

# Censorship, Authority, Control

Sebastien Bishop

A thesis submitted for the degree of

**Doctor of Philosophy**

University College London,

Department of Philosophy

## Declaration

I, Sebastien Bishop, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

## Acknowledgements

I am truly thankful to the many people that helped me in writing of this thesis. At UCL, I benefitted from the supervision and advice of Amanda Greene, Doug Lavin, and James Wilson. In addition, I am grateful to Kalle Grille and Gavin Phillipson, each of whom provided me with detailed feedback on chapters in this thesis.

Sincerest thanks also to Alexander Brown and Jeffrey Howard, who generously examined this thesis and gave me many helpful comments.

I am very grateful for the academic and social support I received from Rachel Harris, Chong-Ming Lim, Kate Fisk, Elena Raimondi, and Robin Wisse.

Special thanks must go to three people. Deryn Bishop – my Mum; a continual source of inspiration and support. Rob Simpson – who has been the best supervisor I could have hoped for. And Jessica Fischer – the best philosopher I know, and without whose support this thesis would simply not exist.

## Abstract

Should the government engage in the censorship of materials that it fears will be dangerously persuasive? In other words, should the government appeal to persuasion-mediated harms, when seeking to justify the regulation of speech? While many register a deep unease with this kind of censorship, typically it is conceded that such censorship is justified when the harms in question are imminent. What about when this is not the case? Several highly influential philosophical accounts propose that this kind of censorship would conflict with our core democratic values, and that the government engages in such regulation at the cost of undermining its own democratic legitimacy.

The primary aim of this thesis is to investigate whether these democracy-based worries can be cashed-out in a way that is both philosophically rigorous and somewhat broadly appealing. Several approaches to problematizing this kind of censorship are rejected. The thesis argues that despite the alarmist language sometimes deployed by philosophers, this kind of censorship need not involve a special interference with the autonomous capacities of citizens. Similarly, it is argued that philosophers have been hasty in dismissing this kind of censorship as insulting, or as constituting a case of unfair viewpoint discrimination. On both counts, it is argued that censorship can be defended through reflecting on the duties that autonomous agents owe to one another.

The thesis concludes that the most productive way of problematising the democratic objection to censorship, is through reflecting on the kind of government-citizen relationship that such censorship establishes. The thesis draws on the insights of relational, feminist theories of autonomy – offered by the likes of Catriona Mackenzie and Marina Oshana – as well as Republican theories of domination, to help better understand the way in which censorship warps and changes the government-citizen relationship.

## Impact Statement

The thesis considers the highly vexed issue of government censorship, with a special focus on the topic of censoring materials that are deemed ‘too persuasive’. The thesis takes seriously the threat posed by certain kinds of persuasive speech acts, with close attention paid to hate speech, climate change denialism, and anti-vaccination speech. While some commentators attempt to problematise government censorship by doubting the existence of persuasion-mediated harms or the effectiveness of such regulation, this thesis considers a rather different approach. In particular the thesis considers deep, deontological objections with this kind of censorship that orbit around the values of democracy and citizen autonomy. Put simply, the thesis considers whether we might have powerful reasons to reject even effective government censorship.

The thesis draws on a rich philosophical literature. On the one hand, a dialogue is constructed with some of the most influential political philosophers of our times. In particular the thesis considers the work of T.M. Scanlon, Thomas Nagel, Ronald Dworkin, and James Weinstein, and the powerful, deontological objections to government censorship they present. While taking seriously the suggestions offered by these philosophers, the thesis largely responds critically - attempting to show that the censorship of persuasive materials need not undermine our democratic or autonomous values. Indeed, it is argued that in several respects, such censorship may in fact promote these values.

On the other hand, the thesis builds upon recent insights offered by relational and feminist theorists of autonomy. The relational approach to autonomy provides many important insights regarding the flourishing of autonomous beings. Too often these insights have been overlooked by philosophical accounts of free speech. The thesis offers that reflecting upon relational approaches to autonomy helps us to clarify both the case for, and against, government censorship. Ultimately it is argued that the most productive way of framing the deep worry that so many express with censorship, is as a worry about the kind of government-citizen relationship such censorship establishes. The thesis also offers important insights regarding how the government should go about engaging in censorship, and how it should frame the censorious burdens it imposes upon citizens.

## Contents

Chapter 1: Introducing persuasion-control	8
§I. Persuasion-control	18
§II. The focus of the thesis	21
§III. Methodology	25
§IV. Deontological arguments	29
§V. Clarifying persuasion	35
§VI. A taxonomy of persuasion-mediated harms	40
a. beliefs and intentions to act	40
b. individually harmful v. cumulatively harmful	42
c. self-harm and harming others	43
d. imminent v. non-imminent harm	44
§VII. Some preliminary challenges to the thesis	45
a. is persuasive speech really harmful?	45
b. the responsibility challenge	46
c. is the ‘ethic of responsibility’ too demanding?	50
Chapter 2: Telling citizens what to think	52
§I. Scanlon’s Argument	53
§II. Autonomy	57
§III. Motives and meaning	65
§IV. Democratic values	68
§V. Unsupported Foundations	69
Chapter 3: Recognising people; recognising flaws	75
§I. Appraisal respect	77
§II. Nagel and status	83
§III. Recognition respect	96
Chapter 4: Sympathy for the devil	104
§I. A basic sketch	106
§II. Analysing Weinstein’s key claim	112
§III. Two objections	117
§IV. Analogies	127
§V. Zombie Weinstein	129

Chapter 5: “This relationship isn’t working”: governments, citizens, and censorship	140
§I. The relational approach to autonomy	141
§II. Persuasion-control and relational accounts	153
§III. The relationship account	155
§IV. Objections and responses	164
a. objections to how we have constructed the government-citizen relationship	164
b. alternatives	165
c. exaggeration	166
Conclusion	170
Bibliography	171

## Chapter 1: Introducing persuasion-control

Persuasion can be a dangerous tool. Broadly speaking persuasion involves a speaker seeking to influence the beliefs or intentions of their audience, through presenting them with reasons.<sup>1</sup> Unfortunately, as imperfect reasoners, we are all vulnerable to being influenced by persuasive speech in all manner of strange, regrettable ways. Let ‘harmful persuasion’ refer to the kind of speech that persuasively moves listeners to develop harmful beliefs or intentions. More precisely, we might say that harmful persuasion involves a speaker adducing reasons in favour of beliefs and intentions that expose others to harm, or at least increase the risk of their being harmed.<sup>2</sup>

---

<sup>1</sup> Richard Perloff (2020). *The Dynamics of Persuasion: Communication and Attitudes in the Twenty-First Century* (New York: Routledge).

<sup>2</sup> Of course, persuasion is just one part of a larger story regarding the harmfulness of speech. Certainly, speech can harm in non-persuasive ways. For instance, copyright infringement might be considered harmful because the speech itself effectively constitutes a form of theft, setting back the proprietary interests of the copyright holder. Threats, when credible, harm via disclosing information that illegitimately gives the listener cause for anxiety and distress. Similarly, blackmail discloses information to the listener in a way that may not only cause them distress, but also narrows down their viable options. Defamation and libel illegitimately involve damaging the reputation of innocent persons.

Hate speech is a trickier case to analyse. In the main text we discuss the ways that hate speech can be harmfully persuasive. But hate speech can be harmful in many ways that have nothing to do with its persuasive effects. For instance, hate speech is straightforwardly harmful when, as is often the case in face-to-face encounters, it rises to the level of harassing or threatening behaviour. Beyond these cases, many have argued that being targeted by hate speech can have serious, negative effects on one’s health. Mari Matsuda, for instance, argues that hate speech can cause targets to internalise deeply damaging racist views about themselves, thus undermining their sense of self-esteem. Matsuda also finds that being targeted by racist hate speech is associated with a number of long-term health problems, such as difficulty in breathing, nightmares, post-traumatic stress disorder, hypertension, psychosis, and suicide. To some extent these harms may be mediated by persuasion - perhaps, in some sense, listeners are persuaded to develop self-loathing attitudes. But a more natural explanation is simply that these negative effects are the result of sub-rationally internalising the hateful views of others, as well as the psychological damage that goes along with being repeatedly exposed to dehumanising, abusive messages. See: Mari Matsuda (1989). “Public response to racist speech: considering the victim's story”, *Michigan Law Review* 87 (8), pp.2320-38; Mari Matsuda (1993). *Words That Wound: Critical Race Theory, Assaultive Speech, and the First Amendment* (Boulder, Colorado: Westview Press). Richard Delgado emphasises how being the subject of hate speech can lead people to internalise highly damaging stereotypes about themselves. The result of this is deep psychological damage in the form of longstanding feelings of humiliation, isolation, and self-hatred. See: Richard Delgado (1982). “Words that wound: a tort action for racial insults, epithets, and name-calling”, *Harvard Civil Rights-Civil Liberties Law Review* 17, pp.133-83. Charles Lawrence notes the shocking, even emotionally disabling effects of being targeted by hate speech. See: Charles Lawrence (1990). “If he hollers let him go: regulating racist speech on campus”, *Duke Law Journal* 39 (3), pp.431-83. Laura Beth Nielsen, surveying a wide range of literature, suggests that exposure to sexist speech is related to a number of mental health risks, including eating disorders and the internalisation of sexist views about one’s self. Nielsen also argues that exposure to hate speech significantly

To help clarify the concept of harmful persuasion, let us reflect upon four (sometimes overlapping) categories of allegedly harmful persuasion – incitement, hate speech, pornography, and anti-science speech. Incitement is a paradigmatic case of harmful persuasion. The most straightforward cases of incitement involves a speaker directly encouraging their audience to commit harmful acts. This might involve the inciter plainly adducing reasons in favour of, say, imminent violence. However, incitement is not limited to cases of imminent harm. Indeed, a speaker plausibly counts as engaging in incitement even if they encourage their audience to engage in harmful acts many months from now.

Still, incitement can be more subtle and complicated than these straightforward cases. One example is when the incitement in question is indirect, rather than direct. As Joseph Jaconelli puts it, whereas direct incitement explicitly urges listeners to perform a certain harmful action, indirect incitement “is more circumspect, consisting of such forms as to state

---

alters the way women feel and act, especially when in public spaces. See: Laura Beth Nielsen (2002). “Subtle, pervasive, harmful: racist and sexist remarks in public as hate speech”, *Journal of Social Issues* 58 (2), pp.265-90.

In a different vein, Mary Kate McGowan argues that certain kinds of speech can be harmful insofar as they alter the norms surrounding how we speak about and treat certain people. McGowan generally focuses on the ways that certain speech acts immediately, here and now, alter our conversational norms. But she also suggests that the same kinds of speech can make a contribution to the broader social norms regarding how we may talk about and treat certain groups. These changes in the norms are not a matter of persuading listeners to act and think in different ways, but rather seem to be a product of the almost unconscious ways in which speech acts can influence us. See: Mary Kate McGowan (2019). *Just Words: On Speech and Hidden Harm* (Oxford: Oxford University Press).

Moreover, both Stephen Heyman and Jeremy Waldron discuss the potential harmfulness of hate speech in terms of dignity. Heyman worries that hate speech demonstrates a harmful lack of recognition that undermines our dignity. While Waldron talks of hate speech robbing us of our sense of assurance – that is, our sense that we are a minimally well respected member of the community. Neither of these accounts necessarily characterise the dignity-threatening aspect of hate speech as a function of its persuasive influence. On the contrary, these philosophers characterise hate speech as simply signalling to their targets, sometimes in a rather distorting manner, the psychologically damaging message that they are despised members of the community. See Steven Heyman (2009). “Hate speech, public discourse, and the first amendment”, in *Extreme Speech and Democracy* (Oxford: Oxford University Press), ed. Ivan Hare and James Weinstein, pp.151-81; Jeremy Waldron (2012). *The Harm in Hate Speech* (Cambridge, Massachusetts; Harvard University Press).

Catharine MacKinnon has argued that even consensually produced pornography may violate the right to equal civil status of women. MacKinnon argues that pornography has a profound influence on how men think about women, and ultimately normalises discrimination and violence against women. This process of ‘normalisation’ may be partly the result of the persuasive influence of pornography, but it is plausibly also a result of the unconscious influence that vivid images eroticising violence against women exert upon some viewers. In addition, MacKinnon also reminds us that not all pornography is produced by consenting adults. See: Catharine MacKinnon (1992). “Pornography, civil rights and speech”, in *Pornography: Women, Violence and Civil Liberties* (Oxford: Oxford University Press), ed. Catherine Itzin, pp.456-512; Catharine MacKinnon (1995) *Only Words* (London: Harper Collins).

that committing a particular crime is morally justified or to be applauded”.<sup>3</sup> Take the case where incitement involve a speaker promoting attitudes and beliefs that set the stage for harmful actions later down the line – for instance, a speaker at a political rally might describe certain ethnic groups as nothing but murderers and rapists, and a threat to our safety. Jeffrey Howard explains how it is a crime in the United Kingdom to “incite harm through fomenting racial and religious hatred – for example, by expressing the view that citizens from certain groups are morally vile and so deserve to be attacked”.<sup>4</sup> Such a law implies that merely inciting hatred and vilification is sufficiently dangerous as to justify legal prohibition. Along similar lines, Jaconelli points out that U.K. law not only curbs the freedom of speakers to encourage terrorist acts, but also limits the freedom of speakers to recruit others to terrorist causes, and is aimed generally at preventing persons from even being drawn into terrorism in the first place.

The underlying thinking here is that speech is not only dangerous when it involves directly encouraging terrorist acts, but also when it is part of the wider radicalisation process that often precedes this. Philip Baugut and Katharina Neumann agree. Interviewing radicalized Muslim prisoners and former Islamists in Germany and Austria, they found that propaganda (in several different forms) was often used to achieve what they call cognitive and behavioural radicalization, which they classify as a precursor to terrorist acts.<sup>5</sup> Similarly, Elihu Richter et al find that indoctrination and the acceptance of hateful ideologies, and thus the kinds of speech acts that help underpin these processes, are upstream drivers of genocidal violence.<sup>6</sup> In a

---

<sup>3</sup> p.248 of Joseph Jaconelli (2018). “Incitement: a study in language crime”, *Criminal Law and Philosophy* 12 (2), pp.245-65.

<sup>4</sup> p.210 of Jeffrey Howard (2019). “Dangerous speech”, *Philosophy & Public Affairs* 47 (2), pp.208-54.

<sup>5</sup> Philip Baugut and Katharina Neumann (2020). “Online propaganda use during Islamist radicalization”, *Information, Communication and Society* 23 (11), pp.1570-92. See also: Sarah Sorial (2011). *Sedition and the Advocacy of Violence: Free Speech and Counter-Terrorism* (New York: Routledge); Sarah Sorial (2015). “Hate speech and distorted communication: rethinking the limits of incitement”, *Law and Philosophy* 34 (3), pp.299-324.

<sup>6</sup> Elihu Richter, Dror Kris Markus, and Casey Tait (2018). “Incitement, genocide, genocidal terror, and the upstream role of indoctrination: can epidemiologic models predict and prevent?”, *Public Health Review* 39 (30), pp.1-22. While not strictly speaking a case of persuasion, Glover suggests that human beings are able to engage more easily in horrifying behaviour, if at the same time they expressively deny the humanity of their victims. One of the features that characterises gross human abuses is symbolic degradation. Symbolic denials of dignity – including the use of de-humanising language – may help psychologically to remove barriers in the way of any disposition to engage in acts of torture and inhuman treatment. See: Jonathan Glover (1999). *Humanity: a Moral History of the Twentieth Century* (London: J. Cape). See also: Irene Couzigou (2019). “The criminalization of online terrorism preparatory acts under international law”, *Conflict and Terrorism*, pp.1-20; UN Office on Drugs and Crime (2012). “The use of the internet for terrorist purposes”, pp.3-11.

