Pr(H).

P4.   Targeting problematic dissent can prevent this speech from entering and circulating through the public knowledge system, or certain parts of this system, and/or can reduce its regularity and reach, and consequently its impacts. As such, targeting dissent can help some, perhaps many, people to more accurately perceive first- and higher-order features of the epistemic landscape, and hence, to form more accurate beliefs about Pr(H). Given the relationship between Pr(H) and P (see P1), some, perhaps many, peoples' attitudes towards P will therefore differ when certain problematic dissent is targeted, compared to a counterfactual world in which this problematic dissent is not targeted and perhaps even unwittingly helped to flourish.

---

[67] Note that dMMI generally treat consensus as a binary, presence-absence phenomenon. I think it's more helpful to understand consensus as a scalar concept, and to talk about the *degree* and *spread* of expert (dis)agreement.

C.    Targeting problematic dissent is a sound strategy for curbing the negative conse-
      quences of problematic dissent on policy in some, perhaps many, circumstances.

In what follows I discuss and defend each of these assumptions.


**(P1) The significance of the perceived probability that a pertinent hypothesis is cor-
rect**

In democratic societies, the development, enactment, and continuity of public policies
(Ps) rests in part, and often to a considerable degree, on the extent to which potential
voters (and to a lesser extent the public in general) *support* P. The extent to which an
individual supports or opposes P partly reflects their subjective assessment of how well
*justified* P is (Druckman and Lupia 2000).[68] This assessment partly turns on whether,
and to what extent, they: believe that there is a need for P; support P's objectives; take
P to be capable of meeting these ends; and are willing to accept trade-offs associated
with P. These beliefs and attitudes are influenced by a person's values and normative
views, in concert with their understanding of how the world is, forecasts about how the
world could be, and best estimates of the likelihoods of such scenarios. For some policy
issues, scientific information will be pertinent.

Where a scientific hypothesis (H) is relevant to P, how probable a person takes H to be
— $Pr(H)$ — can *influence* the person's assessment of how well justified (or not) P is.[69] For
instance, whether, and the degree to which, exposure to second-hand smoke can elevate
risks to non-smokers' health will be relevant to many peoples' subjective assessments of
how well justified a policy that prohibits smoking in workplaces ($P_{PSW}$) is.[70] Therefore,

---

[68] By 'support' I mean that that person would then act in ways that further P, rather than opposing P and
behaving in ways that prevent or reverse P, or at least that their support of P would feature as such in
their deliberations about all-things-considered decisions. I take 'justified' to mean underpinned by good
or legitimate reasons.

[69] Consistent with the role probability plays in rival descriptive theories of decision (i.e. expected utility
theory and prospect theory), as well as in key philosophies of mind (e.g. Clark 2015).

[70] The tobacco industry's own research noted this. *The Roper Report* (1978, cited by CDC 2006) reported
the widespread public belief that 'What the smoker does to himself may be his business, but what the
smoker does to the non-smoker is quite a different matter'. It also predicted that, as the belief that second-
hand smoke exposure could harm non-smokers became more widespread, public support for smoking
restrictions would grow.

the probability that a person assigns to the following hypothesis, for example, can influence their assessment of how well justified $P_{PSW}$ is, and thus whether, and to what extent, they support or oppose $P_{PSW}$:

$H_C$. Exposure to second-hand smoke in workplaces increases a non-smoker's risk of lung cancer, on average, by 15–20%.[71]

*Ceteris paribus*, the higher a person perceives $Pr(H_C)$ to be, the more justified they will take $P_{PSW}$ to be, and the stronger their support (or the weaker their opposition). Importantly, it is possible that once some $Pr(H)$ reaches a certain threshold, a person may then take P to be justified enough to support, or at least be neutral towards, P, all-things-considered. What this threshold is will depend on a person's values, among other factors, and will therefore be unique to each individual. Thresholds are also sensitive to contextual factors, e.g. may shift in response to changes in the status of other hypotheses.

To illustrate, take León, who thinks that it is unlikely that second-hand smoke exposure can increase the average non-smoker's risk of lung cancer by 15–20% (i.e. he takes $Pr(H_C) = 0$–0.33). This, in conjunction with his strong normative beliefs about the importance of smokers' negative freedom and of government interference being undesirable, means that he initially *strongly opposes* $P_{PSW}$. By contrast, Frida is much more accepting of government regulation, and has a very low tolerance for risks of harm being imposed on others. She also takes $Pr(H_C) = 0$–0.33, but given her values she *weakly supports* $P_{PSW}$. Now imagine that at some later time León and Frida both come to accept that $Pr(H_C) >= 0.95$, i.e. that it is extremely likely that $H_C$ is correct. With these updated beliefs (their values remain unchanged), León now *weakly*, and Frida *very strongly*, supports $P_{PSW}$.

**(P2) Non-experts depend on expert testimony and higher-order evidence to form beliefs about Pr(H)**

---

[71] Of course, I am not saying that this is the only hypothesis or consideration someone could take to be important to their deliberations about P. I isolate one hypothesis here for illustrative purposes; ultimately many considerations are entangled in a complex web.

How León and Frida form their beliefs about $H_C$ is important. When H relates to a domain of specialist knowledge, non-experts rely on expert testimony to provide them with the best knowledge about H available, including estimations of Pr(H).[72] Most straightforwardly, non-experts can defer to experts' assertions such that they directly employ experts' claims as premises in their own reasoning. But non-experts can also form impressions about Pr(H) more indirectly, e.g. León and Frida may have originally estimated $Pr(H_C) = 0–0.33$ because the media's reports were frequently along the lines of: 'whilst some medical scientists say it is virtually certain that exposure to second-hand smoke increases the risk of lung cancer in non-smokers by up to 20%, they remain deeply divided, with significant numbers of experts who dispute this claim, and who think there is zero chance this is the case.'

Non-experts also use higher-order evidence — that is, evidence about putative experts and the processes they used to arrive at their claims — to form their beliefs about Pr(H). Higher-order evidence is often conveyed to non-experts by intermediaries, such as the media, when communicating putative experts' first-order claims. It includes (but is not limited to): the degree and/or spread of expert (dis)agreement; who agrees/disagrees, and why; indicators of contenders' competence, vis-à-vis the specific hypothesis; relevant potential interests and biases of contenders; indicators of honesty, and track records of integrity; how and why any consensus formed; the degree of meta-expert agreement, and other indicators of meta-expert and interactive expert appraisal (Goldman 2001; Coady 2006; Anderson 2011; Coady and Corry 2013, pp. 22–34; Lane 2014). That non-experts use higher-order evidence in this way also has empirical support. For example, several studies have demonstrated the causal role of the perceived degree of expert consensus in the belief formation of non-experts (Lewandowsky et al. 2013; van der Linden et al. 2015; et al. 2017; Hamilton 2016; Maibach and van der Linden 2016). These studies found there to be greater convergence between non-experts' and experts' beliefs about the (non)existence of empirical phenomena as non-experts' perceptions of the degree of expert agreement became more accurate.

---

[72] I simply take an expert to be someone whose competence in a domain is secure. They need not be a professional scientist.

When experts disagree about Pr(H), or when a non-expert perceives there to be expert disagreement, higher-order evidence plays an especially important role in helping non-experts to adjudicate between conflicting first-order claims. Upon consideration of higher-order evidence, if a non-expert perceives some putative experts or the processes they used to lack credibility or trustworthiness, they may discount their testimony and take Pr(H) to be higher or lower than stated. This is positive when it helps non-experts to form more accurate beliefs. Perhaps León later accepted that $Pr(H_C) >= 0.95$ partly because he came to understand that only a handful of putative experts accepted $\neg H_C$, and that, although scientists, they lacked the necessary types of domain-specific expertise to competently assess the likelihood that $H_C$ is correct. These evaluations, however, can go awry when the putative higher-order evidence non-experts base their assessments on is itself misleading, inaccurate, or false, as non-experts may then mistakenly discount, buy into, or inflate some testimony.

Critically, given the relationship between a person's beliefs about Pr(H) and their attitudes towards P, a non-expert's awareness and perception of the epistemic landscape, which includes higher-order evidence, can influence the extent to which they support or oppose P. Misperceptions of key features of the epistemic landscape can be difference making to a person's strength of support or otherwise for P, and in some cases, even to whether or not they support P, all-things-considered. For example, if a person misperceives the degree of expert agreement about some Pr(H) to be 'X', when it is in fact 'Z', it is possible they might fail to support P when a degree of agreement 'Y' about Pr(H) would have been sufficient for their support. Or even if they correctly perceive the degree of expert agreement, they may misperceive this view to be that Pr(H) is 'X', when it is in fact 'Z', and they would have supported P if Pr(H) was at least 'Y'.[73] Thus, when non-experts' perceptions of the epistemic landscape are distorted — e.g. when they believe that there is much weaker expert agreement about an empirical matter than there actually is — this can affect the probability they assign to H or $\neg H$, and consequently, their evaluations of P and actions with regard to P and related issues.

---

[73] This is reflected in studies that have found that for some policy issues, presumably those where empirical information is highly relevant, policy attitudes are linked to perceptions about the degree of expert agreement about pertinent empirical facts (e.g. see Ding et al. 2011).

dMMI consider something close when they say: 'Presumably, the idea is that in order for a science-based policy to be justified, there must be good reason for thinking that the science is more likely to be true than to be false. If the existence of scientific consensus is thought to indicate that a theory is likely to be true, then the necessity of consensus for adopting a policy would seem to follow' (p. 600). But they reject this because 'it is not clear that a consensus is the best indicator of truth. Presumably, what is important is not the existence of a consensus per se, but rather the fact that consensus was formed in epistemically appropriate ways' (p. 600). However, note that on my account, $Pr(H)$ need not be >0.5 (i.e. 'more likely to be true than to be false') for a person to consider P to be justified (as illustrated by Frida's initial support for P, despite $Pr = 0$–0.33). Additionally, if a consensus *has* formed in an epistemically appropriate way, then the degree of consensus *is* one piece of meaningful higher-order evidence (Coady 2006), as dMMI themselves acknowledge.

## (P3) Problematic dissent can contribute to non-experts holding false or inaccurate beliefs

A non-expert's ability to form beliefs about $Pr(H)$ that are apportioned to the actual evidence can be adversely affected by problematic dissent in at least three key ways. First, when the erroneous claim 'X', an instance of problematic dissent, is asserted about $Pr(H)$, a non-expert may form the belief 'X' by straightforwardly deferring to this assertion. They may do so because 'X' is made to seem plausible, and the putative higher-order evidence presented makes the testifier appear credible and trustworthy and that there are no additional reason/s to reduce one's credence or hold a different belief. Alternatively, 'X' may be the only claim about $Pr(H)$ available, or the most prominent, within that person's socio-epistemic environment[74]; there is some evidence that we tend to believe that those propositions that are more familiar to us are more likely to be true (Schwarz et al. 2007). Or, a non-expert might accept 'X' because they are: engaged in motivated reasoning (Kunda 1990; Kraft, Lodge and Taber 2015); making inferential

---

[74] This could be due to acts of omission. Under such circumstances, non-experts may not be given fair opportunity to form accurate beliefs about $Pr(H)$. I put this issue to one side for this chapter to focus on actual communicative acts.

errors due to confirmation bias (Nickerson 1998); or because the assertion of 'X' triggers other cognitive phenomena such as cultural cognition (Kahan et al. 2011).

Second, a non-expert might form the inaccurate belief 'Y' about Pr(H), by using the variation between putative experts' competing claims (e.g. 'Z', which is correct, and 'X') to arrive at some value for Pr(H) in between what they perceive to be two equally credible claims. In such cases, dissenting speech, or the way it is presented, modifies non-experts' perceptions of Pr(H) more indirectly by influencing their perceptions of higher-order evidence. When a problematic dissenting claim is asserted, it can be done so in a way that misrepresents (exaggerates) its epistemic credibility, and/or misrepresents (understates) the epistemic credibility of opposing claims. Methods can be framed as reliable when they are unreliable. On the other hand, legitimate scientific practices can be framed as suspicious or duplicitous, when they are in fact uncontroversial. Putative experts can be falsely accused of scientific fraud, or they can be engaged in fraudulent activity without non-experts being any the wiser. How qualified an interviewee is to comment on the status of a particular scientific hypothesis can be exaggerated, or on the contrary, downplayed. The degree of expert (dis)agreement can be skewed; intermediaries and informants can engage in 'false balance' (Boykoff and Boykoff 2004), or plain bias (Manne 2011, 50–54), and thus lead non-experts to perceive experts to be more divided than they really are, and so on.

Third, some dissent can misrepresent experts' testimony about first order facts whilst exploiting genuine higher-order evidence to lend credibility to the false or inaccurate statements made about experts' first-order claims. For example, early in 2013, the *Australian*, Australia's widest circulating national newspaper, ran a front-page story with a headline that asserted: 'Sea rise "not linked to warming"'.[75] This article was supposedly based on the findings of a team of climate scientists, Gregory et al. (2013), and commenced with the pronouncement that: 'The latest science on sea level rises has found no link to global warming.' But this claim was patently false; the article grossly misrepresented Gregory et al.'s research and conclusions, as well as the state of scientific

---

[75] The article no longer appears online, but a copy of the print version can be viewed at: https://doi.org/10.6084/m9.figshare.5481100.v1

knowledge about sea level rise more generally. Dr John Church, one of the study's authors, said of article: 'Sea level clearly is linked to climate change, it clearly is linked to greenhouse gases and that was in the paper quoted by the *Australian*. The quote is, I am sorry, inaccurate' (quoted in Creagh 2013).

Whilst in this case the *Australian* retracted the online version of this article and issued a correction in print, this isn't necessarily sufficient. Retractions and corrections are often ineffective because it can be difficult to un-do false or unsupported beliefs once they have taken hold (Gilbert et al. 1990; Thorson 2015). Discredited and corrected misinformation can also have a powerful ongoing effect on people's memory and reasoning (Lewandowsky et al. 2012), and moreover, sometimes even make matters worse: the 'backfire effect' suggests that correcting false beliefs can work to further entrench those beliefs (Kuklinski et al. 2000; Nyhan and Reifler 2010; Nyhan, Reifler, and Ubel 2013). Corrections can also result in misinformation becoming more salient, which is problematic because, as noted earlier, we tend to believe that propositions familiar to us are likely to be true. This all only serves to worsen the impact of problematic dissent, and of course means that those who actually do intentionally mislead others for their own ends have additional incentives to propagate misinformation in the first place.

It's not hard to imagine then, how certain types of dissent could have led León and Frida to initially perceive expert opinion on $H_C$ to be deeply divided when *in fact* there was a very high degree of expert agreement that $H_C$ was *virtually certain* (i.e. Pr = 0.99–1), and there were good reasons to think this strong consensus had formed in epistemically appropriate ways. Owing to this dissent, León may have failed to support a policy that he would have at least weakly supported, absent this distortion in his understanding of the epistemic landscape.[76] But most importantly, notice that León and Frida *did not hold false philosophies of science-informed policymaking*: they did not take a scientific consensus per se to be necessary or sufficient to justify policy (as evidenced by Frida's initial support of $P_{PSW}$ despite her perception that experts were divided). Nor did they think

[76] This also serves to illustrate that the greater the misalignment of one's *perception* of certain facts and the *actual* facts, the more likely it is that one will be less successful in achieving their own goals, or living their life in a way that is consistent with their values. This is likely to be morally and politically significant when deception is involved, or when the consequences of holding a misperception fall on a person's dependent or on other people. I do not explore this particular issue further here.

that a unanimous expert consensus was required (both supported $P_{PSW}$ when they still perceived there to be experts who rejected $H_C$), that the science needed to be certain (both supported $P_{PSW}$ whilst perceiving there to still be a 0–5% chance $H_C$ was incorrect), or that scientific knowledge was the only factor that could legitimately influence their policy preferences (both Frida and León's values influenced their preferences). The issue here (particularly in the case of León, who initially did not support a policy he would have otherwise supported) was instead that they were unable to form well-calibrated beliefs about $Pr(H)$, primarily because they were unable to accurately perceive the degree of agreement among experts and other higher-order evidence. This was either because they did not have access to the requisite information and evidence, or because they personally lacked the capabilities required to make competent judgments about the epistemic credibility of competing claims and evidence. Being embedded within socio-epistemic environments that failed to provide this evidence and that was rich in misinformation and distortions, they could not distinguish between dissent that did present a real challenge the consensus position and dissent that did not.

**(P4) Targeting problematic dissent can assist non-experts to form accurate beliefs**

I'll now explain how the three strategies dMMI raise as being commonly used to target dissent can assist some (perhaps many) people to more accurately perceive epistemic features pertinent to their practical reasoning. First, imagine a psychology lecturer who chooses to initially *mask*: (1) the discredited claim that non-heterosexuality is itself a mental illness; and (2) the presence of a minority of psychiatrists who dissent from the (now) expert consensus, and official position of societies like the American Psychiatric Association, that non-heterosexuality is not a mental disorder. She chooses to do so because she has good reasons to suspect that many of her students have underlying prejudices towards non-heterosexual people, and as such is concerned that raising this dissent early within the course might trigger unconscious biases, making it harder to achieve learning objectives. Additionally, when she does eventually introduce her students to the past controversy and the presence of a dissenting minority who still believe non-heterosexuality should be classified and treated as a mental illness, she does not teach it as a 'balanced debate' with two equally credible sides; she does not frame the dissidents' position as being on an epistemic par with that of their opponents (nor a

moral one). Based on her experience, her sense is that these measures, in this context, facilitated better engagement with the issues discussed in class and likely helped at least some of her students to form more accurate beliefs than they otherwise would have if she had raised this dissent in the first lesson, or had not emphasized the consensus position throughout the course in certain ways.

Now let's consider how the *South African Medical Journal*'s (SAMJ) *silencing* of certain dissident views on AIDS from that particular forum helped non-experts to form more accurate empirical beliefs. At the time SAMJ made this decision, in 2003, South Africa's AIDS public health crisis was at its peak, with approximately one in four adults with HIV. Its President at the time, Thabo Mbeki, publicly denied the link between HIV and AIDS, as well as the efficacy of anti-retroviral (ARV) medication. He instituted policies denying HIV-positive citizens ARVs in parts of South Africa, and publicly justified and tried to legitimate these policies with reference to what he had accepted as empirical facts. If, in such circumstances, South Africa's most eminent medical journal had continued to publish dissenting articles denying the well-established link between HIV and AIDS, it would have been conferring credibility on Mbeki's (and others') empirical assertions, and tacitly recommending the need for non-experts to give serious consideration to such views. Instead, by indicating that such articles no longer had a place in its journal because the evidence for the link between HIV and AIDS was so overwhelming, the SAMJ was providing non-experts with an important piece of higher-order evidence: that the experts associated with SAMJ did not take claims denying the link between HIV and AIDS to merit serious consideration. Such a move plausibly contributed, either directly or indirectly, to more South Africans coming to doubt the credibility of Mbeki's assertions, and how well justified his policies were, than they otherwise would have if the SAMJ continued to offer up their journal as a forum in which AIDS dissidents could express their views.

Finally, much of the scholarship dMMI characterize as a targeting strategy to *discredit dissenters* has in fact publicly revealed and discussed critical higher-order evidence, some of which had previously been actively suppressed and concealed from the public. Through the work of the scholars dMMI reference, such as Proctor, Michaels, Oreskes

and Conway, Shrader-Frechette, and Elliott (among others), we now have a deeper understanding of the interests and biases, strategies, and misconduct of players in so-called scientific controversies.[77] Likewise, high-quality investigative science journalism in some cases has brought relevant higher-order evidence to light to be scrutinised by the public. Importantly, these scholars and journalists themselves often have various types of expertise — e.g. meta-expertise and interactive expertise[78] — and frequently have unique access to putative experts and evidence. Their scholarship and journalism can therefore be understood as a type of testimony too: about the credibility of different experts and their claims. Sometimes implicit within this testimony is their judgment that a consensus has been reached in epistemically appropriate ways. As such, it is valuable additional higher-order evidence that non-experts can use in forming their own beliefs.

Returning again to León, to the extent that his impression of the epistemic landscape was initially distorted by unsubstantiated dissent, or dissent that misrepresented others' positions or suppressed higher-order evidence, then if counterfactually this speech had been judiciously targeted in a way that successfully reduced its frequency and intensity, or its perceived credibility, it's possible — even probable, perhaps — that León would have perceived significant features of the epistemic and policy landscapes more accurately. Given the relationship between a person's perception of first- and higher-order facts and their evaluation of how well justified policies are, León's initial attitude toward $P_{PSW}$ most likely would have been different; in this case, more supportive. It's also likely that he would have formed his policy attitudes and preferences with greater levels of autonomy, given my arguments in Chapter Three.

For certain issues, these seemingly small differences in each person's perceptions, when combined in aggregate with all similar changes across a population, can result in a significant difference to the trajectory of policy, as I argued in Chapter Four. Take, for example, the 2003 Iraq war, a decision that significantly turned on, and was publicly justified with reference to, the alleged status of just a few key empirical claims, most

---

[77] Leuschner (2016) makes a similar argument.

[78] Following Collins and Evans (2007), I take meta-expertise to be expertise used to judge other expertises and experts, and interactive expertise to be having necessary and sufficient knowledge and skill within a domain such that one can engage with contributory experts fluently, and in sophisticated ways.

famously that Iraq possessed and was developing weapons of mass destruction. In a 2003 study, Kull, Ramsey and Lewis found that the majority (60%) of U.S. citizens held at least one of three key misperceptions about empirical features that were cited by politicians to legitimate their decision to go to war; only 30% had no misperceptions. Moreover, they found that these misperceptions were highly related to peoples' attitudes towards the Bush Administration's decision to go to war, both before and after the war commenced: the presence of misperceptions was the most powerful predictor of support for the war, with those misperceiving 4.3 times more likely to support the war than those who did not misperceive. Further still, the presence of each additional misperception was associated with sharply higher support for the war (one misperception = 2.9 times more likely; two misperceptions = 8.1 times more likely; three misperceptions = 9.8 times more likely). It is plausible that it would have been significantly more difficult for Bush and other politicians to elicit and maintain support for the decision to go to war if the public had not held such misperceptions. By giving citizens incorrect and misleading information, support was gained for a policy. Whilst this example is not from science, I think it helps us to be more imaginative about the likely cumulative consequences on public policy when large proportions of the population within democracies hold significant misperceptions about key empirical matters.[79]

**(C) Targeting problematic dissent can be a sound strategy to curb its negative policy consequences.**

To summarise: problematic dissent can interfere with non-experts' abilities to accurately perceive the contours of epistemic landscapes. This can in turn hinder a non-expert's ability to form accurate beliefs pertinent to their practical decision-making. These inaccurate beliefs mean that some peoples' subjective evaluations of how justified a given policy is differ from what they would be in a counterfactual world in which this dissent was absent or reduced; *this can be true even when they do not hold false philosophies of science-*

---

[79] This was the best data-backed example I could find to demonstrate this point. Not all public policies will have such a tight relationship to empirical features, but many do, and are reasonably perceived as such by significant proportions of populations (e.g. prohibitions on smoking in public places seems to be one).

*informed policy*. Targeting problematic dissent can prevent it from entering and circulating through the public knowledge system, or certain parts of this system, and/or can reduce its regularity, reach, and impact. As such, targeting dissent in pursuit of policy-related ends will not be misguided in some, perhaps many, circumstances. The Probability Argument therefore provides a *prima facie* case, and a *pro-tanto* reason, for targeting problematic dissent about science. Whether or not one should target dissent 'all-things-considered', and if so, how, is a different matter, which I will discuss in §5.6.

### 5.2.2. Responses to two objections

Here I raise and respond to two objections. The first is that even if the Probability Argument shows that targeting dissent can sometimes influence the trajectory of policy, this trajectory will not necessarily bend towards 'sound' policy (remember: 'sound' policy is what dMMI take to be the aim of those who target problematic dissent). For example, if one's definition of 'sound' climate policy is holding global warming to one degree Celsius, it seems unlikely that targeting problematic dissent about empirical claims would help to secure this end. However, if one took a much wider view of what counts as 'sound' climate policy – e.g. for 'action', broadly construed rather than 'inaction' – emerging evidence suggests that more accurate beliefs about higher-order evidence, such as the degree of scientific agreement about the existence of anthropogenic climate change, is an important gateway belief associated with increased support for policy action across the political spectrum (van der Linden et al. 2015).

Alternatively, someone might not be interested in the trajectory of public policy at all, nor in the content of policy being in some objective sense 'sound'; they might question whether such a notion even makes sense, as I argued in the last chapter. Instead, they might hold that what's most important in democratic societies is that when more people have better calibrated perceptions of the epistemic landscape, more people will be subjectively assessing policy as sound or unsound *by the lights of their own value system and reasoning*, rather than because they have distorted perceptions of empirical features. So, rather than being interested in the specific *content* of public policy, as dMMI presume, someone might instead be moved to target certain dissent because they are interested in

securing a particular type of policymaking *process*, perhaps even to secure the democratic legitimacy of science-informed policy, I argued in the previous chapter.