[https://www.unodc.org/documents/frontpage/Use\\_of\\_Internet\\_for\\_Terrorist\\_Purposes.pdf](https://www.unodc.org/documents/frontpage/Use_of_Internet_for_Terrorist_Purposes.pdf)

slightly different vein Ariel Victoria Lieberman looks at the way ISIS uses social media to spread propaganda aimed at young adults. She finds that propaganda glamorising life in ISIS and presenting members as heroic and desirable, is used to win over new followers.<sup>7</sup>

These insights also help ground and contextualise some of the dangers involved in a different kind of harmful persuasion – hate speech. Of course, a good deal of hate speech has nothing to do with persuasion, coming in the form of e.g., face to face abuse, threatening speech, and harassment. For instance, a clear case of hate speech might involve someone verbally racially abusing a stranger in the street. All the same, hate speech also comes in the form of vilifying persons based on their identity or membership in a certain group. Such vilification may be a precursor to violence, when it is used to engage an audience and inflame their feelings of antipathy towards certain groups. Jordan Kiper et al, for instance, studied the effect of hate speech on the way listeners consider ‘out-group’ members.<sup>8</sup> They found that dehumanising language leads to a decrease in empathy for members of the out-group. The authors also argue that attempts by propagandists to exploit historical or social grievances, coupled with a call for revenge in response to these grievances, was particularly liable to decrease out-group empathy, and led in many cases to an increased support for violence. Wiktor Soral et al agree that hate speech increases out-group prejudice. They also suggest that exposure to hate speech can have the effect of desensitising agents to further and more serious acts of verbal abuse.<sup>9</sup> In a similar vein, Ishani Maitra and Mary Kate McGowan argue that hate speech persuades listeners to believe negative stereotypes about certain groups, which in turn removes psychological barriers against engaging in violence against members of those groups.<sup>10</sup> Karsten Müller and Carlo Schwarz focus on online echo-chambers, and find that hate speech can act as a “propagation

---

<sup>7</sup> Ariel Victoria Lieberman (2017). “Terrorism, the internet, and propaganda: a deadly combination”, *Journal of National Security Law and Policy* 9 (1), pp.95-124. Lieberman argues that censorship (including suspending persons from social media sites), alongside other responses such as disseminating counter-propaganda and releasing information that disproves ISIS propaganda, should form part of an effective government response.

<sup>8</sup> Jordan Kiper, Richard Ashby Wilson, Christine Lillie, and Lasana Harris (2019). “Propaganda, empathy, and support for intergroup violence: the moral psychology of international speech crimes”: <https://ssrn.com/abstract=2580521>

<sup>9</sup> Wiktor Soral, Michał Bilewicz, and Mikołaj Winiewski (2018). “Exposure to hate speech increases prejudice through desensitization”, *Aggressive Behaviour* 44 (2), pp.136-46.

<sup>10</sup> Ishani Maitra and Mary Kate McGowan (2012). “Introduction and Overview” in *Speech and Harm: Controversies over Free Speech* (Oxford: Oxford University Press), ed. Ishani Maitra and Mary Kate McGowan, pp. 1-23.

mechanism for violent crimes through enabling the spread of extreme viewpoints”.<sup>11</sup> Finally, Alexander Tsesis argues that when hate speech becomes part of our culturally accepted dialogue, it can promote an attitude of animosity and intolerance against certain groups in such a way as to cultivate an environment supportive of the commission of hate crimes.<sup>12</sup> Ultimately Tsesis seeks to draw a connection between racist, hateful speech acts, and gross injustices like genocide and slavery.<sup>13</sup>

None of this is to imply that hate speech is so powerful as to necessarily lead to violence all by itself. Clearly hate speech is just one part of a much more complicated story about how humans can come to see other humans as acceptable targets of violence. And perhaps hate speech cannot move listeners to act in harmful ways unless a host of other conditions are already met. Still, this needn't imply that hate speech does not have a significant role to play in the violence that it is associated with. Discussing the way people are radicalised into supporting violence, Jonathan Maynard and Susan Benesch write:

Many parts of such radicalizing journeys are not principally ideological: existing policies are replaced by successively more extreme and provocative options, low level violence psychologically habituates perpetrators for more extensive campaigns, complex bureaucratic arrangements develop progressively and expand a group's capacity to organize violence on a mass scale, and so forth. But changes in speech, and the changes in ideologies they reflect and cause, are part of the radicalizing trajectory. In the months or years preceding mass intergroup violence, relevant groups of perpetrators and their supporters come increasingly to see violence as permissible and even necessary. Repertoires of ideas and arguments that encourage such a perception become increasingly formulated and disseminated through speech—spoken, written or

---

<sup>11</sup> Karsten Müller, and Carlo Schwarz (2020). “Fanning the flames of hate: social media and hate crime”: <https://ssrn.com/abstract=3082972>

See also Matthew Williams, Pete Burnap, Amir Javed, Han Liu, Sefa Ozalp (2020). “Hate in the machine: anti-black and anti-Muslim social media posts as predictors of offline racially and religiously aggravated crime”, *The British Journal of Criminology* 60 (1), pp.93-117.

<sup>12</sup> Alexander Tsesis (2013). “Inflammatory hate speech: offense versus incitement”, *Minnesota Law Review* 97 (4), pp.354-405.

<sup>13</sup> Alexander Tsesis (2002) *Destructive Messages: How Hate Speech Paves the Way for Harmful Social Movements* (New York: New York University Press).

otherwise—allowing greater numbers of potential perpetrators to see violence as increasingly thinkable, possible, and justified, and to possess the justificatory resources necessary to convince others.<sup>14</sup>

Finally on the topic of hate speech, note that while the harms discussed so far have all been cases of harmful *actions*, it may be that certain attitudes or beliefs promoted by persuasive hate speech acts, should themselves count as harmful. Consider the case, touched upon above, where hate speech promotes society-wide animosity towards certain groups. Perhaps film and television effectively portrays the Muslim community as animalistic and vile – worthy of animosity. Should this animosity, simply by itself, count as harmful? On the one hand, this may seem like a largely uncontroversial point. After all, defamation laws have long since assumed that people can be harmed by having their reputation damaged, implicitly acknowledging that we might be harmed by how others perceive us. Moreover, a common strand found throughout many political and philosophical accounts – most notably relational accounts of autonomy, as well as accounts of dignity, respect, and humiliation – is that we are in some deep way sensitive to how others think about us.<sup>15</sup> Indeed, several philosophers have problematised hate speech precisely on these grounds. Katharine Gelber and Luke McNamara, for instance, have suggested that promulgating negative stereotypes about certain groups may itself count as harmful.<sup>16</sup> Jeremy Waldron, in at least certain moods, is concerned with the harm of having one’s basic reputation as a member of society in good standing, undermined by hate speech. While Bhikhu Parekh argues that protection of one’s good name and honour are

---

<sup>14</sup> p.76 of Jonathan Maynard and Susan Benesch (2016). “Dangerous speech and dangerous ideology: an integrated model for monitoring and prevention”, *Genocide Studies and Prevention* 9 (3), pp.70-95.

<sup>15</sup> For more on interpersonal recognition and relational autonomy, see: Catriona Mackenzie and Natalie Stoljar (2000). “Introduction: autonomy refigured”, in *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self* (Oxford: Oxford University Press), ed. Catriona Mackenzie and Natalie Stoljar; Catriona Mackenzie (2008). “Relational autonomy, normative authority and perfectionism”, *Journal of Social Philosophy* 39 (4), pp.512-33; Catriona Mackenzie (2013). “The importance of relational autonomy and capabilities for an ethics of vulnerability”, in *Vulnerability: New Essays in Ethics and Feminist Philosophy* (Oxford: Oxford University Press), ed. Catriona Mackenzie, Wendy Rogers and Susan Dodds, pp.33-59.

For more on dignity and humiliation, see: Avishai Margalit (1996). *The Decent Society*. Cambridge, Massachusetts: Harvard University Press; Primo Levi (1987). *If This Is A Man :The Truce* (London: Abacus); Alison Liebling (2011). “Moral performance, inhuman and degrading treatment and prison pain”, *Punishment and Society* 13(5), pp.530-50.

<sup>16</sup> Katharine Gelber and Luke McNamara (2016). “Evidencing the harms of hate speech”, *Social Identities* 22 (3), pp.324-41.

central to the good life, but sees these goods as potentially undermined by the hatred of others.<sup>17</sup>

On the other hand, we must be careful so as not to give an overinclusive definition of harm that counts almost any unfair negative appraisal as harmful. Indeed, one might think that, to some extent, it is important to develop a sufficiently robust sense of one's own worth, such that the ignorant (e.g. racist) contempt of others does not automatically damage one's self-respect.<sup>18</sup> Relatedly, there is the question of just how widespread the social antipathy towards one's social group must be, in order for that antipathy to count as harmful. These issues will be explored in more depth in chapter 5, through a discussion of relational autonomy. For now, it suffices to say that we should be at least open to the idea that being exposed to antipathy from a significant number of one's peers, based on one's identity, may itself count as a harm. And that as such, we can discuss as at least potentially harmful, any kind of hate speech that encourages listeners to see certain groups as disgusting, threatening, untrustworthy, etc.<sup>19</sup>

Pornography represents a related, albeit distinct case of potentially harmful persuasion. Often times when feminists lament the effects of pornography, they are discussing its unconscious effects on the way consumers think about women. All the same, several feminists at least touch upon the idea that part of the toxic influence that pornography has, is the result of its authoritative and vivid depictions presenting a story about women that is then more or less rationally accepted by some consumers. Catharine MacKinnon, for instance, has argued that pornography encourages consumers to see women as mere sexual objects. Judith Hill suggests that this way of representing women is therefore degrading, and harms women qua representing them as mere means.<sup>20</sup> MacKinnon also argued that this influence on the way

---

<sup>17</sup> Bhikhu Parekh (2006). "Hate speech". *Public Policy Research* 12, pp.213-23.

<sup>18</sup> See, for instance: Daniel Statman (2000). "Humiliation, dignity and self-respect", *Philosophical Psychology* 13 (4), pp.523-40.

<sup>19</sup> For more on the harms of hate speech, see: Susan Brison (1998). "The autonomy defense of free speech", *Ethics* 108 (2), pp.312-39; Alexander Brown (2015). *Hate Speech Law: A Philosophical Examination* (New York: Routledge); Alexander Brown and Adriana Sinclair (2019). *The Politics of Hate Speech Laws* (New York: Routledge); Richard Delgado and David Yun (1994). "The neoconservative case against hate-speech regulation - Lively, D'Souza, Gates, Car, D'Souza, Gates, Carter, and the toughlove crowd", *Vanderbilt Law Review* 47 (6), pp.1807-25; Richard Delgado and Jean Stefancic (2001). *Critical Race Theory: An Introduction* (New York: New York University Press); Ishani Maitra (2012) "Subordinating Speech", in *Speech and Harm: Controversies Over Free Speech* (Oxford: Oxford University), ed. Mary Kate McGowan and Ishani Maitra, pp. 94-120.

<sup>20</sup> Judith Hill (1987). "Pornography and degradation", *Hypatia* 2 (2), pp.3-54.

people think about women, contributes significantly to the continuing subordinate position of women. In a similar vein, Gail Dines suggests that pornography is part of a larger system of sexist representations that normalize the economic and political oppression of women.<sup>21</sup> Moreover, many feminists have explored the way in which pornography ‘silences’ women. Again, it is not always clear how much of this is mediated by persuasion. Still, Rae Langton, Jennifer Hornsby, and Caroline West all argue that pornography changes the way (some) consumers think about women, such that they come to misunderstand the words of women.<sup>22</sup> This results in women being effectively ‘silenced’, and thus, in some sense having their freedom of speech undermined.

One final species of harmful persuasion worth reflecting upon, is what we might call anti-science speech. This comes in several forms. Take the issue of climate change denialism. Despite overwhelming evidence for the catastrophic dangers posed by climate change, many still express doubts over the existence of significant anthropogenic climate change, whether the associated harms are really so serious, and whether the costs of fighting climate change outweigh the benefits. The promulgation of these kinds of doubts poses a great danger. Elizabeth Anderson suggests that inaccurate public assessment of scientific topics, including on the issue of climate change, has hampered the capacity of American democracy to respond effectively to urgent problems.<sup>23</sup> In short, a complacent public has allowed the government to stay complacent when it comes to tackling climate change. Brian Leiter agrees. He suggests that “massive public ignorance” around climate change has undoubtedly made it easier for the US government to take no meaningful action in addressing the issue.<sup>24</sup> William Tucker puts matters bluntly: if people remain blinded and continue to ignore the environmental dangers that face us, then the very survival of our species will be imperiled.<sup>25</sup>

---

<sup>21</sup> Gail Dines (2012). “A feminist response to Weitzer”. *Violence Against Women* 18 (4), pp.512-20.

<sup>22</sup> Rae Langton (1993). “Speech acts and unspeakable acts”, *Philosophy & Public Affairs* 22 (4), pp.293-330; Rae Langton and Caroline West (1999). “Scorekeeping in a pornographic language game”, *Australasian Journal of Philosophy* 77 (3), pp.303-19; Jennifer Hornsby and Rae Langton (1998). “Free Speech and illocution”, *Legal Theory* 4, pp.21-37.

<sup>23</sup> Elizabeth Anderson (2011). “Democracy, public policy, and lay assessments of scientific testimony”, *Episteme* 8 (2), pp.144-64.

<sup>24</sup> Brian Leiter (2016). “The case against free speech”, *Sydney Law Review* 38, pp.407-39.

<sup>25</sup> William Tucker (2012). “Deceitful tongues: is climate change denial a crime?”, *Ecology Law Quarterly* 39 (3), pp. 831-94.

Unfortunately, it may be that those preaching climate change denialism have an inherent advantage over those preaching climate change disaster. Howard Koh suggests that many find it difficult to truly engage with the distant and unfamiliar idea of climate change disaster.<sup>26</sup> He found that persuading citizens of the dangers involved in climate change, and convincing them that they can play a role in fighting climate change through their actions, depends on a number of factors, including: framing the issue as a health crisis, emphasising the relatively easy changes to their lives that individuals can make, and tailoring messages to specific groups so that the climate change message is personal. Emma Lozon likewise recommends framing climate change as a health issue, as well as ensuring that the language being used in this discussion is accessible, a co-operative and egalitarian connection with the audience is established, and that the listener is made to feel empowered.<sup>27</sup>

These suggestions underline just how difficult it can be to persuade listeners to appreciate the dangers of climate change, as well as how susceptible people are to being persuaded to the contrary. They also highlight the potential harm of climate change denialism. Anderson, for instance, suggests that the public's failure to grasp the dangerous realities of climate change, is partly the product of misleading media reports. Lozon likewise suggests that the demonisation of experts, as well as inadequate media coverage, are major sources of misunderstanding in public thinking surrounding climate change. Liisa Antilla, having conducted a one-year review (1 March 2003 to 29 February 2004) of US newspaper coverage on climate science, found that the media often promotes misinformation concerning climate change. In particular Antilla points to climate change sceptics being over-represented, mainstream scientists being framed as alarmists, and the inaccurate suggestion that there is still significant debate within the scientific community over anthropogenic climate change, as factors that persuade viewers to continue to question the seriousness of climate change.<sup>28</sup> These findings are especially worrying when paired with another finding of Antilla's; that media coverage is *the* major source of information on climate change for many Americans. Finally, Catriona MacKinnon highlights how climate change deniers use a range of argumentative tactics in order to persuade people of their position. These tactics include:

---

<sup>26</sup> Howard Koh (2015). "Communicating the health effects of climate change", *JAMA Network* 315 (3), pp.239-240.