Intemann (2017) and Pearce et al. (2017) are also concerned about democratic deficits in policymaking caused by quarrels over empirical facts. However, they think that these debates can either by resolved by other means (e.g. Intemann suggests more discussion of evidence and mechanisms than consensus) or side-stepped (e.g. Pearce et al. advocate greater discussion of values and policy options) to make policy progress. However, I'm sceptical that these untested strategies will be as effective as these authors believe. For one thing, I think it's likely that the factual questions they hope to avoid will inevitably re-appear in these conversations. I also don't think they manage to escape notions of consensus. For instance, Pearce et al. (p. 4) claim that political progress was made on the Montreal Protocol largely due to the unexpected discovery of the Antarctic ozone 'hole' rather than the scientific consensus about long-term ozone depletion. But they fail to acknowledge that part of the reason why this was accepted as a 'hot crisis signal' is that there was a strong expert consensus that the 'hole' existed and was unlikely to quickly disappear. Similarly, Intemann (p. 201) urges scientists to communicate features of their theories or models that make them reliable or trustworthy, such as the ability to make novel predictions. But presumably a novel prediction only counts as such, and thus is meaningful in the way Intemann says it is, if the status of the prediction as 'novel' has been backed by their peers. If instead there were significant disagreements among relevant experts over whether or not it counts as novel, or whether the prediction could even count as having been confirmed, then it's not clear how reliable or trustworthy a non-expert ought to take a claim about a novel prediction to be.

The second objection I anticipate is that the Probability Argument is just the 'information deficit model' of science communication and policymaking in disguise. The central tenet of the information deficit model — a model widely rejected within the STS and science studies communities — is that any public scepticism or negative attitudes toward science, or a policy that has been informed by science, can be explained by people just not knowing enough; if they only knew more and better 'facts', they would accept the science, or support a policy that has been developed with reference to those

empirical claims. Because the Probability Argument is motivated by a focus on the communication of information and evidence, I can see how someone might think that it shares this same tenet. However, the Probability Argument does not hold that *every person's* doubt or negative attitudes towards science or policy can be *fully* explained by not knowing enough, or in some cases *at all*. It only suggests that in *some* circumstances *some* people's doxastic attitudes about a scientific hypothesis can be *partly* explained by misperceptions of key features of epistemic landscape, and that in many cases these doxastic attitudes would differ in degree, or kind, if those people were provided with additional information, or if misinformation was reduced or eliminated within their socio-epistemic environments. The Probability Argument suggests that in *some* cases doing so will influence *some* people's *degree* of support or otherwise for a particular policy. These nuances distinguish the Probability Argument from the 'deficit model'.

## 5.3. Is Targeting Dissent Dangerous?

So far, I have argued that dMMI do not succeed in establishing that targeting dissent is misguided, even generally. What they have identified, however, is *one* potential part of the problem (false philosophies of science-informed policy), and *one* aspect of a possible solution (education). I take my argument in the first part of this paper to be sufficient to re-open the debate about the soundness of targeting dissent. Moving forward, I suggest that we focus on developing a better understanding of the specific conditions under which targeting dissent, or not targeting dissent, will be misguided or sound. However, making such judgments is inevitably bound up with questions about the riskiness of the consequences of pursing different courses of action.[80] Therefore, I now turn to dMMI's charge that targeting dissent is *dangerous*, which I understand to mean that doing so can, or is likely to, cause unfavourable or harmful consequences. dMMI argue that targeting

---

[80] In §5.6 I point out that before we can make these judgments, we first need to better understand, and make a case for, our: final values; what values we are and are not willing to trade-off along the way; and what is morally permissible (or even demanded), in a particular context.

dissent is dangerous for three key reasons. Whilst I agree that all three reasons are genuine risks and legitimate concerns, I show that these same dangers can also be present in situations in which dissent is *not* targeted.

### 5.3.1. It is not clear that targeting problematic dissent leads to an increase in its frequency

The first reason dMMI provide to support their contention that targeting dissent is dangerous is that targeting dissent is likely to lead to a greater frequency of problematic dissent (p. 607). They suggest that this then signals to dissenters that contesting the science is an effective strategy, and thus provides them with an incentive to make misleading remarks, spread misinformation, and manufacture doubt. dMMI also claim that targeting problematic dissent suggests to listeners that the dissent targeted *does* appropriately challenge and weaken the consensus position, and therefore that the dissent targeted *should* be considered by non-experts when they are deciding what to believe.

However, it seems at least equally plausible, and perhaps even more plausible in some cases, that *not* targeting certain instances of problematic dissent could lead to an increase in its frequency; if nobody attempted to discourage or reduce the presence and efficacy of problematic dissent, there could be an even larger incentive to create and disseminate it. The presence of voices targeting problematic dissent may currently be operating as counter-vailing forces and disincentives, keeping the frequency and influence of problematic dissent (as well as its consequences) at reduced levels compared to a counterfactual world in which no targeting of dissent took place. Similarly, the targeting of some dissent about science could equally reflect precisely the opposite of what dMMI suggest: it could indicate to listeners that the dissent targeted *does not* in fact challenge or weaken the consensus position, and therefore that non-experts *should not* take it into account when forming their beliefs. I suggested that this was the case when the SAMJ refused to publish articles denying the link between HIV and AIDS. dMMI do not account for the way that claims allowed to circulate through the public knowledge system can become normalized merely by familiarity, and legitimised by the social endorsement they are tacitly given, especially when those actors are perceived to be authoritative or epistemically privileged in some way.

More evidence, especially empirical evidence, is needed to determine whether, and under what circumstances, targeting problematic dissent, or not targeting problematic dissent, is likely to lead to an increase in its frequency.

### 5.3.2. Not targeting problematic dissent can also create 'chilling effects'

dMMI also argue that targeting dissent is dangerous because doing so is likely to result in a 'chilling effect', silencing dissenting voices that genuinely ought to weaken a prevailing expert consensus (p. 608). These chilling effects, say dMMI, occur when scientists feel the need to self-censor to protect their careers and reputations, or when particular research and conclusions are suppressed in peer review, editorial processes, and funding decisions. Whilst there is no doubt that this type of chilling effect does occur, there is growing evidence that the presence of certain types of dissent about science has distinct chilling effects of its own; it too can lead to legitimate voices being stifled or silenced, or scientists being discouraged from making certain claims or researching certain topics (Kempner 2008; Biddle and Leuschner 2015; Lewandowsky et al. 2015; Sharman 2015; Leuschner 2016). Furman (2016) also helpfully provides an example from the policy-making arena: the way that the dissent of then South African President, Thabo Mbeki, about the cause of AIDS and the efficacy of ARVs, suppressed other voices (other dissent) within his party over policy. Where different chilling effects stand in a trade-off relationship, judgments will need to be made about which chilling effect is more problematic, and whose speech ought to be prioritized.

### 5.3.3. Targeting the targeting of problematic dissent also risks stripping away resources needed to guard against special interest science

Lastly, dMMI warn that targeting dissent can strip away the resources needed to guard against the influence of special interest science (p. 607). The thought is that because vested interests can sometimes manufacture an expert consensus to further their own objectives, dissent is critical to our ability to successfully challenge these interests; by targeting problematic dissent in Case X, we may be discouraging the very sort of critique required to oppose the influence of vested interests in Case Y.

However, not targeting problematic dissent, and also targeting the targeting of problematic dissent, could also discourage the very sorts of critique and actions required to oppose vested interests who manufacture dissent rather than consensus, or who are reckless and negligent when speaking about science. In doing so, these actions also seem to risk stripping away a key resource (i.e. the judicious targeting of problematic dissent) sometimes needed to guard against the influence of special interests. dMMI's argument here therefore seems more appropriate as a response to a position that *all* dissent in and about science is problematic and should be targeted, rather than my nuanced and particularistic position that *some* dissent is problematic, and that it can be sound, reasonable and legitimate to target *some* problematic dissent in *some* circumstances. The onus is on dMMI to show that even the judicious targeting of dissent in delimited circumstances will have the effect of stripping away resources in the way that they contend, and further, that their own position does not unjustifiably strip away a resource that is needed and legitimate in some contexts.

## 5.4. Education and Targeting Dissent

Based on dMMI's criteria for an action counting as 'dangerous', it seems that the jury is still out as to whether, in a particular circumstance, targeting dissent or not targeting dissent will be the risker option. But even granting this, and even if one accepts my arguments in §5.3. too, one might still argue that not targeting dissent simply ought to be preferred because it is *more likely* to be *more efficacious* and *less dangerous* than targeting dissenting speech, i.e. it will have a more desirable net effect, all-things-considered. Although I do think that this will be the case in *some circumstances*, in this section I aim to cast further doubt on this position as a more general conclusion by critically examining the soundness of, and the risks associated with, dMMI's proposed alternative: educating policymakers and the public to correct false philosophies of science-informed policy. I will also briefly explore challenges associated with education solutions more broadly.

### 5.4.1. It is unclear how efficacious education that corrects false philosophies of science-informed policy will be

The chief weakness of dMMI's alternative to targeting problematic dissent is that despite claiming that the prevalence of false philosophies of science-informed policy is the 'real problem' (p. 604; p. 609), dMMI do not provide any evidence to support their conclusion that it is *these specific* erroneous beliefs that are *significantly causally responsible* for the negative consequences problematic dissent generates. Particularly important here would be empirical evidence of what proportion of non-experts hold such beliefs, and how correcting these particular false beliefs in some sub-population can be reasonably expected to make a significant difference. As dMMI do not provide this information, what positive impact we can anticipate education to correct false philosophies of science-informed policy to have remains very unclear. This is especially because issues will persist even when people have perfect beliefs about science-informed policymaking if non-experts are still unable to perceive various empirical features relevant to their decision-making with adequate accuracy, as I argued in §5.3.

A further challenge for dMMI's proposal is that it is also possible for problematic dissent to be deployed to undermine non-experts' understanding of science-informed policy. This extends Oreskes and Conway's observation that a key aspect of the Tobacco Strategy has often involved spreading erroneous understandings of scientific practices (2010, p. 34, pp. 267–274). Should the focus turn to educating non-experts about science-informed policy, this is probably what we should expect to happen. Notably, the infamous 'Luntz Memo' included talking points on this front: e.g. he recommended that Republicans emphasize 'the importance of acting only with <u>all</u> the facts in hand' (emphasis in the original) (Luntz 2003, 137–138). dMMI's solution may therefore simply involve pushing the problem back a step, and it's not clear that the problems associated with misinformation and misrepresentation can be overcome without targeting at least some problematic dissent, at some level or stage, or within at least some parts of the public knowledge system.

### 5.4.2. Challenges for any education solution

However, some might still try to maintain that there remains a way we can educate away the problems to avoid the negative consequences of problematic dissent. Towards the

end of their article, dMMI deviate from their main proposal (correcting false philosophies of science-informed policy through education) to claim that we should simply educate non-experts in such a way that *they themselves* are able to 'distinguish between dissent that is appropriately consensus-undermining and that which is not' (p. 604); remember, this is what I noted as being León's fundamental problem (p. 23). While it is unclear what kind of education dMMI envisage here, note that if this was achieved the third premise of the Probability Argument — that some dissent can contribute to non-experts holding false or inaccurate beliefs — would fail. Such a solution seems, at first blush, to be unobjectionable; who would deny such capabilities for all? If everyone could successfully sift truth and falsity for themselves it would of course mean that it would be virtually impossible for anyone to be misled by others' problematic dissent. In fact, it would ultimately mean that there would be no such thing as 'problematic' dissent. But although this solution sounds ideal, once its deeper implications are probed it becomes clear that this suggestion does not adequately attend to several challenges and complexities associated with addressing problematic dissent through education in our non-ideal world.

Firstly, whilst it is currently wildly unclear what kind(s) and degree(s) of expertise people would need to competently evaluate the epistemic merits of dissenting speech, one possibility is that dMMI had in mind something like a mass education campaign to improve peoples' general critical reasoning skills. Scholars have been optimistic that the capacities needed to reason well about competing putative experts' claims are within reach for most people (e.g. see Anderson 2011), especially as they involve skills we already deploy when we assess everyday experts, like mechanics and doctors. Whilst I support such proposals in principle, I have several concerns about how realistic they are in practice, and therefore just how confident we should be about their ability to alleviate the consequences of problematic dissent in the absence of other measures. Firstly, in all likelihood, such efforts will be starting from a low baseline of skills. Wineberg et al. (2016) found that even highly educated and digitally savvy students at Stanford University have very poor reasoning skills in evaluating the credibility of online sources. Of course, it could be the case that those with less formal education are actually better at these tasks. But without this evidence, such findings should at least give us moment for pause. Ensuring everyone develops sufficient capacity for even the most basic of critical tasks,

perhaps especially in societies with stark pre-existing educational inequalities, is likely to be extremely difficult.

Moreover, reaching large numbers of adults, most of whom are no longer in formal education, would be an enormous challenge (as dMMI acknowledge). One of the only extant forums for such education is the mass media. But large parts of the media have played, and continue to play, a significant role in generating, relaying, and amplifying problematic dissent (Boykoff 2011; Painter 2011). In media markets driven by profit that are subject to only minimal regulation, sensationalism, and even falsity, are incentivised. It is therefore hard to see the media being used to carry out any kind of civic education without first at least partly addressing the underlying issues in media culture and norms of public communication that are currently aiding and abetting the mass distribution of misinformation. Even if an alternative channel or forum is created, its educative function could be undermined by the parallel existence of an unreformed network of influential entities that continue to transmit enormous amounts of misinformation and frequently misrepresent others, including educators. But addressing media culture and norms of public communication, including issues associated with social media, is likely to require interventions that count as targeting problematic dissent. For example, while Robert Manne (2011, p. 114) calls for media outlets and journalists to engage in 'courageous criticism' of rivals that practice unethical reporting, even this solution — a form of informal 'industry self-regulation' — could plausibly count as a type of 'silencing' or 'discrediting' of dissenters.[81] Additionally, even possessing the best critical skills does not solve the problem of having adequate access to the right kinds of higher-order evidence to make use of these skills. And higher-order evidence, as I explained in §5.3, can be actively, negligently or recklessly suppressed or distorted, often by the very intermediaries we reasonably rely on to provide this evidence to us.

Lastly, but perhaps most importantly, suggestions that narrowly focus on, and propose improvements via education to individuals' own cognition, in the hope that they will

---

[81] This is precisely how the *Australian* itself perceived Manne's critique of its practices and Manne's calls for others to engage in 'courageous criticism' of the *Australian* (e.g. see reply to Manne by its Editor in Chief titled 'A critic untroubled by facts who seeks to silence dissent' (Mitchell 2011). Ironically, Mitchell ends the article by calling on editors at Fairfax and the ABC not to publish anything written by Manne.

then be capable of sifting truth from falsity about specialist domains of knowledge *themselves*, risk being overly individualistic. Such solutions neglect the socially embedded, and collective, nature of human cognition and knowledge, and overlook several possibilities: that we may have certain epistemic or informational distributive entitlements (Coady 2010) that can be interfered with by certain types of dissent; that we could be subject to a type of epistemic injustice or epistemic oppression when our positive freedom to know something is restricted by particular dissenting communicative acts (as per my chapter 3); that actors who occupy certain nodes within the public knowledge system may have special obligations to supply certain types of information, or to refrain from propagating misinformation or misrepresenting evidence, and that people can be wronged when these actors fail to discharge these responsibilities. If any of these possibilities hold, then the fundamental defect should not be viewed as residing primarily, or even at all, *internally* within individuals' own heads, but *externally* in the socio-epistemic environments they are embedded in, and upon which they reasonably depend to form beliefs about specialist domains of knowledge. Calls for non-experts to be generally left to make their own decisions about the accuracy of speech concerning such domains, and placing the burden on individuals to successfully sift truth and falsity, may even actively foster the abrogation of epistemic responsibilities and create room for exploitation, and political manipulation, to take place. This would in turn aid the consolidation and reinforcement of problematic power structures that condition features of the public knowledge system, which can in turn work to prevent, or considerably delay, social and political paradigm shifts.

Even where education solutions appear to be successful, neglecting the correction of malfunctions and unethical practices within socio-epistemic systems may leave underlying structural problems and institutional dynamics intact and dormant, ready to re-emerge when education efforts can no longer keep pace with science and technology (if they ever could). Similarly, even if the challenges identified thus far could, over the long term, eventually be overcome without recourse to targeting dissent along the way, it would surely take an extremely long time to get to the point where there were no longer any significant negative consequences associated with problematic dissent (if it is even possible to reach this point). By this time, many negative downstream consequences may have already occurred, or be too late to prevent. The impacts that can be traced to

certain dissenting communicative acts — including those non-epistemic consequences emanating from the failure or delay of certain policies being implemented, or actions being carried out — are the sorts of costs that need to be weighed against the anticipated benefits of not targeting problematic dissent. I further explain such evaluations in the next section.

## 5.5. Identifying problematic dissent and choosing a response

Based on my responses to dMMI and the arguments I have defended in this chapter, my conclusion — at least at this stage — is that the best *general strategy* to deploy to counter problematic dissent, is pluralistic: we should simultaneously implement education efforts, whilst also demanding and working towards improvements in the integrity and proper functioning of the informational ecosystems within which individuals are situated, and upon which they reasonably depend. However, what *particular strategy or strategies* we ought to deploy in any given circumstance will be *highly context dependent*. In *some circumstances*, the most efficacious and ethical response will be to target *some* problematic dissent in *some* way. In other circumstances it will be better to respond to problematic dissent by critically evaluating it, and/or by educating listeners on the place of such dissent in policymaking processes. However, we shouldn't think that this will always be desirable or appropriate (e.g. when this dissent itself is likely to silence others), nor always be possible or the first-best option given limited resources of time, space, attention, etc.

My suggestion of embracing a plurality of approaches, rather than just the single strategy dMMI advocate, is further supported by the observation that targeting dissent and education are often not distinct activities. Although we have, up until now, treated targeting dissent and education as separate, unrelated, and mutually exclusive, this is largely a false dichotomy, as these activities are often entangled. Nowhere is this more apparent than in the case of the regulation of speech about tobacco products, which has involved the state curtailing the tobacco industry's freedom to publicly dissent over certain empirical claims that have been accepted as being beyond reasonable doubt. The Final Opinion of the case *United States v Philip Morris USA* even went as far as permanently

enjoining (i.e. permanently *silencing*) the Defendants (tobacco industry representatives):

> from making, or causing to be made in any way, any material false, misleading, or deceptive statement or representation, or engaging in any public relations or marketing endeavor that is disseminated to the United States public and that misrepresents or suppresses information concerning cigarettes (2006, p. 3).

Tobacco companies are now also required by the state to include certain messages about health risks on their products and advertisements. Such measures are at once an example of targeting (silencing) dissent, and of education; the regulation of speech on tobacco products has been credited as one of the most far-reaching and effective aspects of education campaigns (Hammond 2011; U.S. National Cancer Institute and World Health Organization 2016). Thinking again of the examples of targeting dissent raised in this chapter, we can appreciate that each instance of targeting dissent simultaneously had an educative function, and therefore that it is often difficult to neatly separate targeting dissent from education.

Even so, one might worry that the pluralistic approach that I advocate assumes we can: (1) accurately identify 'problematic dissent', especially in the present when we need to act, rather than just in hindsight,[82] and (2) successfully select the most appropriate response/s to problematic dissent in any given circumstance. I have two replies to these very important concerns. First, despite the inherent uncertainties, we can sometimes be very confident, and justifiably so, about certain dissent being problematic, e.g. public dissent sponsored by the tobacco industry denying the link between smoking and cancer. Surely targeting this dissent, as we do today through regulating the tobacco industry's speech, has been and is still valuable, even if most of the damage was done before these regulations were put in place. And given that western tobacco companies are now focusing on expanding into new and emerging markets in low-income countries, where large proportions of populations are not aware of the health risks associated with smoking (WHO 2011), it's not clear that most of the damage has already been done on a

---

[82] Thank you to an anonymous reviewer for *Perspectives on Science* for pressing me on this point.

global scale. The WHO also reports that, based upon the past experiences of other countries, the governments of many of these countries are working towards putting in place measures to avoid the worst impacts of misinformation spread by Big Tobacco.

Second, it's possible that we currently under-estimate the extent to which we can make *good judgments* about dissenting speech about science 'in the moment'. We can evaluate, for example, whether it has had sufficient expression across time and place; whether it has received adequate attention and response; the likelihood that its expression will lead to epistemic or non-epistemic benefits, or harms; whether those expressing it have a claim to expressing it, or being heard, within a particular forum; or whether some dissenting speech might actually suppress the speech and dissent of others.

The alternative to attempting to make good judgments — i.e. 'not interfering' — also seems very unattractive. In many cases, such an approach would lead to endorsing the market as a superior, omnipotent information processor, and in our highly and increasingly resource stratified non-ideal world, this merely ends up prioritising the speech of those who can buy or otherwise capture the means of producing and transmitting ideas.[83]

Of course, we could, and most likely will, make mistakes about what is and is not problematic dissent and about how to appropriately respond to it; these judgments will necessarily be decisions made under uncertainty, and will involve epistemic and non-epistemic risks. As such, they will often be *value-laden decisions*. But so too are decisions to *not target* problematic dissent, and this has generally not been properly acknowledged within the debate so far. Insofar as decisions to consider dissent problematic or not, and

---

[83] Phillip Mirowski (2015, 2017) has characterised this as an expression of 'neoliberal epistemology'. Mirowski holds that neoliberalism *is*, first and foremost, a set of epistemological precepts, recruited in service of a political program. He traces this back to the Mont Pelèrin Society, which Karl Popper was a member of. But interestingly, in a section in *Conjectures and Refutations* (1972) called 'Some Practical Problems: Censorship and Monopolies of Publicity', which was a lecture Popper presented at the Mont Pelèrin Society, he lays out a series of questions that indicate that he was: aware that some speech was potentially problematic ('The problem of propaganda and advertisement versus the spread of information'); concerned about the social consequences of this speech ('The problem of propaganda for cruelty'); but that he had no answers about what to do in response to such speech ('No theses are offered in this section—only problems').

to target it or not, are informed by judgments about how bad the consequences of erroneously doing so are likely to be, there will be no value-free or value-neutral moves; those who advocate not targeting dissent, or the position that we should address bad speech with more speech or education, are often merely assigning hefty weight, and giving priority to, epistemic values, such as a greater probability of obtaining new knowledge, over non-epistemic risks and harms. But this kind of 'Millian orthodoxy' neglects the possibility that sometimes the things we take to be true really are true, and that various non-epistemic benefits can be worth the risk or even cost of some epistemic loss (Schauer 2012). What we need then, are well justified 'all-things-considered' assessments of the status of specific dissent about science, and how we ought to respond to it in different contexts.

As my position is that these 'all-things-considered' assessments and decisions are highly context-dependent, and thus should be made on a case-by-case basis rather than in the abstract,[84] I will not attempt to reach conclusions about particular instances of dissent here. What I will do, however, is provide a sketch of the sorts of considerations I think should factor into these evaluations and decisions, and how future scholarship by philosophers and science studies scholars could contribute.

First, those making assessments and judgments about speech acts about science should try to determine what the interests, rights and legitimate claims of the speakers and listeners are in that particular context. Evaluators should start to think about whether there is the potential for any of these interests or rights to come into conflict. Furthermore, they should consider who holds a duty — legal, professional or moral — to fulfil that interest, right, or claim.

Next, the potential *epistemic* consequences of that speech being uttered or published should be considered, including the rough likelihoods of such consequences, and whether these epistemic risks can be expected to vary for different sub-groups within the audience.

---

[84] Perhaps this even explains why, as dMMI questioningly point out (p. 602), dissent that has differing characteristics is often targeted for what seem to be 'unsystematic' reasons.

From there, one can try to establish what the flow-on effects of these epistemic consequences are likely to be, in terms what *non-epistemic* consequences may emerge, and how likely they are to do so. In this thesis I have highlighted three non-epistemic consequences that I believe are particularly important and central to the question of problematic dissent in democratic societies: effects on personal autonomy, freedom of speech, and the democratic legitimacy of public policy. But there are undoubtedly many more. Again, one will need to ask whether these consequences are likely to differ for different people, and with what further repercussions.

The range of options for responding to this dissent, and the epistemic and non-epistemic risks of each, will also need to be thought through and contrasted. For example, how does allowing someone who has been known to misrepresent his colleagues' research to present at a conference alone in a single slot compare to offering him only a presentation slot on a panel with other experts? Or do his recent violations of research and communicative integrity mean that his presence is likely to do more harm than good, especially when his slot could be given to someone else who presents a new angle of critique? There are lots of possibilities that need to be thought through with nuance.

By this point, the epistemic and moral landscape will hopefully be much clearer. One should be in a much better position to identify a broad range of epistemic and non-epistemic benefits, risks and costs associated with a particular speech act being expressed within a particular forum, in a particular way, by a particular actor, and as well as the benefits, risks and costs of treating that speech in a particular way. Where these costs and benefits stand in a trade-off relationship will hopefully be apparent, but it's worth remembering that these tensions and trade-offs are likely to be different in different parts of the public knowledge system, and to vary depending on who the speakers and listeners are. Think about the potential differences, for example, when a forum is composed of all or mainly contributory or interactive experts, versus the situation in which a room is mostly full of non-experts.