<sup>27</sup> Emma Lozon (2018). "Communicating the health impacts of climate change: from conviction to persuasion", *Open Access Master's Thesis*, Michigan Technological University: <https://doi.org/10.37099/mtu.dc.etr/603>

<sup>28</sup> Liisa Antilla (2010). "Self-censorship and science: a geographical review of media coverage of climate tipping points", *Public Understanding of Science* 19 (2), pp.240-56.

highlighting the uncertainties in climate models and disagreements between climate scientists in order to give the misleading impression that the science is not established and thus that the grounds for action on climate change are less secure than they in fact are, presenting climate change deniers as more respected in the scientific community than they in fact are, and presenting conspiracy theories that encourage listeners to question whether they can even trust the scientific findings presented by mainstream climate change scientists.<sup>29</sup>

Anti-vaccination speech is a separate category of anti-science speech worth reflecting upon. There is a growing literature on just how anti-vaxxers manage to get their message across in such a persuasive way. Tara Smith has found that a number of different arguments – e.g. that vaccines are toxic, that natural immunity is superior to taking a vaccine, how ‘big pharma’ cannot be trusted – have unfortunately proven persuasive to many, and have consequently led to a distrust of medical practitioners and vaccination procedures.<sup>30</sup> Smith also emphasises that these damaging arguments are popularised not simply by the media, celebrities, and powerful groups, but by ordinary citizens too. Meanwhile, Obdulia Torres González has found anti-vaxxers often exploit a lack of specialist knowledge in order to deliver their message persuasively.<sup>31</sup> In addition, Torres González suggests that susceptibility to anti-vaccination messages is a matter of one’s worldview. Those who hold new age world views are particularly liable to be persuaded by anti-vaccination arguments, because of a worldview that includes (amongst other things) a lack of faith in the scientific community, and a relatively strong faith in alternative ways of fighting against diseases including dietary changes and homeopathic treatments.

Those who engage in this kind of anti-science speech are arguably engaging in harmful persuasion, through encouraging their audience to act in ways that risk harming both themselves and others. Jessica Flanigan, for instance, finds that those who refuse to vaccinate

---

<sup>29</sup> Catriona McKinnon (2016). “Should we tolerate climate change denial?”, *Ethics and Global Climate Change* 40 (1) pp.205-16. See also: Leah Ceccarelli (2011). “Manufactured scientific controversy: science, rhetoric, and public debate”, *Rhetoric and Public Affairs* 14 (2), pp.195-228; Peter Jacques, Riley Dunlap, and Mark Freeman (2008). “The organisation of denial: conservative think tanks and environmental scepticism”, *Environmental Politics* 17 (3), pp.349-85.

<sup>30</sup> Tara Smith (2017). “Vaccine rejection and hesitancy: a review and call to action”, *Open Forum Infectious Diseases* 4 (3), pp.1-7.

<sup>31</sup> Obdulia Torres González (2018). “Claves para comprender la resistencia de los colectivos antivacunas una controversia científico-tecnológica pública”, *Revista de Humanidades de Valparaíso* 11, pp.7-37.

themselves compromise herd immunity in the community.<sup>32</sup> She also found that in several cases, the outbreak of a deadly disease was partly the result of parents refusing to vaccinate their children, as well as that vaccine refusal generally increases the rates of transmission for many diseases. Jason Brennan likewise suggests that those who refuse to vaccinate themselves violate the ‘clean hands principle’, by participating in the collective imposition of unjust harm or risk of harm on others.<sup>33</sup> In addition, those parents who refuse to vaccinate their children expose those children to an unacceptable risk, that in the worst cases results in their death.<sup>34</sup> On top of encouraging listeners to expose themselves and others to risk, anti-vaccination speech may be considered wrong in more subtle ways. Mark Navin, for instance, points out that those who refuse to vaccinate may be guilty of free-riding on herd immunity.<sup>35</sup> Anti-vaccination speech, then, may be considered wrongfully persuasive insofar as it encourages this kind of free-riding.

#### §I. Persuasion-control

The above suffices to both help clarify the concept of harmful persuasion, and to establish a strong prima facie case in favour of what will henceforth be referred to as persuasion control – that is, the regulation of speech based on the persuasion-mediated harms that the speech in question threatens. After all, the government plausibly has a (defeasible) duty to intervene whenever innocent citizens are exposed to harm, or even the significant risk of harm. Despite this, many register a deep-seated anxiety towards permitting persuasion-control. It is one thing, so the thought goes, for the government to fight violence and discrimination through attaching penalties to these actions, or to fight climate change disaster by fining polluters and educating citizens, or to fight hate through education programmes. But it is quite another for the government to fight these ills by restricting the promulgation of ideas and preventing citizens from having access to certain persuasive appeals. Naturally, there are a variety of different ways

---

<sup>32</sup> Jessica Flanigan (2014). “A defense of compulsory vaccination”, *HEC Forum* 26, pp.5-25.

<sup>33</sup> Jason Brennan (2018). “A libertarian case for mandatory vaccination”, *Journal of Medical Ethics* 44 (1), pp.37-43.

<sup>34</sup> Johan Christiaan Bester (2017). “Measles vaccination is best for children: the argument for relying on herd immunity fails”, *Bioethical Inquiry* 14, pp.375-84.

<sup>35</sup> Mark Navin (2015). *Values and Vaccine Refusal: Hard Questions in Ethics, Epistemology, and Health Care* (New York: Routledge).

in which one might cash-out this scepticism towards persuasion-control. Two broad approaches can be discerned.

The first involves exploring practical, empirical-based worries with persuasion-control. Some philosophers, for instance, express doubts regarding the strength of evidence demonstrating a relationship between controversial speech (e.g. hate speech) and the purported harmful consequences that follow (e.g. discrimination and violence).<sup>36</sup> A related but more nuanced worry pushes the thought that while the persuasive speech in question (say, pornography or hate speech) may have some relationship with the associated harms (say, identity-based prejudices in society), this relationship is relatively trivial and weak, especially when compared to the larger forces that are more chiefly responsible for these harms.<sup>37</sup> Some suggest that talk of the harms of speech tends to be overblown or speculative.<sup>38</sup> Others worry that the regulation of speech is likely to prove ineffective, or even liable to backfire and undermine its stated aims.<sup>39</sup> Some hold that there are equally effective, less costly alternatives to fighting against the harms of speech.<sup>40</sup> Some worry that more problematic motives might lurk behind even the government's ostensibly reasonable speech regulation rationales.<sup>41</sup> And of

---

<sup>36</sup> Ronald Dworkin (2009). "Foreword", in *Extreme Speech and Democracy* (Oxford: Oxford University Press), ed. Ivan Hare and James Weinstein, pp. v-ix.

<sup>37</sup> Robert Mark Simpson (2013). "Dignity, harm, and hate speech", *Law and Philosophy* 32 (6), pp.701-28.

<sup>38</sup> Eric Heinze (2013). "Hate speech and the normative foundations of regulation", *International Journal of Law in Context* 9 (4), pp.590-617; Nadine Strossen (2012). "Interview with Nadine Strossen", in *The Content and Context of Hate Speech: Rethinking Regulation and Responses* (Cambridge: Cambridge University Press), ed. Michael Herz and Peter Molnar, pp.378-98.

<sup>39</sup> Edwin Baker (2012). "Hate speech", in *The Content and Context of Hate Speech: Rethinking Regulation and Responses* (Cambridge: Cambridge University Press), ed. Michael Herz and Peter Molnar, pp. 57-80; Sandra Coliver (1992). "Hate speech laws: do they work?", in *Striking a Balance: Hate Speech, Freedom of Expression, and Non-Discrimination* (Article 19, International Centre Against Censorship, Human Rights Centre), ed. Sandra Coliver, pp.363-74; Sue Curry Jansen and Brian Martin (2004). "Exposing and opposing censorship: backfire dynamics in freedom-of-speech struggles", *Pacific Journalism Review* 10(1), pp.29-45; Sue Curry Jansen and Brian Martin (2019). "The Streisand effect and censorship backfire", *International Journal of Communication* 9, pp.656-71; Maxime Lepoutre (2019). "Can "more speech" counter ignorant speech?", *Journal of Ethics and Social Philosophy* 16 (3), pp.155-91.

<sup>40</sup> Maxime Lepoutre (2017). "Hate speech in public discourse: a pessimistic defense of counterspeech", *Social Theory and Practice* 43(4), pp. 851-83; Geoffrey Stone (1994). "Hate Speech and the U.S. constitution" *East European Constitutional Review* 78, pp.78-82; Nadine Strossen (2018). *Hate: Why We Should Resist It With Free Speech, Not Censorship* (Oxford: Oxford University Press).

<sup>41</sup> Larry Alexander (1996). "Banning hate speech and the sticks and stones defense", *Constitutional Commentary* 600 (13), pp.71-100; Ronald Dworkin (1981). "Is there a right to pornography?", *Oxford Journal of Legal Studies* 1 (2), pp. 177-212.

course, there is the worry that censorship, while perhaps justifiable now, will lead via a slippery slope to worse and worse forms of government interference.<sup>42</sup>

A second broad approach involves expressing a rather more deontological, principled worry with persuasion-control. On this approach, persuasive speech is said to conflict with our deep moral and political values. The value of autonomy is almost always central to these arguments. For instance, allowing agents to engage one another in persuasive speech is framed as a matter of permitting agents to use their autonomous, intellectual capacities to decide for themselves what to think. These deliberative powers are described as central to what makes us unique and morally valuable, and it is supposed that we each have a deep interest in being free to exercise and develop those deliberative powers, shielded from outside interference.<sup>43</sup> But the argument doesn't end here. Rather, it is suggested that some of our deepest moral and political values require that we take deliberative autonomy seriously. The supposed upshot of all of this is that, unlike the regulation of directly harmful speech (e.g. the regulation of blackmail, threats, harassment, etc), the regulation of harmful persuasion is said to uniquely threaten our moral and political commitments to taking autonomy seriously.

Of course, the ways in which the value of autonomy is construed varies from one philosopher to another, as does the vision of how autonomy connects to other values. Still, what remains constant is the suggestion that reflecting on our deepest values, particularly that of autonomy, establishes a powerful argument against persuasion-control. Take Ronald Dworkin's defence of the First Amendment. In typically strident fashion, he states that "we are a liberal society committed to individual moral responsibility, and any censorship on grounds of content is inconsistent with that commitment".<sup>44</sup> Thomas Nagel describes restricting

---

<sup>42</sup> For a good summary of such arguments, see: chapter 5 of Brown and Sinclair, *'The Politics of Hate Speech Laws'*; Frederick Schauer (1985). "Slippery slopes", *Harvard Law Review* 99 (2), pp. 361-83.

<sup>43</sup> See for instance: Eric Cave (2007). "What's wrong with motive manipulation?" *Ethic Theory Moral Practice* 10 (2), pp.129-44; John Christman (1991). "Autonomy and personal history", *Canadian Journal of Philosophy* 21 (1), pp.1-24; Gerald Dworkin (1988). *The Theory and Practice of Autonomy* (Cambridge: Cambridge University Press); John Stuart Mill (2003) [1859]. *On Liberty*, in *Utilitarianism and On liberty: including Mill's Essay on Bentham' and selections from the writings of Jeremy Bentham and John Austin* (Malden, Massachusetts: Blackwell Publishers), ed. Mary Warnock, pp.88-180. See especially his thoughts on the importance of deliberation as a higher capacity, and its role in the good life; Martin Redish (1982). "The value of free speech", *University of Pennsylvania Law Review* 130, pp.591-645. Consider also how in the U.S., freedom of thought is protected at a constitutional level by the bill of rights.

<sup>44</sup> Ronald Dworkin (1996). *Freedom's Law: The Moral Reading of the American Constitution* (Cambridge, Massachusetts; Harvard University Press).

persuasive appeals as a kind of “thought control” and “worst” amongst the various species of censorship. In a withering assessment he states that “[a]part from its epistemological stupidity this is the ultimate insult not only to the dissenters but to the rest of us, their potential audience, who are not trusted to make up our own minds”.<sup>45</sup> David Strauss characterises persuasion-control as a form of mind-control, and argues that it is an unacceptably manipulative interference with the flow of information.<sup>46</sup> While David Richards argues that “[o]ur principles are...best and most reasonably affirmed when we resist the temptation to respond to bigots in kind and insist on embracing them in an inclusive moral community that recognizes in all persons what some of them might willfully deny to others, the equality of all persons as free and reasonable members of a political community of principle. Protecting the rights of the speakers and speech we hate affirms the deeper fraternal bonds of a political community based on universal human rights”.<sup>47</sup>

## §II. The focus of the thesis

This thesis focuses on the second kind of approach. That is, on the deontological approach. Narrowing our focus further still, the thesis draws upon the works of T. M. Scanlon, Thomas Nagel, James Weinstein, and Ronald Dworkin, in order to consider whether there is a principled conflict between persuasion-control and our democratic values.<sup>48</sup> In particular it is asked whether our democratic values entail that citizens should be free to hear all the available arguments and ideas, and to determine for themselves what to think. In this sense, the arguments canvassed in this thesis are ‘listener-based’ – problematising persuasion-control, not from the standpoint of those seeking to engage in persuasive speech, but rather from the standpoint of those who might engage with the persuasive speech of others.

But why democracy? Why think that governmental interferences with our ability to engage with the persuasive speech of others, in any way conflict with our democratic values? A

---

<sup>45</sup> p.97 of: Thomas Nagel (1995). “Personal rights and public space”, *Philosophy & Public Affairs* 24 (2), pp.83-107.

<sup>46</sup> David Strauss (1991). “Persuasion, autonomy, and freedom of expression”, *Columbia Law Review* 91 (2), pp. 334-71.

<sup>47</sup> pp.821-2 of: David Richards (1999). “Constitutional legitimacy, the principle of free speech, and the politics of identity”, *Kent Law Review* 74 (2), pp.779-822.

<sup>48</sup> As will be explained below, both ‘deontological’ and ‘principled’ have technical meanings here.

common theme in free speech debates is that some of our core democratic values are tied up with our ability to engage in discussion with others. As we will see in chapter 2, T.M. Scanlon suggests that being free to decide for one's self what to believe and how to act, forms the seat of morally responsible autonomous agency.<sup>49</sup> Building upon this insight, he suggests that our democratic institutions ought to address themselves to us at this level of being a morally responsible, thinking agent. In a somewhat similar vein, Ronald Dworkin argued that a constitutive feature of a just political society is that it treats its citizens as responsible agents that are free to think for themselves.<sup>50</sup> Famously John Rawls' commitment to Kantian constructivism led him to hold that a certain political ideal of citizens as "free and equal" (part of which involves viewing citizens as possessing the capacity to form, revise and pursue a conception of the good), should ultimately constrain and inform our first principles of justice.<sup>51</sup> Edwin Baker argued that the values of democracy and a respect for autonomy demand that citizens are afforded broad speech freedoms.<sup>52</sup> And Robert Wolff's brand of political anarchism at least partly rested on his contention that being truly intellectually autonomous was in fact incompatible with also being under an obligation to act as the state commands simply because the state has commanded it.<sup>53</sup> In short, political philosophers have long supposed that part of our democratic vision involves the government appropriately recognising and respecting the autonomous capacities of its citizens. This thesis seeks to explore this common theme and whether it can establish a case against persuasion-control.

Above it was suggested that the arguments canvassed in this thesis are 'deontological'. But what does this mean? In simple terms, it means that they seek to move beyond merely evaluating persuasion-control from the standpoint of how desirable its expected outcomes are, and instead shift towards evaluating it in terms of the related values of democracy and autonomy. In other words, Scanlon, Nagel, Weinstein, and Dworkin are united in thinking that certain forms of censorship conflict with our democratic values. Indeed, as we will see, these philosophers can be read as arguing that permitting certain forms of censorship

---

<sup>49</sup> T.M. Scanlon (1972). "A Theory of Freedom of Expression", *Philosophy & Public Affairs* 1 (2), pp. 204-26.

<sup>50</sup> Ronald Dworkin, 'Freedom's Law: The Moral Reading of the American Constitution'.

<sup>51</sup> John Rawls (1980). "Kantian constructivism in moral theory", *The Journal of Philosophy* 77 (9) pp.515-72; John Rawls (1972). *A Theory of Justice* (Oxford: Oxford University Press).

<sup>52</sup> Edwin Baker (2011). "Autonomy and free speech", *Constitutional Commentary* 27, pp.251-82.

<sup>53</sup> Robert Wolff (1970). *In Defense of Anarchism* (New York: Harper & Row).

ultimately impairs the political legitimacy of the government. More than this, these philosophers argue that the conflict between censorship and our democratic values is a principled one. We discuss what this means in detail below, but for now it suffices to say that the idea is that it is the very nature of act censorship itself, that places it in conflict with our democratic values. This idea is quite distinct from the suggestion that censorship is sometimes, or as a matter of contingent fact tends to, have certain negative side-effects that conflict with our democratic values. These philosophers are also unified in that they, although not necessarily exclusively, consider these questions of democracy and free speech from the standpoint of the (would-be) listener). That is to say, unlike, say, Baker, the philosophers that this thesis discusses in detail, argue that censorship in-principle conflicts with our democratic values, because of the way it impacts those who might listen and consider the censored arguments.<sup>54</sup> That is to say, these philosophers argue that censorship fails to properly respect, recognize, treat fairly, etc, *the would-be listeners* affected by this regulation. Naturally, these arguments touch upon what it means to be an autonomous, adult, responsible agent, who has the power to consider arguments for themselves and come to their own conclusions. It is notable that this model of the political citizens orientates virtually all the arguments found in this thesis.

Moreover, the arguments presented in this thesis are united in seeking to justify what we might call a strong freedom of speech approach. According to strong freedom of speech approaches, speech (or some particular kinds of speech, such as persuasion) deserve special protection from government interference. On such an approach, government interferences with our speech are acceptable, if ever, in special cases only. Pointing out that a certain kind of speech is harmful does not, on this approach, necessarily show that government interference is warranted. Reflecting on the deontological nature of the arguments in question helps to explain how they might establish such a strong freedom of speech approach. For if regulating speech (or a certain kind of speech) involves undermining our core democratic values, then it may be that regulation is unacceptable even in some cases where it helps prevent serious harms. In this way, the strong freedom of speech advocate needn't deny that the speech they seek to protect is dangerous or harmful. They may well concede, for instance, that incitement, hate speech, violent pornography, climate change denialism, etc. are harmful. Nonetheless, they

---

<sup>54</sup> Edwin Baker (1989). *Human Liberty and Freedom of Speech* (Oxford: Oxford University Press)

judge that these harms are at least sometimes worth tolerating in the name of democracy. As one might say, this is ‘the price we pay for democracy’.

Note that in this way, there is a rhetorically powerful, almost inspirational tone to the deontological freedom of speech arguments this thesis canvasses. One of the key factors that unifies the arguments of Scanlon, Nagel, Weinstein, and Dworkin, is that advocates of a strong freedom of speech position, evoke deep values and principles. They ground their position in the meaning of democracy, and presents us with sketches of what the liberal, democratic state ought to be like. And like the supporters of fair elections and just criminal trials, they tell us that these values are so important as to be worth suffering for. With such proselytising argumentation in its corner, perhaps it is little wonder that free speech has attracted widespread and loyal support.

Of course, this kind of free speech approach is far from uncontroversial. Some will object, for instance, that this principled free speech approach originates from a less than noble source, such as an unduly indignant attitude towards even reasonable governmental interference, or an ultimately unjustified distrust in the government’s ability to handle speech issues. Some will worry that this kind of striking free speech approach is historically contingent; doomed to be looked back upon by future generations as unjustified and unenlightened. Another worry is that the arguments offered for this approach are dialectically inefficacious; music to the ears of the converted, but unlikely to win over non-believers. For instance, while several of the philosophers discussed in this thesis may want to characterise persuasion-control as a form of thought-policing of mental coercion, these descriptions will strike many as unhelpful poetry. Moreover, one might question whether the values of democracy and autonomy straightforwardly tell in favour of freedom of speech and against persuasion-control, or whether a more complicated and nuanced picture is more appropriate. Susan Brison, for instance, explains how many of the autonomy-based arguments used to justify a strong freedom of speech approach, may in fact be ‘inverted’ so as provide arguments in favour of limited government censorship.<sup>55</sup>

Finally, while persuasion-control comes in many different forms, this thesis discusses the insights of Scanlon, Nagel, Weinstein, and Dworkin, as they apply to a number of core

---

<sup>55</sup> Brison ‘The autonomy defence of free speech’.

examples: hate speech (including the promotion of terrorism), pornography, anti-scientific speech (including climate change denialism and anti-vaccination speech). The justification for this focus is partly practical, and partly argumentative. On the one hand, these are a cluster of cases in which we can claim to have some philosophical expertise, and thus can discuss in relation to theories of free speech with the appropriate level of philosophical nuance. It is our opinion that virtually any case of dangerous speech is bound to be controversial and complicated, and so we think it is important to focus on a cluster of cases that we have substantial research experience in, if we are to do these cases any kind of justice. On the other hand, we find these cases to be both practically important, and philosophically interesting. While it can be tempting to focus on highly abstract cases, the advantage of focusing on these very much alive cases, is that they lend themselves to being described with a certain richness and nuance, that serves, we think, for better philosophical discussion.

### §III. Methodology

The primary aim of this thesis is to investigate whether a deontological, democracy-based objection to persuasion-control, can be cashed-out and contextualised in a way that responds to these worries. That is to ask: can we cash-out an objection to persuasion-control that is both philosophically rigorous and somewhat broadly appealing? At this point some discussion of the methodology deployed in this thesis is necessary. In the most basic sense, testing the philosophical rigor of an argument involves investigating whether that argument is logically valid, and whether the premises of that argument are true. Unfortunately, when we are dealing with deep and subtle arguments, even this is no mean feat. Often our critical evaluation revolves around the complicated concepts deployed in the arguments – concepts like autonomy, democracy, respect, recognition. With this in mind, what becomes most important is to ascertain how the philosopher construes and uses these concepts in their argument. Ultimately we need to discover whether the value really grounds the conclusion the philosopher is trying to derive. In practice, this involves the process of reflective equilibrium wherein we consider whether the philosopher’s theorising about the value, coheres with our considered judgements and intuitions about that value, as well our judgements and intuitions more generally. In the case where there is a lack of coherence, we must consider whether the philosopher’s theorising should be modified, or whether we should reconsider our judgements

and intuitions. Equally important, is whether the philosopher has explored the value in their argument in a consistent manner. In the most simple case, this involves checking that the philosopher has not directly contradicted themselves when discussing the value. More challenging cases involve considering whether the philosopher has ignored certain implications of their discussion of the value, and whether the philosopher has focused on one kind of threat to the value while ignoring others.

The above considerations speak to the idea what constitutes a ‘philosophically rigorous’ argument. But what of testing whether an argument is ‘somewhat broadly appealing’. How should we investigate this matter? For starters, we can ask whether we can, through reflective equilibrium, reach a state of coherence between the philosopher’s theorising and our intuitions. We suggested above that this was a matter of coherence, but it also sheds light on how well the philosopher’s theorising accommodates our intuitions, and thus how likely readers are to find the theorising to be appealing. Beyond this, the thesis often evaluates an argument by asking: at which point does the argument rely on brute assumptions? Often when we are dealing with complicated philosophical arguments that evoke deep values such as democracy, legitimacy, respect, etc, it is inevitable that the argument will involve brute assumptions. After all, it is unclear how one could categorically ‘prove’ that, say, our democratic institutions should fundamentally be about ensuring that hypothetical reasonable citizens could consent to the government’s actions, rather than securing actual consent from citizens. Both these democratic models have their strengths and weaknesses, and reasonable persons are wont to disagree on such subtle, philosophically complicated matters. With this in mind, we can make sense of how broadly appealing a theory is, by considering at what level in the argument we find these brute claims, and whether anything can be offered in favour of these brute claims. Granted, we may not be able to justify these brute claims from the ground-up (they are ‘brute’, after all), but this doesn’t mean that no philosophical evidence can be offered. Again, reflective equilibrium proves to be a useful tool insofar as we can reflect upon how well these brute intuitions marry with our wider judgements and intuitions. Can we, for instance, tell a story about these brute claims that shows them to cohere with other values we hold? Do these brute claims cohere with our intuitions and considered judgements? Or would accepting these brute claims require us to rethink our other values? These are the kinds of questions this thesis asks, when considering the ‘broad appeal’ of a philosophical argument.

In terms of methodology, another important question is how this thesis approaches dealing with the concepts (e.g. harm, autonomy, respect, etc) that are so central to its investigation. We take the meaning of these concepts to be not only complicated, but contested by philosophers and non-philosophers alike. After all, there is no one single meaning of concepts like ‘autonomy’ or ‘harm’. How then, should we understand these concepts as they are used in this thesis? On the point, we do well to follow Seana Shiffrin.<sup>56</sup> In her analysis of the concept of paternalism, she argues that from the standpoint of conducting a moral analysis, we should not worry too much about whether our use of ‘paternalism’ is extensionally accurate (in line with how the word is used in everyday speech). Similarly, we should not be overly concerned with ensuring that our use of the word perfectly coheres with how the word is used by philosophers. On the contrary, what really matters is that our moral analysis of the concept, *as we understand it*, makes salient philosophical and moral points. In other words, when we are looking at Scanlon’s work on speech and autonomy, we should be concerned so much with whether his use of these terms is extensionally accurate or not. Rather we should focus on whether Scanlon is right to suppose that the phenomena he isolates and calls speech, is valuable on the grounds that it promotes what he understands as people’s ‘autonomy’. In this way, we can philosophically evaluate the argument that Scanlon is making without having to worry unduly about the labels he deploys. Similarly, when we spoke earlier of conducting reflective equilibrium, what matters is that e.g. Scanlon’s theorising about what he calls autonomy, is in some sense coherent with our intuitions and convictions. In this context, what matters here, for instance, is that Scanlon’s theorizing does not directly conflict with say our deeply held intuitions about the needs of people or the role of the government. What is not important, is whether Scanlon’s use of ‘autonomy’ matches up with how we commonly use the word ‘autonomy’. Indeed, in this sense, it is perfectly possible for Scanlon’s theorising about ‘autonomy’ to look very different to our moral convictions about what we understand by autonomy. Put simply, the thesis tries to focus on the philosophical analysis being offered by some of the great political philosophers, and whether this contradicts with our general convictions, and not so much on the terminology being deployed.

---

<sup>56</sup> Seana Shiffrin (2000). “Paternalism, unconscionability doctrine, and accommodation *Philosophy & Public Affairs* 29 (3), pp.205-50.

With these points clarified, what follows is a brief synopsis of the chapters that make up this thesis. Chapter 2 reflects upon T.M. Scanlon's treatment of censorship in his 1972 article, 'A theory of freedom of expression'. The chapter considers some of the more straightforward ways in which persuasion-control might undermine the autonomy of citizens. It is argued that contrary to the alarmist language sometimes deployed by philosophers, persuasion-control does not necessarily involve a special interference with the autonomy of citizens – at least not in the conventional senses under consideration. The chapter also reflects upon Scanlon's worry that persuasion-control conflicts with our democratic values, and thus constitutes a threat to government legitimacy. It is argued that while reflecting on our democratic values may be a promising way to approach problematising censorship, Scanlon's own assumptions about democracy are unhelpfully brute and in need of unpacking.

With this in mind, chapters 3-5 each explore a different way of framing the conflict between persuasion-control and our democratic principles, in an attempt to unpack Scanlon's democratic concern. Chapter 3 explores the worry that persuasion-control constitutes an unacceptable insult to its citizens, or in some sense involves a failure to properly recognise and respect citizens. The chapter draws on Nagel's discussions of rights and status. In the end it is argued that, contra Nagel, persuasion-control may in fact be compatible with a rather rich appreciation and recognition of citizens. And that reflecting on the duties citizens owe one another is key in this respect.

Chapter 4 considers whether persuasion-control conflicts with a certain democratic vision of how the government ought to operate in matters of speech. The chapter reflects upon James Weinstein's worry that censorship involves some form of viewpoint discrimination. While it is suggested that the arguments presented by Weinstein do not succeed, his account nonetheless tips us off to a justified worry about the stigmatising effect that persuasion-control might have on 'innocent' citizens. The chapter closes by considering how we might go about mitigating, or at least reframing, this issue of stigma.

Finally, chapter 5 attempts to synthesise the approaches from chapters 2 and 3; arguing that we should problematise persuasion-control in terms of its impact upon the government-citizen relationship. Drawing on relational accounts of autonomy presented by the likes of Catriona Mackenzie, it is suggested that persuasion-control threatens the standing of citizens within the government-citizen relationship.

The purpose of this first, introductory chapter, other than providing a general outline of the thesis, is to clarify some of the key concepts around which the thesis orbits. This chapter also responds to some of the worries a reader might have at this early stage. The remainder of this chapter proceeds as follows: §IV clarifies what we mean when we talk about ‘deontological’ and ‘principled’ arguments in favour of persuasion-control. Clarifying this issue allows us to outline in more detail the main conclusions of this thesis. §V takes a closer look at the concept of ‘persuasion’. §VI outlines a taxonomy of persuasion-mediated harms, in order to further clarify the kind of harms this thesis is interested in. §VII draws the chapter to a close by canvassing and responding to a series of preliminary worries a reader might have.

#### §IV. Deontological arguments

The arguments canvassed in this thesis have been characterised as ‘deontological’. But like so many ideas in philosophy, the concept of ‘deontology’ has been interpreted in subtly different ways by a wide variety of different philosophers. What, precisely, does ‘deontological’ mean in this context? What does it mean to argue in favour of freedom of speech, from the deontological value of democracy?

For our purposes, the following three points are worth highlighting. First, deontological arguments revolve around values that have normative weight, independent of the boons in welfare or other consequentialist benefits they might secure. Making a promise, for instance, obligates the promise-maker, irrespective of the happiness that keeping the promise might secure. As such, we might say that promises exert deontological normative force.

Second, deontological values are supposed to be in some sense partly immune to being compared with and outweighed by consequentialist (e.g. welfare-based) counter-considerations. For instance, the promise-maker is supposed to be bound by her promise, even when she knows that breaking the promise might accrue more overall happiness. In a similar vein, one might plausibly think that democracy and justice obligate us to abide by fair election procedures and fair trials, even when we are confident that rigging the election and fixing the trial would maximise overall social welfare. Granted, there may well be limitations on this kind of deontological reasoning. All but the very staunchest deontologists will admit that it is sometimes permissible to break a promise or even rig an election, if doing so is necessary to avoid a sufficiently huge catastrophe. Nonetheless, the essential point remains; in standard

cases deontological values, unlike consequentialist goods, do not form just another part of the calculation.<sup>57</sup>

Third, it should be noted that even some deontological arguments have a certain consequentialist flavour to them. In particular, some arguments highlight how government censorship has consequences that can be evaluated negatively from the standpoint of our deontological, democratic values. For instance, Alexander Meiklejohn suggests that robust free speech protection helps facilitate e.g., discussion, criticism, exchange of information, which in turns helps facilitate the democratic notion of self-governance e.g., in the form of informed decision-making.<sup>58</sup> In essence, then, we have a deontological argument in favour of free speech and against persuasion-control, that is based on the positive consequences of embracing the free speech approach and the negative consequences of permitting persuasion-control. By and large, the arguments canvassed in this thesis do not have even this consequentialist flavour.