What is left then, is the task of weighing up what's more important within a particular context to make an 'all-things-considered' decision about the status of a particular dissenting speech act (i.e. problematic/normatively inappropriate, or not, and perhaps the degree to which it is), and how that speech act, or potential speech act, ought to be

treated within that particular part of the public knowledge system. The decisionmaker should ask themselves which values and risks, in the specific context of the decisions they must make, are weightier than others, and why.

Once a decision has been made, it can then, where appropriate, be justified to others by being transparent about the content and process of the reasoning underpinning this decision, including: the values, empirical assumptions, and evidence base that informed it; the values encapsulated by evaluations of dissent as 'problematic' or 'not problematic' and charges that certain ways of responding to dissent are 'dangerous'; the values one thinks a particular response to dissent can help achieve, or risks undermining; and the values one is and is not willing to trade off to secure other values. Such explanations should avoid empty appeals to 'free speech' and the like, and instead clearly articulate what a right to free speech is understood to be constituted by, for whom, and how the decision is expected to further that value.

Because these activities are normative, it seems that moral, social and political philosophy will need to play a much greater role in these debates than they have to date. In particular, as the debate about dissent in and about science parallels, but lags well behind, progress that has been made in the free speech debate more broadly, it would benefit from closer connections to this scholarship.

Scholars should also make greater use of empirical research to provide credible evidence of the potential for different strategies to achieve particular objectives, or to have certain consequences or side-effects, in different contexts. However, at the same time, there is a need to critique this research, especially the validity of its conclusions, and how far findings are warranted to be generalised. We should also be mindful that the evidence required might not yet exist, that there may be genuinely competing theories and models within this literature, or that different models, theories and findings apply to different individuals, sub-populations, or scenarios.

Before concluding I would like to make a brief comment on terminology choice. Throughout this chapter I used dMMI's language of 'targeting dissent' to aid the tracing of arguments. However, 'targeting' is a loaded word, as it commonly refers to the object of an attack. I think the use of such language has the effect of unfairly and unduly

influencing a reader's attitude towards such responses before they have had an opportunity to properly engage with the substantive arguments, and largely via an unconscious mechanism. Therefore, I suggest the term 'checking' should be used in its place, where to check dissent means to halt or slow its progress, or prevent it from entering, some part of the public knowledge system. An added benefit of employing this terminology is that because checking also means to examine something to determine its accuracy, quality, or appropriateness, it reminds us that before we check dissent (in the first meaning of the word), we must check it (examine it), carefully and thoroughly, reflect on it, and ensure we have good reasons for our subsequent actions.

## *Conclusion*

I have argued that dMMI's argument in 'Who's Afraid of Dissent' is unsuccessful: it is far from clear that checking problematic dissent about science is even generally misguided and/or dangerous. Under some conditions, checking dissent will not be misguided: the Probability Argument shows that checking dissent can help non-experts to more accurately perceive critical features of epistemic and policy landscapes, and therefore assist people to form well-calibrated factual beliefs relevant to their normative policy-related decisions. It also demonstrates that checking dissent can help to secure other ends unrelated to policy content. In criticizing dMMI's claim that checking dissent is dangerous, I did not argue that checking dissent does not carry risks; rather, I demonstrated that the arguments forwarded by dMMI also apply to *not* checking dissent, and highlighted the reasons we have to doubt that dMMI's proposed education alternative will be efficacious and riskless in several contexts.

Once a fuller account of the potential negative risks and consequences associated with dissenting speech acts about science are brought into sharper relief, and doubt is cast on the putatively risk-free nature of refraining from checking dissent and addressing the problem alternatively through education, it becomes very plausible that certain ways of checking certain types or instances of dissent could, in some circumstances, be the *less misguided and dangerous* action, or be complementary to, or even sometimes an aspect of, education. We should not rule out the soundness or reasonableness of checking dissent,

or of any other response to problematic dissent, *a priori*, but rather should adopt a nuanced and contextual all-things-considered approach.

# CHAPTER SIX:
## CONCLUSION

In response to the overwhelming focus in the literature to date on the *epistemic* consequences of dissent for *science and scientists*, my aim in this thesis was to provide a better understanding of the *non-epistemic* consequences of problematic scientific dissent for *non-experts* and *society* more broadly. Specifically, I wanted to articulate what is *politically problematic* about speech acts about scientific issues that misrepresent experts' testimonies or that otherwise contribute to non-experts being misled about the contours of epistemic landscapes. I delivered a three-part answer to this question: some speech acts about science are politically problematic because they threaten three core liberal democratic values: (i) personal autonomy, (ii) freedom of speech and (iii) democratic legitimacy of public policy.

In Chapter Two I defended the position that false (and inaccurate) beliefs that are the product of problematic speech acts and omissions about science can compromise personal autonomy; holding erroneous beliefs can undermine a person's control over their choice of values as well as whether their actions realise their values. However, I argued that only those false beliefs that are formed *non-voluntarily* are autonomy-compromising. The (non)voluntary status of the etiology of false beliefs depends on whether a person possessed a fair opportunity to form accurate factual beliefs or to resist forming inaccurate factual beliefs.

I proposed a set of jointly sufficient conditions to identify these circumstances. An empirical belief can be considered to have been voluntarily formed if and only if the following conditions all hold:

(1) In light of what is known at the time, or what one can legitimately expect to be known, adequate evidence is available within the public knowledge system, and a person has effective access to this evidence;

(2) The body of putative evidence a person has access to is generally presented with integrity; there are no significant misrepresentations or misleading uses of evidence within important parts of the public knowledge system;

(3) A person is able to deploy their reflective capacities.

Chapter Two therefore illuminated the way that the integrity of our personal autonomy is not solely an individual enterprise and instead significantly turns on those relationships we are legitimately dependent on to acquire certain types of knowledge and information. Respect for the autonomy of others demands that certain actors within knowledge and information systems meet their professional — and general moral — obligations to provide others with certain types of information and to refrain from transmitting misinformation. When these actors fail in their communicative responsibilities, one important way that people are wronged is by virtue of the way in which such acts and omissions can undermine their personal autonomy.

Following on from Chapter Two's conclusions, Chapter Three then asked: what, if any, implications does such a loss of autonomy have for misinformed persons' freedoms of speech? I identified and analysed a new freedom-constraining silencing phenomenon that I called *pre-emptive silencing*. Pre-emptive silencing occurs when people's freedom of thought has already been or is, in the course of a social interaction, compromised by the speech of other agents. The conditions that result in a person's forming beliefs non-voluntarily thus also serve to curtail the potential scope of a person's future speech and therefore what they can do with their words.

Chapter Three also argued that we have just as much reason to regard pre-emptive silencing as a free speech violation as we do to regard other types of silencing as such. Moreover, we have reason to believe that pre-emptive silencing is a widespread phenomenon, especially in the context of scientific expert-intermediary-non-expert relations. Since it is a speech-related and communication-based harm and/or wrong, we have ample reason to attend to the phenomenon of pre-emptive silencing.

The aim of Chapter Four was to explore how the non-epistemic consequences identified in the two previous chapters could impact upon the development of science-informed public policy. I argued that concerns about the impact of problematic dissent and other

speech acts about science on public policy and democratic values are best understood as concerns about the democratic legitimacy of public policy. Although much of the extant literature frames these concerns in ways that accord with Rational Epistemic Proceduralist accounts of democratic legitimacy, I argued that Fabienne Peter's account of Pure Epistemic Proceduralism yields an alternative, and more attractive, diagnosis of the how problematic scientific speech can interfere with the democratic legitimacy of public policy.

Peter's account of democratic legitimacy avoids controversial assumptions about the correctness of the content of public policy and instead focuses only on the procedure it was made through. Peter's Pure Epistemic Proceduralism rests on a link between procedural fairness and epistemic value. By requiring that public deliberation tied to policy-making is carried out under conditions of *social equality* and characterised by *epistemic fairness*, it captures the important idea that inclusive public deliberation is valuable not just for the ideal of political fairness that it embodies, but also for its epistemic contribution to public deliberation and democratic decision-making. Using Kolodny's more specific conditions for the social equality aspect of political fairness and Longino's criteria for epistemic fairness, I demonstrated that some dissent about science can interfere with the degree to which such conditions are met and, hence, can undermine the democratic legitimacy of public policy.

Finally, in Chapter Five I turned to the question: how should we respond to the existence and circulation of problematic dissent and other problematic speech acts and omissions about science within our public knowledge and information systems? I closely examined de Melo-Martin and Intemann's argument that 'targeting' the dissenting speech acts is misguided and dangerous and argued that their argument was ultimately unsuccessful, as it is far from clear that checking problematic dissent about science is even generally misguided and/or dangerous.

 I developed the 'Probability Argument' to show that certain ways of checking the presence or reach of problematic dissent within knowledge and information systems are not misguided because doing so can often help non-experts to more accurately perceive critical features of the epistemic and policy terrain and can therefore assist them to form well-calibrated factual beliefs relevant to their decision-making. In replying to de Melo-

Martin and Intemann's charge that checking dissent is generally dangerous, I did not argue that doing so does not carry risks; rather, I demonstrated that the arguments they offered for this claim also apply to *not* checking dissent. Additionally, I highlighted several reasons we have to doubt that their proposed education alternative will be efficacious and riskless in non-ideal, practical contexts.

Once a fuller account of the potential negative risks and consequences associated with dissenting speech acts about science are brought into sharper relief, and doubt is cast on the putatively risk-free nature of refraining from checking dissent and addressing the problem alternatively through education, it becomes very plausible that certain ways of checking some types or instances of dissent can, in some circumstances, be the *less misguided and dangerous* action or be complementary to, or even sometimes an aspect of, education. We should not rule out the soundness or reasonableness of checking dissent, or of any other response to problematic dissent, *a priori*. Instead, we should adopt a nuanced and contextual all-things-considered approach. I concluded by sketching a basic framework of the reasoning process different actors should undergo when evaluating dissent and making decisions about how they ought to respond to it.

The key direction I think future scholarship on dissent and other speech acts about science should take is a series of detailed case studies of particular speech acts in particular contexts. These case studies should offer 'all-things-considered' evaluations of the ethics of speech acts and various responses to them in that context. Such a case study approach would help to identify 'exemplars' of beneficial and pernicious dissent and scientific speech as well as well- and poorly-judged responses to these speech acts. Given the radical context-dependency of these judgments, I think this approach is more suitable than trying to come up with a set of abstract general communication principles.

There is also more work to be done to further connect debates about dissent and speech acts about science to other debates and bodies of academic literature at the interface of philosophy of science and political epistemology, for example, the literature on epistemic responsibilities (e.g. Code 1987) and epistemic trust (e.g. Grasswick 2010, 2014). At the interface of philosophy of science and political philosophy/theory, a particular area of need, I think, is in developing a better understanding of the professional duties of various actors vis-à-vis science communication, such as those who occupy important

nodes in public knowledge systems, e.g. scientists, politicians, journalists, and media managers. Considering the rapid advancement of science and technology in our societies, together with widening inequalities and our present struggles to find broad agreement on common factual foundations upon which democratic deliberations can be conducted, I think that communicative ethics—especially in the realm of specialist information—is an important area for philosophers to continue to contribute to.

# BIBLIOGRAPHY

*Chapter One*

Beatty, J., & Moore, A. 2010. Should We Aim For Consensus?. *Episteme*, 7(3): 198-214.

_____. 1928. Manipulating Public Opinion: The Why and the How. *American Journal of Sociology*, 33(6): 958-971.

_____. 1928. *Propaganda*. New York: H. H. Liveright.

_____. The Engineering of Consent. *The Annals of the American Academy of Political and Social Science*, 250(1): 113-120.

_____. 1965. Biography of an Idea: Memoirs of Public Relations Counsel Edward L. Bernays. Simon and Schuster.

Biddle, J. B., & Leuschner, A. 2015. 'Climate Skepticism and the Manufacture of Doubt: Can Dissent in Science Be Epistemically Detrimental?' *European Journal for Philosophy of Science*, 5(3), 261-278.

Biddle, J. B., Leuschner, A., & Kidd, I. J. 2017. 'Epistemic Corruption and Manufactured Doubt: The Case of Climate Science.' *Public Affairs Quarterly*, 31 (3):165-187.

Carey, A. 1995. *Taking the Risk out of Democracy: Propaganda in the US and Australia*. Sydney: University of New South Wales Press.

Caspari, E. W., & Marshak, R. E. 1965. The Rise and Fall of Lysenko. *Science*, 149(3681): 275-278.

deJong-Lambert. W. 2012. *The Cold War Politics of Genetic Research: An Introduction to the Lysenko Affair*. Springer Netherlands.

de Melo-Martín, I, and K. Intemann. 2014. 'Who's Afraid of Dissent? Addressing Concerns About Undermining Scientific Consensus in Public Policy Developments.' *Perspectives on Science*, 22(4): 593–615.