---

<sup>57</sup> The distinction drawn here between deontology and consequentialism, largely rests on the kind of values these systems are interested in. In particular, the distinction rests on the idea that deontology is about deontological values such as democracy, while consequentialist deals with welfarist values. The implication is that these two values are different in kind; that they function differently and belong to different schools of value. To further fortify this distinction, we noted that deontological and consequentialist values are not straightforwardly comparable, and cannot be straightforwardly traded-off against one another. This was demonstrated by observing how deontological values like the value of promise-keeping and democracy, are in some sense immune to consequentialist (e.g. welfarist) counter considerations. In particular, deontological values are supposed to, at least to some degree, suspend our welfarist counter calculations.

Some will object to this. In particular, sophisticated consequentialists will argue that deontological values like promise-keeping and democracy can in fact be subsumed under the same metric. In particular they will argue that both may be subsumed under the metric of 'the good' (or perhaps simply 'what is valuable in the world'), and that the clear-thinking consequentialist simply wishes to promote the good. To fortify this suggestion about 'the good', the sophisticated consequentialist will try to draw comparisons between deontological values and welfarist values in order to demonstrate their compatibility and kinship. They will point out that both are values. They will also point to how deontological values (e.g. of promise-keeping and democracy) can plausibly be sacrificed in the name of promoting consequentialist goods such as welfare, so long as the welfare gains in question are sufficiently huge.

In response to this, and in defence of our original distinction, let us offer two points. First, the observation about how deontological and welfarist values can sometimes be traded off against one another, does not necessarily supports the sophisticated consequentialist's case. Granted, if ostensibly deontological values and consequentialist values could always be traded off against one another, this would support the sophisticated consequentialist's position. But the idea that sometimes these values are comparable and sometimes are not, at best leaves the question undecided. At worst, it tells against the sophisticated consequentialist's position, especially seeing as they lack a non-arbitrary way of distinguishing between which do and do not permit of trading off. Second, we needn't solve these subtle and complicated debates about deontology, in order to accept our thesis. Admittedly, it is helpful to use the label of deontology in order to help characterise the arguments canvassed in this thesis. All the same, if it turns out that these seemingly deontological arguments can in fact be 'consequentialised' in some subtle way, this will not undermine the substance of the arguments being presented.

For further discussion on sophisticated consequentialism, see: Peter Railton (1984). "Alienation, consequentialism, and the demands of morality", *Philosophy & Public Affairs* 13 (2), pp.134-71.

<sup>58</sup> Alexander Meiklejohn (1948). *Free Speech and Its Relation to Self-Government* (New York: Harper and Brothers).

Rather, they suppose that persuasion-control lies in conflict with our democratic values, regardless of what positive or negative consequences it might have. For instance, chapter 3 explores the argument that persuasion-control communicates something insulting about those whose freedom it limits, as well as the argument that persuasion-control betrays a lack of recognition for those same citizens. These worries do not revolve around the supposed consequences of persuasion-control. On the contrary, the idea here is that certain speech regulations are simply of a *form* that should worry those who care about democracy. Or, to put it another way, we might say that such arguments take there to be a principled conflict between persuasion-control and our democratic values. As such, let us say that this thesis predominantly focuses on arguments from democracy that are both ‘deontological’ and ‘principled’.

Even at this very early stage, a reader might have worries with this kind of deontological approach to problematising censorship. Let us here note and respond to two such worries. First, one might wonder whether the deontological critique of persuasion-control asks too much; namely that we absolutely prohibit persuasion-control in all instances. After all, it might be difficult to see how a philosopher like Nagel or Scanlon could possibly leave the door open to permitting persuasion-control, while simultaneously painting it as in deep, principled conflict with some of our most important values. For many, including even those who share the worry that persuasion-control might be deeply problematic, such an absolute prohibition will seem too extreme.

In response, let us note that it is indeed possible to provide a deontological and principled critique of persuasion-control, while still leaving the door open to some cases of permissible persuasion-control. Three points will help to substantiate this suggestion. First, it is helpful to recall the distinction between all things considered wrongs, and pro tanto wrongs. The deontological critic of persuasion-control may cash out her position by holding that even the best, all things considered justified persuasion-control, nonetheless comes at a significant pro tanto cost to our democratic values. That is to say, cases of all things considered justified conduct that involve us betraying our deontological values, “leave us with something to regret”.<sup>59</sup> Something has gone wrong when we decide to break a promise or cheat on a test or

---

<sup>59</sup> This phraseology is taken from an email Ronald Dworkin sent to Jeremy Waldron, which would ultimately go on to frame some of the debates between Waldron and James Weinstein. See: James Weinstein (2017). “Hate speech bans, democracy, and political legitimacy”, *Constitutional Commentary* 32 (3), pp.527-83; Ronald Dworkin

rig an election, even when doing so is all things considered justified. In these cases we are left with a kind of negative moral residue that, while justified, is neither undone nor scrubbed away. Of course, much rides on the deontological value in question. Breaking a promise made to a close friend leaves one with more to regret than breaking a promise made to a stranger. Rigging an election warrants even more regret still. Our deontological critic will have made her point if she can show that, for instance, the deontological values sacrificed in the censorship case render it more akin to the election case than the promise case.

Second, our critic can plausibly hold that free speech is tied up with democracy in such a way that weighty countervailing considerations are required if censorship is to be rendered all things considered justified. Along these lines, our critic may insist that the government should generally erect high thresholds for what is considered justified interference e.g. in the form of requiring strong evidence of the harmfulness of speech that is being interfered with, requiring that the harms in question are of a sufficiently great magnitude, requiring that alternative, equally costly (or less costly) means of combatting the harms in question are unavailable, requiring strong evidence that the harms in question will be alleviated by the proposed suppression of speech.

Third, our critic may hold that permitting censorship requires some kind of reimagining of what the liberal, democratic project is in which we are involved. It may be that the censorship, while all things considered justified, is incompatible with certain attractive and popular conceptions of democracy. If this is so, then our critic will have shown that those who support persuasion-control may have to reimagine the democratic project, or give a renewed description of the democratic project, in order to explain why persuasion-control can be considered a democratically justified measure.

The second worry a reader might have, involves a scepticism about the kind of high-minded, deontological arguments being presented in favour of free speech. Stanley Fish, for instance, has suggested that deontological arguments in favour of free speech are likely to collapse into consequentialist rationales, or else collapse altogether.<sup>60</sup> That is, that despite all

---

(2012). "Reply to Jeremy Waldron", in *The Content and Context of Hate Speech: Rethinking Regulation and Responses* (Cambridge: Cambridge University Press), ed. Michael Herz and Peter Molnar, pp. 341-4.

<sup>60</sup> Stanley Fish (1994). *There's No Such Thing as Free Speech; And It's a Good Thing Too* (New York: Oxford University Press).

the deontological hot air and purple prose, free speech regulations are justified, if at all, by the fact that in the long run they secure the best consequences. Something similar is at play in Brison's suggestions that some of these deontological arguments in favour of free speech can be 'inverted' and re-assessed from the standpoint of the damage that free speech does to our deontological values, as well as Susan Easton's critique of autonomy-centred arguments in favour of freedom of speech.<sup>61</sup>

However, it is worth noting that this kind of attempt to collapse deontological free speech rationales into consequentialist rationales may well be more costly than it first appears. A whole host of our formal political and legal commitments, which also form the backdrop for any discussion of these free speech issues, embrace the notion that deontological commitments ought to guide our political theory. Consider the steadfast commitment to dignity outlined in the first lines of the German constitution, or the more subtle commitment to democracy and self-governance expressed in the U.S. constitution, or the numerous unconditional commitments and values that loom so large in the European Court of Human Rights' 'European convention on human rights'.<sup>62</sup> These considerations do not defeat the worry canvassed above. Someone might still reflect upon the universal declaration of human rights and its protection of freedom of speech alongside its stated commitment to dignity and equality, and conclude that these deontological values are a lot of hot air.<sup>63</sup> Even so, it is clear that talk of deontological commitments have become entrenched in our political theorising. As such, without very strong argumentative evidence, we should be wary of so quickly abandoning talk of the deontological value of free speech.

Having clarified the deontological nature of the scepticism this thesis investigates, this section closes by briefly sketching the two main conclusions that this thesis will defend: (i) the most common deontological objections to government censorship turn out to fall rather flat. For instance, the oft repeated worry that censorship is insulting turns out to rest on a rather impoverished, overly rationalistic view of persons. Similarly, demands for the government to treat citizens as morally responsible tend to collapse into an unrealistic vision of citizens as

---

<sup>61</sup> Susan Brison, 'The autonomy defense of free speech'; Susan Easton (1995). "Autonomy and the free speech principle", *Journal of Applied Philosophy* 12 (1), pp. 27-39.

<sup>62</sup> <https://www.btg-bestellservice.de/>; <https://constitutionus.com/>

<sup>63</sup> <https://www.un.org/en/universal-declaration-human-rights/>

excellent moral reasoners, that rather misses the point of a functioning democratic state. Moreover, several of the approaches to problematising censorship canvassed in this thesis are revealed to be myopic insofar as they focus either just on the role of the government in a democracy or just the role of the citizen in a democracy.

(ii) a better approach to problematising censorship combines these two types of critique. Taking inspiration from the richer, feminist, relational accounts touched upon above, it is argued that a more productive way of substantiating our sceptic's position is through looking at what kind of government-citizen relationship we should strive for in a democracy. Clearly this involves a shift away from thinking in isolation about the roles of the government and citizen in a democracy, and towards thinking about how these two roles relate to one another. It also involves, in contradistinction to someone like Steven Heyman, a shift away from just thinking about how our relationship with other citizens impacts upon our dignity and standing.<sup>64</sup> Ultimately it is suggested that censorship puts pressure on the government-citizen relationship. In particular, permitting persuasion-control may unbalance the government-citizen relationship insofar as it grants governments a special kind of asymmetric moral authority over their citizens, in judging what kind of beliefs are permissible and what kind of beliefs are beyond the pale. The thesis concludes that precisely how much this changes and damages the government-citizen relationships, and whether we should tolerate this kind of change, depends on how transparent the government is in its operations, and how effectively it meets the needs of its citizens.

In the end then, the thesis takes a somewhat sympathetic view of the principled anxiety many feel towards censorship. This is a somewhat appropriate conclusion for a thesis that started life as an attempt to try and better understand the kind of anxiety that many feel towards censorship. Still, this conclusion does lie in tension with some parts of the thesis. We have already, for instance, touched upon a whole range of different accounts of the harm that persuasion can do. These harms establish a strong case for government intervention. Indeed, in chapter 5 we will discuss how insights from the relational autonomy literature help contextualise some of these harms. Moreover, it has not escaped our attention that in modern political discourse, this kind of anxiety is often expressed by those with less than noble aims.

---

<sup>64</sup> Heyman, 'Hate speech, public discourse, and the first amendment'.

Robert Mark Simpson, for instance, points to the rank hypocrisy that underpins recent calls by conservative politicians to take freedom of speech seriously.<sup>65</sup> Indeed, all of us are probably aware of how often talk of the importance of freedom of speech is underpinned by a rather ugly individualism that frames virtually any government interference as outrageous, and denies that citizens have duties to protect one another. As will become clear, this thesis wholeheartedly rejects this individualism. In its place is presented a rather more grounded vision of citizens as flawed reasoners who must co-operate with one another, if flourishing is to be achieved on any kind of large scale. We are, as Mackenzie points out, essentially vulnerable.<sup>66</sup> We are vulnerable in our reasoning. We are vulnerable to our complicated and varied needs. And we are vulnerable in our dependence on one another. All of this speaks in favour of construing government regulation, including speech regulation, as a force for good. We do not thrive left to our own devices, but rather benefit from proportionate and sensible government interference.

Nonetheless, there is something deeply interesting in the deontological anxiety towards censorship. Notwithstanding the points touched upon just above, a legitimate concern for how citizens stand before their government can usefully inform how we discuss the problem of speech harms. For even those of us who are generally sympathetic towards government interference and recognise the damage that speech can do, may still be duly anxious about granting the government the moral authority to influence what kinds of beliefs and intentions its citizens may permissibly develop. Indeed, those of us who view the democratic state's role as fundamentally tied up with a duty to respond to vulnerabilities and relieve demeaning conditions, may in fact have special reason to reject permitting the government to have this kind of moral authority over us. Perhaps in the end, responding to speech-mediated harms requires that we reconceptualise what it means to serve as a citizen under a liberal, democratic government. At the very least, it is likely to impose additional, special duties on the state to ensure they engage in speech control in an appropriate manner.

## §V. Clarifying persuasion

---

<sup>65</sup> <http://justice-everywhere.org/general/with-friends-like-these/>

<sup>66</sup> Mackenzie, 'The importance of relational autonomy and capabilities for an ethics of vulnerability'.

One key issue still in need of clarification, is what precisely we mean when we refer to ‘persuasion’. At the start of this thesis, ‘persuasion’ was characterised as an attempt to influence the beliefs or intentions of an audience, via presenting them with reasons. This is indeed central to the definition of persuasion, as we understand the term in this thesis, and helps us to understand what we refer to as the dual nature of persuasion.

What do we mean by persuasion’s dual nature? On the one hand, persuasion involves a speaker attempting to move the listener to change their mind. Thus Sara Rubinelli characterizes persuasion as “communication process whose purpose is to change the mental state of the persuadee”.<sup>67</sup> In this thesis we focus on the way persuasion is used to move listeners to change their beliefs and intentions – which is slightly broader than the way Rubinelli glosses the term ‘mental state’, but we agree with the core of her insight here. In other words, persuasion involves the speaker attempting to exert and influence on the listener, and ultimately move the listener to change the way they think. This is clearly a distinctive aspect of persuasion. What is so special about persuasion, though, is how, on the other hand, this change occurs in a way that simultaneously preserves the freedom and control of the listener. Persuasion, as Isabella Poggi puts it, is a non-coercive means of moving someone else to accept a goal.<sup>68</sup> When an agent is persuaded to change their beliefs or intentions, they are simultaneously moved by the persuader, while also themselves being the one who, in a significant sense, has authorised and authored this change. That is to say, in the paradigmatic case of persuasion, the speaker offers an argument to the effect that the listener should change their mind, and the listener chooses whether to accept this argument. It is this dual nature of persuasion – how the ‘persuadee’ is moved while remaining in autonomous control – that marks it out as a special concept. It is also this dual nature that distinguishes persuasion from coercive means of getting others to change their mind, such as through hypnosis or blackmail.

Moreover, both aspects of this dual nature are connected closely to our characterization of persuasion as involving the presentation of reasons. Successful persuasion involves the audience accepting reasons presented by the speaker, in favour of certain beliefs or intentions.

---

<sup>67</sup> Sara Rubinelli. (2013). “Argumentation as rational persuasion in doctor-patient communication”, *Philosophy & Rhetoric* 46 (4), pp.550-69.

<sup>68</sup> Isabella Poggi (2005). “The goals of persuasion”, *Pragmatics and Cognition* 13 (2), pp.297-336.