_____. 2018. *The Flight Against Doubt*. Oxford: Oxford University Press.

Epstein, S. 1995. The Construction of Lay Expertise: AIDS Activism and the Forging of Credibility in the Reform of Clinical Trials. *Science, Technology, & Human Values*, 20(4): 408-437.

Ewen, S. 1996. *PR!: A Social History of Spin*. New York: Basic Books.

Feyerabend, P. 1975. *Against Method: An Outline of an Anarchistic Theory of Knowledge*, 1st edn. (London, New Left Books).

Furman, K. 2016. *AIDS Denialism in South Africa: A Case Study in the Rationality and Ethics of Science Policy.* Unpublished Doctoral Dissertation. London School of Economics and Political Science.

Kitcher, P. 2011. *Science in a Democratic Society*, Amherst, NY: Prometheus Press.

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

_____. 2002. *The Fate of Knowledge.* Princeton University Press.

Michaels, D. 2008. *Doubt is Their Product.* New York: Oxford University Press.

Mill, J. S. 1988. The Collected Works of John Stuart Mill, Volume XXVI – Journals and Debating Speeches Part I, J. M. Robson (Ed.). London: Routledge and Kegan Paul.

_____ 1991. "On Liberty" in J. Gray (Ed.), *On Liberty and Other Essays.* Oxford: Oxford University Press [originally published 1859].

Moore, A. 2017. *Critical Elitism: Deliberation, Democracy, and the Problem of Expertise.* Cambridge University Press.

Mouffe. C. 1996. Democracy, Power and the Political. In S. Benhabib, ed., *Democracy and Difference: Contesting the Boundaries of the Political.* Princeton University Press, pp. 245-256.

Oreskes, N. and Conway, E. 2010. *Merchants of Doubt.* London: Bloomsbury.

Orwell, G. 2013. *George Orwell: A Life in Letters.* WW Norton & Company.

Popper, K. 1959/2007. *The Logic of Scientific Discovery.* London: Routledge.

Popper, K. 1972. *Conjectures and Refutations: The Growth of Scientific Knowledge.* 4th Edition. London: Routledge and Kegan Paul.

Proctor, R. N. 1996. *Cancer Wars: How Politics Shapes What We Know and Don't Know About Cancer.* Basic Books.

Proctor, R. N., & Schiebinger, L. 2008. *Agnotology: The Making and Unmaking of Ignorance.* Stanford University Press.

Proctor, R. N. 2011. *Golden Holocaust: Origins of the Cigarette Catastrophe and the Case for Abolition.* University of California Press.

Rossianov, K. O. 1993. Editing Nature: Joseph Stalin and the "New" Soviet Biology. *Isis*, 84(4): 728-745.

Shelden, M. 1991. *Orwell: The Authorised Biography.* London: Heinemann.

Solomon, M. 2001. *Social Empiricism.* Cambridge, MA: MIT Press.

Solomon, M. 2006. Groupthink versus The Wisdom of Crowds: The Social Epistmology of Deliberation and Dissent. *The Southern Journal of Philosophy*, XLIV: 28–42.

Solomon, M. 2008. *Norms of Dissent.* Contingency and Dissent in Science Project, Centre for Philosophy of Natural and Social Science, London School of Economics and Political Science.

Stanchevici, D. 2017. *Stalinist Genetics: The Constitutional Rhetoric of TD Lysenko.* Routledge.

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

World Health Organisation (WHO). 2011. *Global Adult Tobacco Survey: Indonesia Report 2011.* World Health Organization, Regional Office for South East Asia. Accessed at: http://www.searo.who.int/tobacco/data/gats_indonesia_2011.pdf

_____. 2018. *Tobacco Fact Sheet.* 9 March 2018. Accessed at: http://www.who.int/news-room/fact-sheets/detail/tobacco

## *Chapter Two*

Anderson, E. 2011. Democracy, Public Policy, and Lay Assessments of Scientific Testimony. *Episteme*, 8(02): 144–164.

Anscombe, G. E. M. 1957/2000. *Intention.* Harvard University Press.

Arendt, H. 1967. Truth and Politics. *New Yorker*, February 25 1967: 49-67.

Arpaly, N. 2005. Responsibility, Applied Ethics, and Complex Autonomy Theories. In: *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy*, 162-180.

Avorn, J. 2006. Dangerous Deception – Hiding the Evidence of Adverse Drug Effects. *New England Journal of Medicine*, 355(21): 2169-2171.

Audi, R. 2001. Doxastic Voluntarism and the Ethics of Belief. In M. Steup ed., *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue.* Oxford University Press, pp. 93-111.

Barnes, D. E., Hanauer, P., Slade, J., Bero, L. A., & Glantz, S. A. 1995. Environmental Tobacco Smoke: The Brown and Williamson Documents. *JAMA, 274*(3): 248-253.

Barnes, A. 1997. *Seeing Through Self-Deception*, Cambridge: Cambridge University Press.

Bok, S., 1978. *Lying: Moral Choice in Public and Private Life*, New York: Random House.

Carson, T. L. 2010. *Lying and Deception: Theory and Practice.* Oxford University Press.

Coady, D. 2006. When Experts Disagree. *Episteme* 3(1–2):68–79.

Coady, D., and R. Corry. 2013. *The Climate Change Debate: An Epistemic and Ethical Enquiry.* Basingstoke: Palgrave Macmillan.

Colburn, B. 2008. Debate: The concept of voluntariness. *Journal of Political Philosophy,* 16(1): 101–111.

_____. 2010. *Autonomy and Liberalism.* Routledge.

Dworkin, G. 1988. *The Theory and Practice of Autonomy.* Cambridge University Press.

Editors. 2013. What is the Purpose of Medical Research? *The Lancet,* 381: 9864.

Frankfurt, Harry G. 1971. Freedom of the Will and the Concept of a Person, *Journal of Philosophy,* 68: 5-20.

Friedman, M. 2003. *Autonomy, Gender, Politics,* New York: Oxford University Press.

Gardner, M. N., & Brandt, A. M. 2006. 'The Doctors' Choice Is America's Choice': The Physician in US Cigarette Advertisements, 1930–1953. *American Journal of Public Health,* 96(2): 222-232.

Goldacre, B. 2014. *Bad Pharma: How Drug Companies Mislead Doctors and Harm Patients.* Macmillan.

Goldberg, S. C. 2017. Should Have Known. *Synthese,* 194(8): 2863-2894.

Goldman, A. 2001. Experts: Which Ones Should You Trust? *Philosophy and Phenomenological Research,* 63(1):85–110.

Hausman, D. M. 2011. *Preference, Value, Choice, and Welfare.* Cambridge University Press.

Killmister, S. 2013a. Autonomy and False Beliefs. *Philosophical Studies,* 164(2): 513-531.

_____. 2013b. Autonomy and the Problem of Socialization. *Social Theory and Practice,* 39(1): 95–119.

_____. 2017. *Taking the Measure of Autonomy: A Four-Dimensional Theory of Self-Governance.* Routledge.

Lane, M. 2014. When the Experts Are Uncertain: Scientific Knowledge and the Ethics of Democratic Judgment. *Episteme* 11(1):97–118.

Light, D.W., Lexchin, J., and Darrow, J.J. 2013. Institutional Corruption of Pharmaceuticals and the Myth of Safe and Effective Drugs. *The Journal of Law, Medicine & Ethics,* 41(3): 590 – 600.

MacKenzie, C., & Stoljar, N. 2000. *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self.* Oxford University Press.

Mahon, J. 2015. The Definition of Lying and Deception. In: *Stanford Encyclopaedia of Philosophy*, (Winter 2016 Edition), Edward N. Zalta (ed.).

McKenna, M. 2005. The Relationship Between Autonomous and Morally Responsible Agency. In: *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy*. Cambridge University Press.

Mele, A. R. 1995/2001. *Autonomous Agents: From Self-Control to Autonomy*. Oxford University Press.

Michaels, D. 2008. *Doubt is Their Product: How Industry's Assault on Science Threatens Your Health.* Oxford University Press.

Mill, J.S. 1991. "On Liberty" in J. Gray (Ed.), *On Liberty and Other Essays*. Oxford: Oxford University Press [originally published 1859].

Olsaretti, S. 1998. Freedom, Force and Choice: Against the Rights-Based Definition of Voluntariness. *Journal of Political Philosophy*, 6(1): 53-78.

_____. 2004. *Liberty, Desert and the Market: A Philosophical Study.* Cambridge University Press.

_____. 2008. Debate: The Concept of Voluntariness – A Reply. *Journal of Political Philosophy*, 16(1): 112–121.

_____. 2018. Voluntariness, Coercion, Self-Ownership. In David Schmidtz and Carmen E. Pavel, eds., *The Oxford Handbook of Freedom*. Accessed online at: DOI: 10.1093/oxfordhb/9780199989423.013.24

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

Pollay, R. W., & Dewhirst, T. 2002. The Dark Side of Marketing Seemingly "Light" Cigarettes: Successful Images and Failed Fact. *Tobacco Control*, 11(1): i18-i31.

Post, R. 1993. Managing Deliberation: The Quandary of Democratic Dialogue. *Ethics*, 103(4): 654-678.

Proctor, R. N. 1995. *Cancer Wars: How Politics Shapes What We Know and Don't Know About Cancer.* Basic Books.

Proctor, R. N. 2011. *Golden Holocaust: Origins of the Cigarette Catastrophe and the Case for Abolition.* University of California Press.

Rosen, G. 2003. Culpability and Ignorance. *Proceedings of the Aristotelian Society* 103(1): 61-84.

_____. 2008. Kleinbart the Oblivious and Other Tales of Ignorance and Responsibility. *The Journal of Philosophy*, 105(10): 591-610.

Saul, J. 2012. *Lying, Misleading, and What Is Said*. Oxford: Oxford University Press.

Scanlon, T. M. 1998. *What We Owe to Each Other*. Cambridge, Mass.: Harvard University Press.

Schwarz, N., L.J. Sanna, I. Skurnik, and C. Yoon. 2007. Metacognitive Experiences and the Intricacies of Setting People Straight: Implications for Debiasing and Public Information Campaigns. *Advances in Experimental Social Psychology* 39:127–161.

Shiffrin, S. V. 2014. *Speech Matters: On Lying, Morality, and the Law.* Princeton: Princeton University Press.

Taylor, J. S. 2009. *Practical Autonomy and Bioethics*. Routledge.

Vitz, R. 2018. Doxastic Voluntarism. *Internet Encyclopaedia of Philosophy*. Accessed at http://www.iep.utm.edu/doxa-vol/#H2, June 2018.

Watson, Gary. 1975. Free Agency. *Journal of Philosophy*, 72: 205-20.

Whiteside, T. 1971. *Selling Death: Cigarette Advertising and Public Health*. New York: Liveright.

Yaffe, G. 2011. Intoxication, Recklessness, and Negligence. *Ohio St. J. Crim. L.*, 9: 545.

## *Chapter Three*

Austin, J. L. 1955/1975. *How To Do Things With Words.* Second Edition. Cambridge: Harvard University Press.

Begon, J. 2015. What Are Adaptive Preferences? Exclusion and Disability in the Capability Approach. *Journal of Applied Philosophy*, 32(3): 241-257.

Berlin, I. 1969. *Four Essays on Liberty*. Oxford University Press.

Biko, S. 2002 [1978]. *I Write What I Like: Selected Writings*. University of Chicago Press.

Brison, S. J. 1998. The Autonomy Defense of Free Speech. *Ethics*, 108(2): 312-339.

_____. 2000. Relational Autonomy and Freedom of Expression. In *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self*, MacKenzie and Stoljar, eds., Oxford University Press.

Chambers, C. 2007. *Sex, Culture, and Justice: The Limits of Choice*. University Park: Penn State University Press.

Crasnow, S. and Waugh J. 2013. Introduction: Philosophers and Popular Culture. In S. Crasnow and J. Waugh (Eds.), *Philosophical Feminism and Popular Culture*. Plymouth: Lexington Books.

de Melo-Martín, I, and K. Intemann. 2018. *The Flight Against Doubt*. Oxford: Oxford University Press.

Dworkin, R. 1993. Women and Pornography, *The New York Review of Books* 40: 36-42.

Everhart, D., & Hameed, S. 2013. Muslims and Evolution: A Study of Pakistani Physicians in the United States. *Evolution: Education and Outreach*, 6(1): 2.

Fanon, F. 1952/1970. Black Skin, White Masks. London: Paladin.

Finlayson, L. 2014. How to Screw Things With Words. *Hypatia*, 29(4): 774-789.

Fish, S. 1994. *There's No Such Thing As Free Speech: And It's a Good Thing, Too.* Oxford University Press.