Such acceptance involves the exercise of the audience's rational agency, and as such is typically framed as a autonomous choosing of the subsequent beliefs or intentions.<sup>69</sup> Note that in some sense, this is a little misleading. It would be an exaggeration to suggest that agents have total control over their beliefs. We cannot, for instance, change our beliefs at will; if I currently believe in a Christian god, I cannot through the force of my will decide to stop believing in this god for the next hour, and then continue believing once again after the hour up. Nonetheless, following the Kantian tradition of identifying rational choice with freedom, philosophers have tended to assume that so long as we develop our beliefs as the result of a rational, intellectual process – say following being rationally persuaded by a friend – then those beliefs are held autonomously.

Similarly, the reason that persuasion exerts some kind of force on listeners, such that it might move them to change their beliefs or intentions, is because persuasion involves the speaker presenting the listener with reasons. The most obvious cases of persuasion involve a speaker explicitly spelling out reasons for their audience. A philosophy paper, for instance, might seek to persuade the reader that some theory should be rejected, by describing a major flaw in that theory. Similarly, my partner might seek to persuade me that we should move to a new area by presenting a list of benefits that would go along with our moving. These are cases of 'rational persuasion', and would seem to exert force on the listener to change their mind simply in light of spelling-out some potential rational error in their thinking.

Still, by no means is all persuasion so dry. Stories, rhetoric, and even poetry can plausibly present reasons to an audience too.<sup>70</sup> Indeed, these forms of expression may elicit emotional responses from their audience, but this does not necessarily mean that these forms of expression are not 'persuasive', or that they undermine the autonomous control of the listener. Eliciting an emotional response from a person, can be one in which we can present that person with a reason to alter their beliefs or intentions. Emotional responses may, for instance, acts as guides to what we really care about. They can also render especially clear to us,

---

<sup>69</sup> Michael Cholbi (2017). "Paternalism and our rational powers", *Mind* 126 (501), pp.123-53; George Tsai (2014), "Rational persuasion as paternalism", *Philosophy and Public Affairs* 42 (1), pp.78-112.

<sup>70</sup> Wordless music and pictures are a trickier case. It may be argued that these forms of expression can indeed have propositional content, and so may present reasons to their audience. We do not discuss this issue here, as it does not impact upon the success or failure of any of the main arguments analysed in this thesis.

the salience of some reason.<sup>71</sup> Consider, for instance, the case where a friend of mine convinces me to go to a restaurant, by vividly and colourfully describing just how excellent the last meal we had there was. In this case, the use of poetic language serves, not to undermine my use of reason, but to highlight certain considerations and remind me of just what is at stake in this case.

Relatedly, in everyday life, emotions certainly play an important role in whether a listener finds a piece of expression to be persuasive. For instance, whether hate speech is persuasive seems to ride not simply on whether the audience rationally accept the argumentative reasons being presented to them, but also whether the hate speech engages with their emotions in the ‘right’ kind of way. One emerging theme in the psychological literature on why people engage in and are persuaded by hate speech, is that hate speech reinforces group-identity and feelings of solidarity. The thought, roughly put, is that people forge an identity within a particular in-group, by distinguishing themselves from a particular out-group, and that hate speech is useful in this sense.<sup>72</sup> Put another way, to reinforce the sense of an ‘us’, people identify and demonise a ‘them’. Or as Cormac McCarthy’s Judge Holden puts it: ‘What joins men together...is not the sharing of bread but the sharing of enemies’.<sup>73</sup> Thomas Szanto, for instance, finds that engaging in hate speech creates a sense of togetherness with fellow haters, and entrenches a ‘shared habitus’.<sup>74</sup> Agneta Fischer et al agree, describing hate as reassuring and self-protective. They also add that hate is most likely to be effectively persuade its audience, when it persuades the audience that the target of hate is dangerous, and that they (the audience) are powerless.<sup>75</sup> Finally, the psychological research would seem to suggest that in

---

<sup>71</sup> For more discussion on the interconnectedness of reason and emotion, see: Miranda Fricker (1991). “Reason and emotion”, *Radical Philosophy Archive*, issue 057; Nadine Jung, Christian Wranke, Kai Hamburger, Markus Knauff (2014). “How emotions affect logical reasoning: evidence from experiments with mood-manipulated participants, spider phobics, and people with exam anxiety”, *Frontiers in Psychology* 5, pp.570-81. For more discussion on the intuitive pull of allowing that certain emotions can autonomously guide our decision-making, see Christine Tappolet (2006). “Autonomy and the emotions”, *European Journal of Analytic Philosophy* 2 (2), pp.45-59. For more on the connection between emotional reasoning and agency, see: Christine Tappolet (2006). *Emotions, Value, and Agency* (Oxford: Oxford University Press).

<sup>72</sup> Michael Peters (2020). “Limiting the capacity for hate: hate speech, hate groups and the philosophy of hate”, *Educational Philosophy and Theory* (forthcoming).

<sup>73</sup> Cormac McCarthy (1992). *Blood Meridian, or, The Evening Redness in the West* (Vintage International: New York).

<sup>74</sup> Thomas Szanto (2018). “In hate we trust: the collectivization and habitualization of hatred”, *Phenomenology and the Cognitive Sciences* 19, pp.453-80.

<sup>75</sup> Agneta Fischer, Eran Halperin, Daphna Canetti, and Alba Jasini (2018). “Why we hate”, *Emotion Review* 10 (4), pp.309-320.

terms of whether an audience is persuaded by an act of hate speech, much rides on whether the hate speech involves a strong narrative. On this topic, the research of Elissa Lee and Laura Leets is fascinating. They draw a distinction between ‘high narrative’ hate speech that involves plots, events and main characters, and ‘low narrative’ hate speech that does not involve these features, and does not link actions and events together. Their research suggests that, generally speaking, adolescent individuals find high narrative hate speech more persuasive than low narrative”. Interestingly, however, Lee and Leets also found that while low-narrative, explicit hate messages were more less persuasive, the effect of this speech on those persuaded by it was more long-lasting. The authors hypothesise that this is because while low narrative hate speech is more likely to be evaluated less favorably initially, those who do accept this speech are more likely to do so as a result of careful processing and issue-relevant thinking – and once a message is accepted in this way, it tends to stick.<sup>76</sup>

Persuasion, then, is a diverse category. All the same, it should be clear that not all harms that follow from speech, count as persuasion-mediated harms. As Strauss notes not every inducement through speech counts as persuasion.<sup>77</sup> For instance, threats and blackmail do not typically count as acts of persuasion. Granted, some might suggest that threats and blackmail must involve some degree of persuasion – after all, doesn’t the speaker have to persuade his audience to take their threat/blackmail seriously? Doesn’t the blackmailer, for instance, have to persuade his audience to believe that he will be hurt unless he complies with the demands being made? There may be some truth to this. All the same, once the audience has been persuaded to take the blackmail/threat seriously, what moves them to action is not the persuasive force of the blackmail or threat, but rather the coercive force of wishing to avoid a harm. Or, to put it another way, once taken as credible, blackmail and threats do not provide independent reasons for action. This is what separates out threats and blackmail from acts of persuasion.

Moreover, the thesis sides with Strauss in assuming that highly deceptive forms of speech should not count as genuinely persuasive, and hence government restrictions on these forms of speech should not count as persuasion-control. After all, we said earlier that persuasive speech

---

<sup>76</sup> Elissa Lee and Laura Leets is fascinating (2002). “Persuasive storytelling by hate groups online examining its effects on adolescents”, *American Behavioral Scientist* 45, pp.927-57.

<sup>77</sup> Strauss, ‘Persuasion, autonomy, and freedom of expression’, p.334.

moves listeners via engaging their deliberative faculties, providing them with a chance to rationally and emotionally reflect on what has been said. Indeed, this turns out to be a crucial part of why the likes of Strauss, Scanlon, Nagel, Weinstein, and Dworkin, think that such speech should be protected; for each of them, in different ways, argues that agents should be free to *autonomously* engage with speech and decide for themselves what to think, with governmental failure to bestow this freedom on its citizens variously described as e.g. a failure of respect, a failure of recognition, insulting, unfair, etc. The general argument here, roughly put, is that while the harmfulness of such speech establishes a strong case for regulation, reflecting on the way such speech engages (and indeed, only harms as a result of engaging) the autonomous capacities of agents, establishes a powerful case for permitting such speech – a case that is crucially strengthened once we reflect, also, on our democratic ideals. What is relevant for our purposes here, is that speech acts that move their listeners via deceiving them are probably best characterised as *subverting* the listener’s deliberative faculties. Because of this, the arguments canvassed in this thesis only provide partial, if any, support for deceptive speech. As such, we are minded in our discussions to set aside such cases, and focus only on those cases – persuasion-mediated harms – where the arguments in this thesis function best. Of course, if it turns out that any of the arguments in this thesis in fact does establish a case for tolerating harmful persuasion, we can then fruitfully ask whether this argument also establishes a case for tolerating harmful deceptive speech.

#### §VI. A taxonomy of persuasion-mediated harms

As we have seen then, this thesis is especially concerned with persuasion-control and cases of persuasion-mediated harm. With this in mind, let us sketch some of the different kinds of harm that persuasion can lead to, and the different ways in which it can lead to these harms. In particular four distinctions are worth highlighting. The point of this taxonomy is to get clearer on what we mean when we talk about harmful persuasion, and to clarify the focus of this thesis.

a. *beliefs and intentions to act*<sup>78</sup>

The thesis follows Scanlon by taking ‘harmfully persuasive speech’ to denote persuasive speech has the power to move listeners to develop harmful beliefs and harmful intentions.<sup>79</sup> Talk of ‘harmful beliefs’ is liable to attract some controversy. After all, some might ask how a mere belief or attitude could possibly be harmful? This issue is touched upon throughout the thesis, and especially in chapter 5. For now let us note that there are a number of different beliefs that intuitively seem to have the power to substantially set back the interests of others. For instance, it is surely harmful to belong to a society where one is exposed to (unjustified) widespread hatred and antipathy from others. That is to say, that among the other harms that such hatred may lead to, we should also consider the hatred itself to be harmful. Building upon this idea, Steven Heyman suggests that hate speech might be harmful insofar as it spreads hatred against certain groups of persons, and in turn leads to their social standing being lowered.<sup>80</sup> The underlying thought here seems to be that we are the kind of creatures that benefit from living in a community where we enjoy at least a basic level of recognition and respect from our peers, and hence we are harmed when this kind of basic recognition is stripped from us. It is also intuitively plausible to suppose that our beliefs and attitudes can prove harmful to ourselves, as when we are self-hating, fail to recognise our own worth, find ourselves to be ugly and repulsive, torture ourselves excessively over our mistakes, etc. In so far as speech can help promote such self-harming beliefs, we can reasonably talk about that speech being harmful.<sup>81</sup>

---

<sup>78</sup> Note that the disjuncts in this distinction are not exclusive. That is to say, it is possible for A to persuade B to develop both harmful beliefs and intentions. This point about non-exclusivity also applies to the other distinctions made below.

<sup>79</sup> Scanlon, ‘A theory of freedom of expression’. We say ‘inspired by’, because Scanlon in fact only discusses *false beliefs*, as opposed to beliefs more generally, as being harmful. Furthermore, Scanlon’s discussion seems to imply that he is referring to beliefs that are harmful *because they are false and might be mistaken as true*. By contrast, we wish to talk more generally about harmful beliefs.

<sup>80</sup> Steven Heyman, ‘Hate speech, public discourse, and the first amendment’.

<sup>81</sup> See Mackenzie and Stoljar, ‘Introduction: autonomy refigured’; Mackenzie, ‘Relational autonomy, normative authority and perfectionism’; Mackenzie, ‘The importance of relational autonomy and capabilities for an ethics of vulnerability’; Matsuda ‘Public response to racist speech: considering the victim’s story’; Martha Nussbaum (2000). *Women and Human Development: The Capabilities Approach* (Cambridge: Cambridge University Press); Martha Nussbaum (2003). “Capabilities as fundamental entitlements: Sen and social justice”, *Feminist Economics* 9 (2-3), pp.33-59; Frederick Schauer (1993). “The phenomenology of speech and harm”, *Ethics* 103 (4), pp.635-53.

Far less controversial is the idea that expression can lead listeners to go and act in harmful ways. Speakers may use their words to inspire listeners to commit acts of violence, harassment, discrimination, polluting, breaching of the peace, and indeed any other kind of harmful act one cares to imagine. Indeed, it is an acknowledgement of this fact that has led, to our knowledge, every developed legal system on earth to in some way regulate ‘incitement’.<sup>82</sup>

Of course, oftentimes what happens is that someone is inspired to develop a certain belief, which in turn primes them to act in a harmful way. How else could one inspire another person to commit acts of violence, other than through convincing them first to develop beliefs such as ‘this act of violence is justified’ and ‘this act of violence is necessary’? Likewise, acts of discrimination and harassment are often underpinned by e.g. discriminatory attitudes about the value of the victims involved. Even more subtly, feminist authors have long argued that certain beliefs about women lead to the testimony of women being ignored or misunderstood.

Why focus so narrowly on cases where speech persuasively inspires harmful beliefs and actions? Why does the thesis focus on whether we should tolerate cases of e.g. hate speech, pornography, anti-science speech that might inspire listeners to think and behave in harmful ways, when there are other kinds of harmful speech that we might focus on? For instance, why not also consider whether we should tolerate, in the name of democracy, blackmail, threats, undue influence, manipulation, or deceptive speech? The answer is that the democratic arguments in favor of free speech canvassed in this thesis, function best when used to defend (harmfully) persuasive speech. These arguments, as discussed earlier, revolve around the idea that citizens should be free to think for themselves and to autonomously decide what to think – and that failure to permit agents this freedom, amounts to e.g. a failure to properly respect them, recognize their autonomous nature, unfair treatment, etc. The key idea underpinning these arguments, is that while persuasive speech may well be harmful, there are compelling reasons to tolerate such speech, that relate to the way persuasion engages our rational, autonomous powers. As such, and in contradistinction, these arguments function rather less well, and indeed potentially fail to function at all, when used to defend harmful speech that does not engage the rational, autonomous powers of agents.

---

<sup>82</sup> Jaconelli, ‘Incitement: a study in language crime’.

*b. individually harmful v cumulatively harmful*

Another distinction we can draw relates to the nature of the harms that persuasive speech can lead to. In particular we can draw a distinction between beliefs and actions that are individually harmful, as opposed to beliefs and actions that are merely cumulatively harmful. We've already seen many examples of the former; threats, blackmail, harassment, for instance, are all harmful by themselves. When persuasive appeal is likely to lead to these kinds of harms, there is a clear, *prima facie* case for engaging in censorship.

Still, sometimes matters are not so simple. Some beliefs and actions are such that they are merely 'cumulatively harmful'. That is to say, they are only harmful when taken in conjunction with other similar beliefs and actions. A classic example might be that of polluting activities that, while harmless when taken in isolation, are harmful when considered altogether. Another case might involve the harm of living in a culture where one is hated because of one's race/religion/ethnicity/etc. If only one person holds such a hateful attitude towards you, then likely this falls below the threshold of being harmful. Still, if a significant number of persons in society hold such hateful views towards you and others like you, then plausibly this does count as harmful. After all, it is no small matter to live in a community where a significant number of people hate you for no good reason.