Furman, K. 2016. *AIDS Denialism in South Africa: A Case Study in the Rationality and Ethics of Science Policy.* Unpublished Doctoral Dissertation. London School of Economics and Political Science.

_____. Forthcoming. Moral Responsibility, Culpable Ignorance, and Suppressed Disagreement. *Social Epistemology.* Online first October 2018: https://www.tandfonline.com/doi/full/10.1080/02691728.2018.1512173

Gilmore, J. 2011. Expression as Realization: Speakers' Interests in Freedom of Speech. *Law and Philosophy*, 30(5): 517-539.

Green, L. 1998. Pornographizing, Subordinating, and Silencing. In *Censorship and Silencing: Practices of Cultural Regulation*, R.C. Post, ed., pp. 285–311. Los Angeles: The Getty Research Institute for the History of Art and the Humanities.

Heinze, E. 2016. Hate Speech and Democratic Citizenship. Oxford University Press.

hooks, b. 1992. *Black Looks: Race and Representation.* Boston: South End Press.

Hornsby, J. 1993. Speech Acts and Pornography. *Women's Philosophy Review* 10: 38–45.

_____. 1994. Illocution and Its Significance. In *Foundations of Speech Act Theory: Philosophical and Linguistic Perspectives*, ed. Savas L. Tsohatzidis. New York: Routledge.

Hornsby, J., & Langton, R. 1998. Free Speech and Illocution. *Legal Theory*, 4(1), 21-37.

Hudson, S. V. 1998. "Recreational Television: The Paradox of Change and Continuity Within Stereotypical Iconography", *Sociological Inquiry* 68 (2), 242-57.

Jacobson, D. 1995. Freedom of Speech Acts? A Response to Langton. *Philosophy & Public Affairs*, 24(1): 64-78.

Kukla, R. 2018. That's What She Said: The Language of Sexual Negotiation. *Ethics*, 129(1): 70-97.

Langton, R. 1993. Speech Acts and Unspeakable Acts. *Philosophy & Public Affairs*, 293-330.

MacFarlane, L. J. 1966. On Two Concepts of Liberty. *Political Studies*, 14(1): 77-81.

MacKinnon, C. A. 1987. *Feminism Unmodified*. Cambridge: Harvard University Press.

_____ 1993. *Only Words*. Cambridge: Harvard University Press.

Maitra, I. 2004. Silencing and responsibility. *Philosophical Perspectives* 18 (1): 189–208.

 _____. 2009. Silencing speech. *Canadian Journal of Philosophy* 39 (2): 309–38.

Matsuda, M. J. 2018. *Words That Wound: Critical Race Theory, Assaultive Speech, and the First Amendment*. London: Routledge.

McGowan, M. K. 2009. Oppressive Speech. *Australasian Journal of Philosophy*, 87(3): 389-407.

_____. 2014. Sincerity Silencing. *Hypatia*, 29(2): 458-473.

_____. 2017. On Multiple Types of Silencing. In Mikkola Mari, ed., *Beyond Speech: Pornography and Analytic Feminist Philosophy*. Oxford University Press.

Mele, A. R. 1995/2001. *Autonomous Agents: From Self-Control to Autonomy*. Oxford University Press.

Nussbaum, M. C. 2001. *Women and Human Development: The Capabilities Approach*. Cambridge University Press.

Oreskes, N., and Conway, E. 2010. *Merchants of Doubt*, New York, NY: Bloomsbury Press.

Post, R. C. 1990. Racist Speech, Democracy, and the First Amendment. *Wm. & Mary L. Rev.*, 32: 267-328.

Proctor, R. N., & Schiebinger, L. 2008. *Agnotology: The Making and Unmaking of Ignorance*. Stanford University Press.

Scanlon, T. 1972. A Theory of Freedom of Expression. *Philosophy and Public Affairs*, 1(2): 204–26.

Schauer, F. 1985. Slippery Slopes. *Harvard Law Review*, 361-383.

_____. 2014. Out of Range: On Patently Uncovered Speech. *Harvard Law Review Forum*, 128: 346-353.

Scheufele, D. A. 2014. Science Communication as Political Communication. *Proceedings of the National Academy of Sciences*, 111(4): 13585-13592.

Shiffrin, S. V. 2014. *Speech Matters: On Lying, Morality, and the Law*. Princeton: Princeton University Press.

Simpson, R. M. 2013. Intellectual Agency and Responsibility for Belief in Free-Speech Theory, *Legal Theory* 19 (3): 307-30.

_____. 2015. 'Speech Matters: On Lying, Morality, and the Law.' *Notre Dame Philosophy Reviews*. 28 April 2015. Accessed at: https://ndpr.nd.edu/news/speech-matters-on-lying-morality-and-the-law/

_____. 2017. 'Free Speech' is a Blunt Instrument. Let's Break It Up. *Aeon*. 31 March 2017. Accessed at: https://aeon.co/ideas/free-speech-is-a-blunt-instrument-lets-break-it-up

Weisberg, D. S., Landrum, A. R., Metz, S. E., & Weisberg, M. 2018. No Missing Link: Knowledge Predicts Acceptance of Evolution in the United States. *BioScience*, 68(3): 212-222.

West, Caroline. 2003. The Free Speech Argument Against Pornography. *Canadian Journal of Philosophy* 33(3): 391–422.

## Chapter Four

Baker, Erik and Naomi Oreskes. 2017. It's No Game: Post-Truth and the Obligations of Science Studies. *Social Epistemology Review and Reply Collective* 6(8): 1-10.

Bayer, R. 1987a. Homosexuality and American Psychiatry: The Politics of Diagnosis. Princeton University Press.

Bayer, R. 1987b. Politics, Science, and the Problem of Psychiatric Nomenclature: A Case Study of the American Psychiatric Association Referendum on Homosexuality. *Scientific Controversies: Case studies in the Resolution and Closure of Disputes in Science and Technology*. Engelhardt Jr, H. Tristram, Hugo Tristram Engelhardt, and Arthur L. Caplan (Eds.), pp. 381-400. Cambridge: Cambridge University Press.

Berlin, I. 1969. *Four Essays on Liberty*. Oxford University Press.

Biddle, J. B., & Leuschner, A. 2015. Climate skepticism and the manufacture of doubt: can dissent in science be epistemically detrimental?. *European Journal for Philosophy of Science*, 5(3), 261-278.

Biddle, J. B., Leuschner, A., & Kidd, I. J. 2017. Epistemic Corruption and Manufactured Doubt: The Case of Climate Science. *Public Affairs Quarterly*, 31 (3):165-187.

Collins, P. H. 1986. Learning from the outsider within: The sociological significance of Black feminist thought. *Social Problems*, 33(6), s14-s32.

Cohen, J. 1986. An Epistemic Conception of Democracy. *Ethics*, 97(1): 26-38.

de Melo-Martín, I, and K. Intemann. 2013. Scientific Dissent and Public Policy. *EMBO Reports,* 14(3): 231–235.

_____ 2014. Who's Afraid of Dissent? Addressing Concerns About Undermining Scientific Consensus in Public Policy Developments. *Perspectives on Science,* 22(4): 593–615.

_____. 2018. *The Flight Against Doubt.* Oxford: Oxford University Press.

Ding, D., E.W. Maibach, X. Zhao, C. Roser-Renouf, and A. Leiserowitz. 2011. "Support For Climate Policy and Societal Action Are Linked to Perceptions About Scientific Agreement." *Nature Climate Change* 1(9):462–466.

Douglas, H. 2009. *Science, Policy, and the Value-Free Ideal.* University of Pittsburgh Press.

Drescher, J. 2015. Out of DSM: Depathologizing Homosexuality. *Behavioral Sciences,*5(4): 565-575.

Estlund, David. 2008. *Democratic Authority: A Philosophical Framework.* Princeton: Princeton University Press.

Furman, K. 2016. *AIDS Denialism in South Africa: A Case Study in the Rationality and Ethics of Science Policy.* Unpublished Doctoral Dissertation. London School of Economics and Political Science.

Goodin, R. 1995. Laundering preferences. In *Utilitarianism as a Public Philosophy* (Cambridge Studies in Philosophy and Public Policy, pp. 132-148). Cambridge: Cambridge University Press.

Hartsock, N. C. 1983. The feminist standpoint: Developing the ground for a specifically feminist historical materialism. In *Discovering Reality* (pp. 283-310). Springer, Dordrecht.

Harding, S. G. 2004. *The Feminist Standpoint Theory Reader: Intellectual and Political Controversies.* Psychology Press.

Hershovitz, S. 2003. Legitimacy, Democracy, and Razian Authority. *Legal Theory*, 9(3): 201-220.

Intemann, K. 2017. "Who Needs Consensus Anyway?" *Public Affairs Quarterly*, 31(3): 189–208.

Kitcher, P. 2011. *Science in a Democratic Society*, Amherst, NY: Prometheus Press.

_____. 2016. Masking the meaningful. *Global Policy*, 7: 5-15.

Kolodny, N. 2014a. Rule Over None I: What Justifies Democracy? *Philosophy & Public Affairs*, 42(3), 195-229.

Kolodny, N. 2014b. Rule over none II: social equality and the justification of democracy. *Philosophy & Public Affairs*, 42(4): 287-336.

Landemore, H. 2012. *Democratic Reason: Politics, Collective Intelligence, and the Rule of the Many.* Princeton University Press.

List, Christian and Robert Goodin. 2001. Epistemic Democracy: Generalizing the Condorcet Jury Theorem. *Journal of Political Philosophy* 9: 277 – 306.

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

_____. 2002. *The Fate of Knowledge.* Princeton University Press.

Millar, R. J., Fuglestvedt, J. S., Friedlingstein, P., Rogelj, J., Grubb, M. J., Matthews, H. D., ... & Allen, M. R. 2017. Emission budgets and pathways consistent with limiting warming to 1.5 C. *Nature Geoscience*, 10(10), 741.

Nash, E. J. 2018. In Defense of "Targeting" Some Dissent about Science. *Perspectives on Science*, 26(3): 325-359.

Oreskes, N. 2008. The devil is in the (historical) details: Continental drift as a case of normatively appropriate consensus?. *Perspectives on Science*, 16(3), 253-264.

Oreskes, N., and Conway, E. 2010. *Merchants of Doubt*, New York, NY: Bloomsbury Press.

_____ 2014. *The Collapse of Western Civilization: A view from the Future*. Columbia University Press.

Pearce, W., R. Grundmann, M. Hulme, S. Raman, E. Hadley Kershaw, and J. Tsouvalis. 2017. Beyond Counting Climate Consensus. *Environmental Communication* 11(6): 723–730.

Peter, F. 2007. Democratic legitimacy and proceduralist social epistemology. *Politics, Philosophy & Economics*, 6(3): 329-353.

Peter, F. 2008a. *Democratic Legitimacy*, New York: Routledge.

_____. 2008b. Pure Epistemic Proceduralism. *Episteme*, 5(1): 33-55.

_____2013. "The Procedural Epistemic Value of Deliberation," *Synthese*, 190(7): 1253–1266.

Raz, J., 1986. *The Morality of Freedom*, Oxford: Clarendon Press.

Sen, A. 1999. 'Democracy and Social Justice', paper presented to the Seoul Conference on Democracy, Market Economy and Development, 26–27 February 1999.

Solomon, M. 2001. *Social Empiricism*. Cambridge, MA: MIT Press.

Solomon, M. 2006. Groupthink versus The Wisdom of Crowds: The Social Epistmology of Deliberation and Dissent. *The Southern Journal of Philosophy*, XLIV: 28–42.

Solomon, M. 2008. *Norms of Dissent.* Contingency and Dissent in Science Project, Centre for Philosophy of Natural and Social Science, London School of Economics and Political Science.

van der Linden, S. L., C.E. Clarke, and E.W. Maibach. 2015. "Highlighting Consensus Among Medical Scientists Increases Public Support For Vaccines: Evidence From a Randomized Experiment." *BMC Public Health* 15:1207.

Wylie, A. 2013. Why Standpoint Matters. In *Science and Other Cultures* (pp. 34-56). Routledge.

Wynne, B. 1989. Sheep Farming after Chernobyl: a Case Study in Communicating Scientific Information, *Environment*, 31(2): 10–15, 33–9.

Zacka, B. 2017. *When the State Meets the Street: Public Service and Moral Agency*. Harvard University Press.

## *Chapter Five*

Anderson, E. 2011. Democracy, Public Policy, and Lay Assessments of Scientific Testimony. *Episteme* 8(2):144–164.

Biddle, J. B., and A. Leuschner. 2015. Climate Skepticism and the Manufacture of Doubt: Can Dissent in Science Be Epistemically Detrimental? *European Journal for Philosophy of Science* 5(3):261–278.