Perhaps best identified as a subset of cumulatively harmful speech, is speech that generally contributes to a harmful culture. Call this contextually harmful speech. A.W. Eaton, for instance, presents what she calls a probabilistic account of the harm of pornography, wherein exposure to pornography is neither necessary nor sufficient for its putative injuries, but rather raises the chances of harm depending on context.<sup>83</sup> In some cases, hate speech may function as contextually harmful, without necessarily harming specifiable individuals, through contributing to the creation or sustenance of a social climate in which harms and disadvantages redound to members of vulnerable social groups.<sup>84</sup> Here the idea is that speech may be contextually harmful, or at least wrongful, not only by generally raising the chances of harm taking place, but also by contributing to a larger harmful culture.

---

<sup>83</sup> A.W. Eaton (2007). "A sensible antiporn feminism", *Ethics* 117 (4), pp. 674-715; Amanda Cawston (2019). "The feminist case against pornography: a review and re-evaluation", *Inquiry* 62 (6), pp.624-58.

<sup>84</sup> Simpson, 'Dignity, harm, and hate speech'.

### *c. self-harm and harming others*

It is helpful to draw a distinction between cases where A persuades B to develop beliefs and intentions to act that stand to harm others, and cases where A persuades B to develop beliefs and intentions that stand to harm B. Put simply, we are here talking about persuading someone to harm themselves, versus persuading them to harm another person. Hopefully it has already been made clear that persuasion can lead to both kinds of harms. For instance, persuasion might lead to the former kind of harm when the speaker is inciting the listener to murder, harass, discriminate, breach the peace, indulge in polluting activities, etc. While persuasion may lead to the latter kind of harm when it involves speakers encouraging listeners to think about themselves in damaging ways e.g. to question their own self-worth, to hold self-hating views, to fail to appreciate their full potential, to torture themselves excessively over mistakes, to be ashamed about their sexuality, etc. In the main, this thesis will focus on cases where speakers try to encourage listeners to develop beliefs and intentions that are harmful to others. These cases have been the most salient in law and in philosophical debates surrounding free speech. Focusing on these cases also has the advantage of allowing us to set aside some of the complexities that go along with discussing the issue of paternalism.

Taken together, we can see that these three distinctions relate to the nature of the harms that expression, and especially persuasion, can lead to. Below three more distinctions are introduced. This time, however, the distinctions relate to nature of the persuasion in question, and how it is connected with these harms.

### *d. imminent v non-imminent harm*

Another important distinction we can draw is between speech that leads to imminent harm, versus speech that leads to non-imminent harms. Here the distinction is drawn in terms of temporal distance between the expression in question, and the resulting harm. Generally speaking, it is assumed by commentators that there is a stronger case, all other things equal, for interfering with speech that threatens imminent harms, compared with non-imminent harms. Indeed, it is commonplace for even the staunchest sceptic of censorship to permit the censorship of speech that threatens imminent harm. *Brandenburg*, for instance, established a precedent in U.S. law for interfering with speech that is likely to lead to imminent lawless behaviour, while no such precedent exists for interfering with non-imminent lawless

behaviour.<sup>85</sup> Broadly speaking the thesis follows this convention insofar as it focuses on whether there are compelling democratic reasons to abstain from censorship in the case of non-imminent harms.

Some might find this distinction puzzling. Why hold that it is more permissible to interfere with speech that will lead to harm today, compared to speech that will lead to harm next month? What difference could the temporal distance of the harm from the speech in question, possibly make? Jeffrey Howard provides the most plausible answer.<sup>86</sup> He points out that one of the key questions we must ask when considering whether to engage in the censorship of harmful materials, is whether there are alternative, less coercive measures that we can take to fight against these potential harms. With this in mind Howard suggests that, all other things equal, speech that threatens imminent harm leaves the government with fewer alternative measures with which to alleviate the potential harm in question. After all, speech that leads to harms further down the line may be responded to by the government in a number of ways. There is the chance for the government to engage in counter-speech, education, to change the mind of the would-be harmer, to provide protection for the potential victim(s), and so on. By contrast, these measures are typically unavailable when dealing with speech that threatens to cause harm right here and now. None of this is to imply that we should not regulate speech that is more remotely connected to the harms that follow from it. As we have seen, non-imminent speech harms may be serious. And even if regulation is not the only tool for protecting citizens in these cases, it may be one of the most effective.

## §VII. Some preliminary challenges to the thesis

Even at this early stage in our investigation, a reader may have some preliminary worries. This chapter closes by considering three particularly interesting challenges that a reader might have regarding this thesis, and the deontological approach to problematising censorship.

*a. is persuasive speech really harmful?*

---

<sup>85</sup> *Brandenburg v. Ohio* (1969), 395 U.S. 444.

<sup>86</sup> Howard, 'Dangerous Speech'.

One worry a reader might have even at this early stage, concerns whether speech is really harmful in the way that this thesis has thus far presupposed. Granted, certain kinds of speech seem directly harmful in a way that is incontestable. Blackmail, threats, harassment, and intimidation, for instance, likely all fall within this category. But other kinds of speech are alleged to be harmful in a more indirect manner. One might especially be sceptical regarding the possibility of so called ‘persuasion-mediated harms’. What evidence, they might ask, is there in favour of widespread persuasion-mediated harms? What evidence is there for the idea that hate speech, climate change denialism, and calls for ignoring the advice of public health experts, are in fact dangerous? What evidence is there to indicate that these forms of speech might in fact be effective in persuading listeners and leading to an increased exposure to risk? These are complicated empirical questions. Rather than assess the impressive literature on these issues, this thesis instead focuses on principled, deontological arguments in favour of permitting dangerous persuasive speech.<sup>87</sup> If it turns out that persuasion doesn’t cause harm, then the arguments for free speech presented in this thesis will stay the same, and perhaps even strike us as more powerful. But if investigation does indicate that sometimes persuasion causes serious harms, then this simply emphasises the importance of attending carefully to the deontological arguments canvassed in this thesis. After all, these arguments purport to establish a strong case for tolerating even dangerous speech. They suppose that, as with fair elections and just criminal trials, our deepest values tell in favour of supporting free speech even when the chips are down.<sup>88</sup> This thesis investigates whether this kind of robust freedom of speech advocacy stands up to critical scrutiny.

In addition, it is worth reflecting on how a number of common-sense observations support the idea that we should at least be open-minded about the possibility of speech being capable of causing harms – and hence why such deontological arguments in favour of free speech may strike us as important. For instance, if persuasive speech possessed no power to influence the thoughts and actions of others, it is doubtful that we would have evolved to develop such complicated linguistic abilities. Moreover, given the various different forms of

---

<sup>87</sup> Brown and Sinclair ‘*The Politics of Hate Speech Laws*’; Gelber and McNamara, ‘Evidencing the harms of hate speech’; Eric Heinze (2016). *Hate Speech and Democratic Citizenship* (Oxford: Oxford University Press); Hugo Mercier (2017). “How gullible are we? A review of the evidence from psychology and social science”, *Review of General Psychology* 21 (2), pp.103-22; Scott Straus (2007). “What is the relationship between hate radio and violence? Rethinking Rwanda’s “radio machete””, *Politics & Society* 35 (4), pp.609-63.

<sup>88</sup> Dworkin, ‘Is there a right to pornography?’.

speech – spoken word, music, clothing, art, the lessons our parents teach us, the jokes our friends make – and the cultural significance of some of these forms of speech, it would seem almost remarkable if speech in no way influenced our thoughts and behaviour. Finally, one might note, as indeed several defenders of freedom of speech themselves have noted, that it is precisely because speech has the power to influence others, that speech is special and *prima facie* worth protecting. In short, this is a thesis that takes speech seriously – both the good and the bad.

*b. the responsibility challenge*

A different kind of challenge concerns how we ascribe responsibility to speakers that encourage others to think and act in harmful ways. A very simple version of this challenge holds that when A convinces B to kill C, it is B and not A who is subsequently responsible for C's death. On this simple version, the thesis is flawed insofar as it holds that speakers should be held responsible or punished for the consequences of their speech.

This simple version of the challenge fails because it overlooks the possibility that A and B can share responsibility for C's death. Some might worry that sharing the responsibility for C's death between A and B, ends up diluting the responsibility of A and B. But even this worry is misguided. Responsibility is not a finite resource. There is no need to hold B any less responsible or culpable or punishable for his role in C's death, just because A also played a role. B's criminal sentence, for instance, need not be reduced upon learning of A's role in the crime.

A more plausible version of the responsibility worry holds that, by and large, even when speakers express themselves in a way that indirectly leads listeners to cause harm, these speakers tend to have had a relatively marginal effect on their listeners. In particular, the speaker's influence is dwarfed by the wider cultural and historical influences on the listener that caused them to act/think in harmful ways. As such, so this worry goes, it would be wrong to single out speakers as responsible for the harms in question, or to hold them as appropriate targets of censorship.<sup>89</sup>

---

<sup>89</sup> Robert Mark Simpson touches upon something like this worry in: Robert Mark Simpson, 'Dignity, harm, and hate speech'. However, Simpson's worry is part of a more nuanced critique of Stephen Heyman's (Steven Heyman, 'Hate speech, public discourse, and the first amendment') work in particular.

In response, let us first note that this worry is distinct from the worry that speech does not in fact have harmful consequences. That is to say, the worry before us admits that speakers may have a significant causal role in the production of harmful beliefs and actions, but nonetheless holds that we would be mistaken to hold those speakers as particularly responsible. With this clarification in mind, I'm not sure the worry is particularly compelling. Granted, speakers are not wholly responsible for the harms that they promote. But there may still be good reason to hold them responsible, or at least apt targets for censorship. For instance, perhaps the government has good reason to think that censorship will prevent these speakers from inciting harm. Or perhaps the government believes that punishing these speakers will discourage others from committing harms. Or perhaps the government thinks that censoring a particular speaker, coupled with other cases of censorship and a larger educative scheme, will help promote more tolerant attitudes in the next generation of thinkers. The wider lesson seems to be that so long as one has good reason to suppose that interfering with these citizens in particular will be an effective means to addressing significant harms, then this seems to justify singling these speakers out in particular.

Moreover, there may be something attractive about asking that citizens take responsibility for what they say and the harms that they contribute to. Alexander Brown and Adriana Sinclair, for instance, talk about an 'ethic of responsibility' that should guide political figures to be duly mindful about what they say and the harms their words can cause.<sup>90</sup> This thesis sides with the suggestion that political figures have special reason to abide by an ethic of responsibility, because of the power and privilege that their position confers upon them. Still, plausibly all speakers with a reasonable amount of power and influence should similarly be guided by such an ethic of responsibility. Even more radically, it has become increasingly popular to suppose that all of us, as citizens, have certain special duties and obligations. Along these lines, why not hold that as citizens with an interest in participating in a functioning democracy, we each of us have various responsibilities and duties towards our fellow citizens, including the demand to be guided by an ethic of responsibility?<sup>91</sup>

---

<sup>90</sup> Chapter 7 of Brown and Sinclair, *The Politics of Hate Speech Laws*.

<sup>91</sup> For more on this, see: Will Kymlicka and Wayne Norman (1994). "The return of the citizen: a survey of recent work on citizenship theory", *Ethics* 104 (2), pp.352-81.

One final version of this responsibility worry focuses on the issue of stigma. On this version, the problem isn't so much that censorship singles out certain speakers as responsible in some way. It is that in reality censorship singles out certain speakers as criminally liable for government interference, and the stigma that goes along with this. This is especially problematic when the speakers in question are only marginally responsible for the harms that are associated with their speech. It is one thing, so the argument goes, to argue that people should be responsible and thoughtful when it comes to what they say, but it is another thing to hold that people should be stamped with the label of 'criminal', just because they have contributed to a wider milieu that moved someone in a way that is harmful.

There are broadly speaking two ways in which to respond to this worry; one grasps the nettle, and the other seeks to avoid it. The nettle-grasping option is to say that those that fail to be guided by an ethic of responsibility and express themselves in a way that foreseeably contributes to listeners developing harmful beliefs or intentions, are therefore responsible for those harms in a way that renders them justified targets for even stigmatising censorship, so long as this is an effective means to redressing significant harms. In other words, 'they had it coming'. If they didn't want to be subject to stigmatising censorship, then maybe they shouldn't have expressed themselves in such a harmful manner.

Still, there is another kind of response that is worth considering for the sake of those who remain unconvinced by the above. The nettle-avoiding response is to point out that the government can take a less stigmatising approach to censorship, whereby it views those it censors as 'innocent' agents that must nonetheless be interfered with in order to prevent harms. That is to say, so long as we are confident that censoring John will help prevent harm, we needn't worry ourselves overly with ascriptions of moral responsibility or culpability. We can, for all intents and purposes, consider John a neutral target of necessary harm-preventing government action. This removes the need to think of John in a stigmatising way.

Some will disagree with this. As H.L.A. Hart explained, while some might hold that criminal punishments are justified so long as they serve a purpose (e.g. deter crime, prevent harms, etc.), others will hold that punishments are only justified so long as they are inflicted upon persons who have committed a moral wrong.<sup>92</sup> In chapter 4 we will consider in more

---

<sup>92</sup> H.L.A. Hart (1959). "Prolegomenon to the principles of punishment", *Proceedings of the Aristotelian Society* 60, pp.1-26. A good example of the view that criminal punishments are justifiable only so long as they are imposed on

detail the ‘serves a purpose view’. For now, it will suffice to sketch the prima facie intuitive appeal of the ‘serves a purpose view’, and how a number of cases can be evoked to help support this intuition. Consider the following cases: the government requires that citizens wear a mask when entering a shop, in order to help fight against the spread of a deadly virus; the police receive a tip that there is a bomb about to go off in a football stadium, and so evacuate the stadium and search those in attendance; upon realising that a terrorist attack is imminent, the government shut off all electricity in an area; a terrorist organisation strap a bomb on to an innocent person, and the police must now, for the sake of this person’s safety and the safety of those in the surrounding area, forcibly hold down this person and defuse the bomb; as we have seen countless times in Hollywood movies, a police officer is chasing a baddie, and must commandeer an innocent citizen’s vehicle in order to continue the chase. All these cases involve innocent citizens who have caused no harm and committed no moral wrong, having their liberty curtailed by the state as part of a justified and proportionate harm-prevention strategy. If one wishes to insist that those that use their words in a way that exposes others to danger are not in fact committing a harm or guilty of a moral wrong, then we can simply view them as similar to the innocents that are interfered with in these cases.<sup>93</sup>

*c. is the ‘ethic of responsibility’ too demanding?*

The final challenge concerns the ‘ethic of responsibility’ touched upon just above, and whether it might be unreasonably demanding. The ethic asks that speakers be mindful of what they say and be held at least partly responsible for the harms that flow from their speech. But depending on how we interpret this statement, the ethic of responsibility might be thought to run into some problems. For instance, consider the case where John says ‘hi’ to his neighbour, and because the neighbour is mentally unstable, he becomes enraged by this friendly greeting and murders an innocent bystander. Does the ethic of responsibility demand that John be held responsible for these harms? Does it make a difference if John is aware that his neighbour might react to the greeting like this? Does it make a difference if John knows that his neighbour is mentally unstable?

---

those who have done wrong, is: Evan Simpson (2006). “Responsibilities for hateful speech”, *Legal Theory* 12 (2), pp.157-77.

<sup>93</sup> For more on this, see: Michael Otsuka (1994). “Killing the innocent in self-defense”, *Philosophy and Public Affairs* 23 (1), pp.74-94.

These are difficult questions, and unfortunately this thesis will not provide the kind of full account of responsibility necessary in order to arrive at fully satisfying answers. Instead, let us briefly suggest that when we ask that agents be held responsible for persuasion-mediated harms their words might cause, what we are really asking is that they (e.g.) make a reasonable effort to reflect upon the content of their words, reflect upon what kind of beliefs and intentions these words are designed to promote, and consider whether these beliefs and intentions are likely to prove harmful or dangerous.