Biddle, J. B., Kidd, I.J., and A. Leuschner. 2017. Epistemic Corruption and Manufactured Doubt: The Case of Climate Science. *Public Affairs Quarterly* 31(3):261–278.

Boykoff, M. T., and J.M. Boykoff. 2004. Balance as Bias: Global Warming and the US Prestige Press. *Global Environmental Change* 14(2):125–136.

Boykoff, M. T. 2011. *Who Speaks for the Climate?: Making Sense of Media Reporting on Climate Change*. Cambridge: Cambridge University Press.

Centers for Disease Control and Prevention (CDC). 2006. *The Health Consequences of Involuntary Exposure to Tobacco Smoke: A Report of the Surgeon General*. Atlanta: Office on Smoking and Health.

Clark, A. 2015. *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford: Oxford University Press.

Coady, D. 2006. When Experts Disagree. *Episteme* 3(1–2):68–79.

Coady, D. 2010. Two Concepts of Epistemic Injustice. *Episteme* 7(2):101–113.

Coady, D., and R. Corry. 2013. *The Climate Change Debate: An Epistemic and Ethical Enquiry*. Basingstoke: Palgrave Macmillan.

Collins, H. and R. Evans. 2007. *Rethinking Expertise*. Chicago: University of Chicago Press.

Cook, J. 2017. Response by Cook to 'Beyond Counting Climate Consensus'. *Environmental Communication* 11(6):733–735.

Cook, J. and S. van der Linden. 2017. "Facts Versus Feelings Isn't the Way to Think About Communicating Science." *The Conversation*, 5 July 2017: https://theconversation.com/facts-versus-feelings-isnt-the-way-to-think-about-communicating-science-80255.

Cook, J., N. Oreskes, P.T. Doran, W.R.L. Anderegg, B. Verheggen, E.W. Maibach, J.S. Carlton, S. Lewandowsky, A.G. Skuce, S.A. Green, D. Nuccitelli, P. Jacobs, M. Richardson, B. Winkler, R. Painting, and K. Rice. 2016. "Consensus on Consensus: A Synthesis of Consensus Estimates on Human-Caused Global Warming." *Environmental Research Letters* 11(4):048002.

Creagh, S. 2013. "Solar Forcing Effect on Climate Change 'Extremely Small': IPCC scientist." *The Conversation*, 14 January 2013: https://theconversation.com/solar-forcing-effect-on-climate-change-extremely-small-ipcc-scientist-11589

_____. 2014. "Who's Afraid of Dissent? Addressing Concerns About Undermining Scientific Consensus in Public Policy Developments." *Perspectives on Science* 22(4):593–615.

Ding, D., E.W. Maibach, X. Zhao, C. Roser-Renouf, and A. Leiserowitz. 2011. "Support For Climate Policy and Societal Action Are Linked to Perceptions About Scientific Agreement." *Nature Climate Change* 1(9):462–466.

Druckman, J. N., and A. Lupia. 2000. "Preference Formation." *Annual Review of Political Science* 3(1):1–24.

Furman, K. 2016. *AIDS Denialism in South Africa: A Case Study in the Rationality and Ethics of Science Policy*. Unpublished Doctoral Dissertation. London School of Economics and Political Science.

Gilbert, D. T., D.S. Krull, and P.S. Malone. 1990. "Unbelieving the Unbelievable: Some Problems in the Rejection of False Information." *Journal of Personality and Social Psychology* 59(4):601–613.

Godlee, F., J. Smith, and H. Marcovitch. 2011. "Wakefield's Article Linking MMR Vaccine and Autism Was Fraudulent." *British Medical Journal* 342:c7452.

Goldman, A. 2001. "Experts: Which Ones Should You Trust?" *Philosophy and Phenomenological Research* 63(1):85–110.

Gregory, J. M., N.J. White, J.A. Church, M.F.P. Bierkens, J.E. Box, M.R. Van den Broeke, ..., and L.F. Konikow. 2013. "Twentieth-Century Global-Mean Sea Level Rise: Is the Whole Greater Than the Sum of the Parts?" *Journal of Climate* 26(13):4476–4499.

Hamilton, L. C. 2016. "Public Awareness of the Scientific Consensus on Climate." *SAGE Open* October–December 2016:1–11.

Hammond, D. 2011. "Health Warning Messages on Tobacco Products: A Review." *Tobacco Control* 20(5):327–337.

Intemann, K. 2017. "Who Needs Consensus Anyway?" *Public Affairs Quarterly* 31(3):189–208.

Kahan, D. M., H. Jenkins-Smith, and D. Braman. 2011. "Cultural Cognition of Scientific Consensus." *Journal of Risk Research* 14(2):147–174.

Kempner, J. 2008. "The Chilling Effect: How Do Researchers React to Controversy?" *PLOS Medicine* 5(11):e222.

Kraft, P. W., M. Lodge, and C.S. Taber. 2015. "Why People 'Don't Trust the Evidence': Motivated Reasoning and Scientific Beliefs." *The ANNALS of the American Academy of Political and Social Science* 658(1):121–133.

Kuklinski, J. H., P.J. Quirk, J. Jerit, D. Schwieder, and R.F. Rich. 2000. "Misinformation and the Currency of Democratic Citizenship." *The Journal of Politics* 62(3):790–816.

Kull, S., Ramsay, C., & Lewis, E. 2003. Misperceptions, the Media, and the Iraq War. *Political Science Quarterly*, 118(4): 569-598.

Kunda, Z. 1990. "The Case for Motivated Reasoning." *Psychological Bulletin* 108(3):480–498.

Lane, M. 2014. "When the Experts Are Uncertain: Scientific Knowledge and the Ethics of Democratic Judgment." *Episteme* 11(1):97–118.

Leuschner, A. 2016. "Is it Appropriate to "Target" Inappropriate Dissent? On the Normative Consequences of Climate Skepticism." *Synthese.* Online First 15 November 2016: https://doi.org/10.1007/s11229-016-1267-x pp. 1–17.

Lewandowsky, S., U.K.H. Ecker, C.M. Seifert, N. Schwarz, and J. Cook. 2012. "Misinformation and Its Correction: Continued Influence and Successful Debiasing." *Psychological Science in the Public Interest* 13(3):106–131.

Lewandowsky, S., G.E. Gignac, and S. Vaughan. 2013. "The Pivotal Role of Perceived Scientific Consensus in Acceptance of Science." *Nature Climate Change* 3(4):399–404.

Lewandowsky, S., N. Oreskes, J.S. Risbey, B.R. Newell, and M. Smithson. 2015. "Seepage: Climate Change Denial and Its Effect on the Scientific Community." *Global Environmental Change* 33:1–13.

Luntz, F. c. 2003. "*The Environment: A Cleaner, Safer, Healthier America*". Pp. 131–146 in *The Luntz Research Companies-Straight Talk*. Unpublished Memo. Copy accessed at: https://www.sourcewatch.org/index.php/File:LuntzResearch.Memo.pdf on 6 October 2017.

Maibach, E. W., and S.L. van der Linden. 2016. "The Importance of Assessing and Communicating Scientific Consensus." *Environmental Research Letters* 11(9):091003.

Manne. R. 2011. *Bad News: Murdoch's Australian and the Shaping of the Nation.* Collingwood: Black Inc.

Mirowski, P. 2017. 'The Logic of Fake News'. Lecture at Goldsmiths University, London. 27 October 2017. Audio can be accessed here: http://www.perc.org.uk/project_posts/logic-fake-news-philip-mirowski-28th-october/

Mitchell, C. 2011. "A Critic Untroubled By Facts Who Seeks to Silence Dissent." *The Australian*, September 17 2011. Accessed at: http://www.theaustralian.com.au/business/media/a-critic-untroubled-by-facts-who-seeks-to-silence-dissent/news-story/a4b03dc8568e4c6e8fdedc032a656ece on 14 January 2018.

Nickerson, R. S. 1998. "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises." *Review of General Psychology* 2(2):175–220.

Nyhan, B., and J. Reifler. 2010. "When Corrections Fail: The Persistence of Political Misperceptions." *Political Behavior* 32(2):303–330.

Nyhan, B., J. Reifler, and P.A. Ubel. 2013. "The Hazards of Correcting Myths About Health Care Reform." *Medical Care* 51(2):127–132.

Oreskes, N. 2004. "The Scientific Consensus on Climate Change." *Science* 306(5702):1686–1686.

_____. 2017. Response by Oreskes to 'Beyond Counting Climate Consensus'. *Environmental Communication* 11(6):731–732.

Oreskes, N., and E.M. Conway. 2010. *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming.* London: Bloomsbury.

Painter, J. 2011. *Poles Apart.* Oxford: Reuters Institute for the Study of Journalism.

Pearce, W., R. Grundmann, M. Hulme, S. Raman, E. Hadley Kershaw, and J. Tsouvalis. 2017. "Beyond Counting Climate Consensus." *Environmental Communication* 11(6): 723–730.

Popper, K. 1972. *Conjectures and Refutations: The Growth of Scientific Knowledge.* 4th Edition. London: Routledge and Kegan Paul.

Schauer, F. 2012. "Social Epistemology, Holocaust Denial, and the Post-Millian Calculus." Pp. 129–43 in *The Content and Context of Hate Speech: Rethinking Regulation and Responses.* Edited by M. Herz and P. Molnar. Cambridge: Cambridge University Press.

Schwarz, N., L.J. Sanna, I. Skurnik, and C. Yoon 2007. "Metacognitive Experiences and the Intricacies of Setting People Straight: Implications for Debiasing and Public Information Campaigns." *Advances in Experimental Social Psychology* 39:127–161.

Sharman, A. 2015. *The Impact of Controversy on the Production of Scientific Knowledge.* Working Paper (No. 207). London: Grantham Research Institute on Climate Change and the Environment.

Thornton, P. 2013. "On Letters From Climate-Change Deniers." *Los Angeles Times*, 8 October 2013. Accessed at: http://beta.latimes.com/opinion/opinion-la/la-ol-climate-change-letters-20131008-story.html

Thorson, E. 2016. "Belief Echoes: The Persistent Effects of Corrected Misinformation." *Political Communication* 33(3):460–480.

U.S. National Cancer Institute and World Health Organization. 2016. *The Economics of Tobacco and Tobacco Control.* National Cancer Institute Tobacco Control Monograph 21. NIH Publication No. 16-CA-8029A. Bethesda: U.S. Department of Health and Human Services, National Institutes of Health, National Cancer Institute; and Geneva: World Health Organization.

United States v. Philip Morris USA Inc., 449 F. Supp. 2d 1 (D.D.C. 2006). *Final Judgment and Remedial Order #1015*, 16 August 2006.

van der Linden, S. 2016. "Why Doctors Should Convey the Medical Consensus on Vaccine Safety." *Evidence-Based Medicine* 21(3):119.

van der Linden, S. L., A.A. Leiserowitz, G.D. Feinberg, and E.W. Maibach. 2014. "How to Communicate the Scientific Consensus on Climate Change: Plain Facts, Pie Charts or Metaphors?" *Climatic Change* 126(1–2):255–262.

van der Linden, S. L., C.E. Clarke, and E.W. Maibach. 2015. "Highlighting Consensus Among Medical Scientists Increases Public Support For Vaccines: Evidence From a Randomized Experiment." *BMC Public Health* 15:1207.

van der Linden, S., A. Leiserowitz, and E.W. Maibach. 2016. "Communicating the Scientific Consensus on Human-Caused Climate Change is an Effective and Depolarizing Public Engagement Strategy: Experimental Evidence from a Large National Replication Study." Available at SSRN: http://dx.doi.org/10.2139/ssrn.2733956

van der Linden, S., A. Leiserowitz, S. Rosenthal, and E.W. Maibach. 2017. "Inoculating the Public against Misinformation about Climate Change." *Global Challenges* 1(2):1600008.

van Niekerk, J.P. 2003. "Politics Must Move Mainstream On AIDS." *South African Medical Journal* 93(3):154.

Wineberg, S., S. McGrew, J. Breakstone, and T. Ortega. 2016. *Evaluating Information: The Cornerstone of Civic Online Reasoning.* Stanford Digital Repository. Accessed at: http://purl.stanford.edu/fv751yt5934

World Health Organisation (WHO). 2011. *Global Adult Tobacco Survey: Indonesia Report 2011.* World Health Organization, Regional Office for South East Asia. Accessed at: http://www.searo.who.int/tobacco/data/gats_indonesia_2011.pdf

*Chapter Six:*

Code. L. 1987. *Epistemic Responsibility*. New England: Brown University Press.

Grasswick, H. E. 2010. Scientific and Lay Communities: Earning Epistemic Trust Through Knowledge Sharing. *Synthese*, *177*(3): 387-409.

_____. (2014). Climate Change Science and Responsible Trust: A Situated Approach. *Hypatia*, 29(3): 541-557.