Typically those that engage in harmful persuasion – hate speakers, climate change denialists, anti-vaccination conspiracists, those that preach the disobedience of public health advice – are well aware of the kind of beliefs and actions their speech might promote. The whole point of such speech is to promote e.g. hatred of minorities, less empathy for certain causes, discrimination, distrust of scientific advice, an increase in disobedience of scientific advice. The problem is that these speakers do not particularly care about the kind of harms or dangers these beliefs and actions can lead to, or have failed to properly consider the issue with enough care. Yet this does not exculpate the speaker or relieve them of responsibility for the harms in question, just as a racist murderer is not exculpated on the grounds that he considers his victims to have no moral value, or a con artist is not forgiven on the grounds that he did not care about those innocent people he was ripping off. Admittedly, sometimes speakers may be surprised about just how much damage their speech leads to. Maybe they meant to promote hatred, but not murder. Or perhaps they wanted the government to take a harder line against immigration, but did not intend for violence to break out in the streets. In these cases, we must make sensible judgements about what kind of consequences agents could reasonably be expected to foresee. Just as we might have to make a sensible judgement about whether a speeding driver who unintentionally kills an innocent person ought to have taken greater precaution, or whether a foreman at a construction site should have been more aware of the dangerous conditions he was forcing his workers to operate in, we can also make sensible judgements about when speech becomes dangerous in a way that speakers ought to be held responsible.

## Conclusion

This first chapter has sought to clarify the aims of the thesis, as well as defend some of its key underlying assumptions. Chapter 2 considers T.M. Scanlon's treatment of persuasion-control. The chapter seeks to unpack the Scanlonian contention that persuasion-control might involve a serious interference with the autonomous capacities of citizens, and because of this undermines the government's democratic legitimacy.

## Chapter 2

### Telling citizens what to think

T. M. Scanlon's 1972 essay "A theory of freedom of expression" looms large over this thesis.<sup>94</sup> Not only does Scanlon focus on persuasion-mediated harms, he explores the idea – central to this thesis – that the regulation of speech aimed at preventing these harms lies in conflict with our democratic values.<sup>95</sup> Moreover, this essay occupies an interesting position in the philosophical literature on freedom of speech. On the one hand, philosophers like Brison have offered persuasive critiques of the 1972 account, with Scanlon himself later going on to disavow key aspects of his original argument.<sup>96</sup> On the other hand, the account has had an

---

<sup>94</sup> Scanlon, 'A theory of freedom of expression'.

<sup>95</sup> Scanlon's primary influences in this area are Immanuel Kant and Alexander Meiklejohn. See: Immanuel Kant (1996) [1785]. *The Groundwork for the Metaphysics of Morals*, in *Practical Philosophy* (Cambridge: Cambridge University Press), ed. Mary Gregor; Alexander Meiklejohn, 'Free Speech and Its Relation to Self-Government'; Alexander Meiklejohn (1960). *Political Freedom* (New York: Harper and Brothers).

<sup>96</sup> Susan Brison's critique of Scanlon may be found in: Brison, 'The autonomy defense of free speech'. Scanlon disavows several aspects of the 1972 account in: T.M. Scanlon (1979). "Freedom of expression and categories of expression", *Pittsburgh Law Review* 40, pp. 519-50.

enduring impact on the freedom of speech literature, with many philosophers seeking to explore his claim that censorship can impair the political legitimacy of the government that engages in it – though most, if not all, of these accounts attempt to cash-out this connection in a way that departs from Scanlon’s own suggestions.<sup>97</sup>

This chapter reflects upon Scanlon’s treatment of persuasion-control. At its core, the chapter traces Scanlon’s exploration of whether persuasion-control amounts to a distinctive, especially objectionable form of government interference. Close attention is paid both to Scanlon’s discussion of autonomy, as well as his comments on democratic legitimacy. Ultimately it is argued that while Scanlon’s provides a thought-provoking account of the conflict between persuasion-control and our democratic values, his assumptions about the latter in particular are rather brute and stand in need of additional substantive argument.

The chapter proceeds as follows: §I offers a basic outline of Scanlon’s critique of persuasion-control. While this basic outline gives us a sense of Scanlon’s argument, it is concluded that several important details are still in need of clarification. With this in mind, §II asks whether we can skip all this talk of democracy and legitimacy, and simply construct a direct autonomy-based argument against persuasion-control. It is concluded that this strategy is unlikely to work, as our autonomous capacities are not undermined in any particularly special way by persuasion-control. §III explores a rather different way of framing the conflict between persuasion-control and autonomy. Developing some of Scanlon’s comments on the issue, it is argued that persuasion-control can be usefully framed as an attempt to tell citizens what they may and may not think (i.e. what kind of beliefs and intentions they may permissibly develop). §IV argues that, by itself, this way of framing persuasion-control still falls short of establishing a powerful, principled objection. The rest of the section dives deeper into Scanlon’s discussion of democratic legitimacy, to see whether this helps problematise persuasion-control. §V draws matters to a close by suggesting that, while this is a promising approach, and may in fact be defended from several sensible objections, this reading nonetheless rests on a brute conception of democratic legitimacy that stands in need of further supporting argumentation. Providing this argumentation will be the main task of chapters 3, 4, and 5.

---

<sup>97</sup> Some particularly important philosophical treatments of the freedom of speech issue that have been influenced by Scanlon’s thoughts on legitimacy, include: Baker, ‘*Human Liberty and Freedom of Speech*’; Dworkin, ‘*Is There a Right to Pornography?*’; Heinze, ‘*Hate Speech and Democratic Citizenship*’; Robert Post (2017). “*Legitimacy and hate speech*”, *Faculty Scholarship Series* 524, pp.651-9.

## §I. Scanlon's argument

Let us start with a basic sketch of the principled objection to persuasion-control, levelled by Scanlon. At the foundation of Scanlon's work on freedom of expression and censorship lies an account of government legitimacy. Drawing on the likes of Kant and Meiklejohn, Scanlon argues that governments may only treat their citizens in ways that leave those citizens capable of viewing themselves as autonomous beings.<sup>98</sup> This, so Scanlon argues, is a matter of "democratic political theory".<sup>99</sup> These claims about legitimacy will be discussed in more detail §4 and §5. But even at this early stage it is clear that the Scanlonian vision of a legitimate democratic state appears to be one in which governments must address themselves to citizens in a certain way. In particular they must treat citizens in ways that shows due deference towards citizens and their autonomous capacities. Here the word 'deference' is used quite precisely – for as we will see, Scanlon's argument may well require that the government, at least in some areas, quite literally defer to the judgement of their citizens.

Of course, much rides on what Scanlon precisely means when he uses the word 'autonomous', and what he sees as required in order for citizens to view themselves as autonomous. Scanlon argues that in order to correctly view one's self as autonomous, one must enjoy sovereign control over two crucial powers:

- i) the power to decide for one's self what to believe, and
- ii) the power to weigh up competing reasons for actions.<sup>100</sup>

---

<sup>98</sup> Scanlon, 'A theory of freedom of expression', p. 214.

<sup>99</sup> Scanlon, 'A theory of freedom of expression', p.226.

<sup>100</sup> Scanlon, 'A theory of freedom of expression' p. 215. The first power is simple enough. Citizens must be sovereign in forming beliefs, deciding what they believe is true, what they believe is false, etc. Scanlon doesn't seem to think this extends to being able to form beliefs that harm others, but only to beliefs that harm one's self. However, we are interested in an even more ambitious account that involves autonomous agents being able to develop any beliefs, including those that harm others. The second is a little trickier. It can't be that citizens must be sovereign in deciding how to act, although Scanlon does occasionally get very close to glossing the second aspect of autonomy this way. The reason this can't be right, is that Scanlon himself says that governments have the authority (within reason of course) to coerce their citizens, and that being subject to coercion is compatible with exercising one's second autonomous power. A better way then of glossing the second autonomous power, is as a matter of being sovereign in forming intentions. As we will see in more detail below, Scanlon thinks that the government can forcibly prevent John from committing a murder by restraining him. Yet they may not pursue the same end by controlling what information John is exposed to so that John never gets to form the intention to murder in the first place.

The key idea here is that autonomous agents enjoy a certain kind of independence and intellectual autonomy; they are free to make up their own minds about what beliefs are worth having and what actions are worth performing. Particularly when we are talking about action, this may be a rather circumscribed freedom. Scanlon does not think, for instance, that intellectually autonomous citizens ought to be free to execute any action they see fit – citizens should not be free to engage in murder and battery. But what they must be free to do, so the argument seems to go, is come to their own conclusions about what actions are worth performing in the first place, even if they are ultimately constrained from following through with those conclusions in certain cases.

In light of all of this, one might expect Scanlon to formulate an overarching thesis about the *value* of freedom of expression, or perhaps provide an account of all of the cases in which censorship is and isn't acceptable. Scanlon does neither. Instead he offers a relatively restrained *negative* thesis that posits that two specific cases of government censorship leave citizens unable to view themselves as intellectually autonomous, and hence must be “absolutely” and “exceptionlessly” prohibited.<sup>101</sup> Summing up his thesis in what he terms the *Millian Principle*, Scanlon rejects certain rationales for government censorship. In particular the government may not appeal to a desire to prevent either of the following two harms:

- i) *the harms to certain individuals which consist in their coming to have false beliefs as a result of those acts of expression,*
- ii) *the harmful consequences of acts performed as a result of those acts of expression, where the connection between the acts of expression and the subsequent harmful acts consists merely in the fact that the act of expression led the agents to believe (or increased their tendency to believe) these acts to be worth performing.*<sup>102</sup>

The first part of the principle is subtly reflexive in its wording. Upon close inspection, it merely

---

<sup>101</sup> Scanlon, 'A theory of freedom of expression', pp. 215, 224.

<sup>102</sup> Ibid, p. 213. The name Scanlon opts for – the 'Millian principle' – derives from how, according to Scanlon, the conclusions of the principle can be seen as an extension of the thesis Mill defends in Chapter II of *On Liberty*. Of course, while the conclusions Scanlon draws may overlap with Mill's views, the arguments he provides for these conclusions are quite distinct.

prohibits the government from engaging in censorship on the grounds that it is worried John will be persuaded to develop *self-harming* beliefs. However, it is unclear why Scanlon does not likewise prohibit censorship aimed at preventing John from developing beliefs that are harmful to others. At the very least, the justification offered by for this principle could be used to defend John's freedom to develop such other-harming beliefs. Matters become even more confusing when we note that the second part of the principle is not limited in this way. With this in mind, we suggest that the first part of the Millian Principle be modified so that it generally prohibits the government from interfering with persuasive speech that it fears will inspire harmful beliefs. This may involve something of a departure from what Scanlon intended. But by extending Scanlon's principle in this way, we are given a distinctively principled rejection of governmental interference with our beliefs, as well as protecting the coherence between the first and second part of the Millian Principle.

The second part of the principle prohibits the government from engaging in censorship on the grounds that it is worried that some piece of expression will persuade John to act in a harmful way. This time the principle is referring to both self-harming and other-harming conduct. In essence the principle requires that citizens are free to engage one another in persuasive appeals, regardless of the dangerous intentions or actions that these appeals might lead to. Putting these two parts together, the very slightly modified Millian Principle we are considering prohibits the government from engaging in censorship, whenever the underlying reason for that censorship is a worry about harms that flow from citizens being persuaded.<sup>103</sup> Or even more simply: the principle holds that citizens must be free to engage in harmful persuasion.<sup>104</sup> This modified principle echoes David Strauss' 'persuasion principle', summed up

---

<sup>103</sup> Against Scanlon, many believe that at least some particularly harmful acts of advocacy should be illegal. Acts of terrorism and mass murder may be so terrible as justify censoring even attempts to persuade others to perform these actions. Indeed, at least in England, the 2007 Serious Crimes Act qualifies such acts as criminal offenses. In Scanlon's favour however, there are other cases of harmful advocacy where *prima facie* it seems right to think ought to be exempt from censorship. Thus, consider Scanlon's case of a decrease in the level of personal safety that is brought about as a result of a piece of political propaganda, or perhaps a religious tract, that inspires great social upheaval and rebellion (see pp.211-212). As Scanlon rightly says, there is something to be said for the idea that citizens should be free to make up their own minds about how to act in this case, even if this has bloody consequences. Still, rather than settle this dispute at the level of intuition, this chapter rather analyses the arguments provided by Scanlon in favour of his view.

<sup>104</sup> In this sense, Scanlon provides what we might call a 'listener-based' account of free speech, in that he is arguing that the democratic legitimacy of the government depends on it persevering the freedom of citizens to listen to various ideas and arguments.

in his conclusion that “the government may not restrict speech because it fears, however justifiably, that the speech will persuade those who hear it to do something of which the government disapproves”.<sup>105</sup> Throughout this thesis we refer to cases of governmental interference that violate the modified Millian Principle and Strauss’ persuasion principle, simply as *persuasion-control*.

When the government engages in harmful persuasion, Scanlon supposes that they impair, if not decisively undermine, their own democratic legitimacy. The details of this democratic vision will be explored in more detail soon, but even now several important features can be highlighted. At the centre of Scanlon’s democratic vision, lies an ideal of citizens as empowered and self-directing. In particular the view picks out our intellectual autonomy – the processes by which we come to decide what to believe and how best to act – as in some sense beyond the reach of a democratic government. This isn’t quite the same as claiming that the government can never engage in persuasion-control. At one point Scanlon suggests that in certain emergency scenarios, the government may justifiably limit the expression of arguments that it deems too dangerously persuasive.<sup>106</sup> Perhaps, for instance, violent civil war is on the cusp of breaking out, and widespread censorship is the only way of dampening the flames of revolution. Here a kind of Hobbesian-security rationale or utilitarian reasoning may well justify the government’s interference, thinks Scanlon. But there’s a sting in the tail. In such a case the government’s authority undergoes an important shift, with the government now operating as a kind of necessary authoritarian entity, rather than a properly democratic government.

Several important ambiguities still remain: precisely how does persuasion-control leave us unable to view ourselves as autonomous? Does persuasion-control constitute a distinctive threat to our intellectual autonomy? Is intellectual autonomy really so important to our vision of how the democratic state ought to operate? To help shed light on these questions, §II-IV each explore different ways of interpreting the Scanlonian argument here. §II kicks things off by looking at whether we can simply leave aside talk of democracy and government legitimacy, and

---

<sup>105</sup> p.334 of Strauss, ‘Persuasion, autonomy, and freedom of expression’.

<sup>106</sup> p.225. Scanlon at least considers that this might be potentially permissible – although he expresses reluctance to allow even this much.

instead provide a direct autonomy-based critique of persuasion-control based on Scanlon's argument.

## §II. Autonomy

In this section we explore whether it is possible to leave aside talk of democracy and legitimacy, and simply construct a direct autonomy-based argument against persuasion-control. That is to say, can we construct a direct autonomy-based argument that establishes a principled and powerful objection against persuasion-control? In the end, we suggest that the answer is 'no'.

The simplest way to cash out the idea that persuasion-control undermines our intellectual autonomy, is as a claim about the way persuasion-control damages our intellectual capacities. For Scanlon this would involve reflecting upon the way persuasion-control impairs our capacity for deliberating on what to believe and how to act. However, this turns out to be a tricky claim to parse. For starters, the claim cannot be that persuasion-control necessarily damages the intellectual capacities themselves. After all, the government may selectively silence certain dangerous ideas from being expressed, without wreaking any long-term damage on our autonomous capacities. There's no reasons why, for instance, banning incitement to violence should necessarily impair the bare capacity of citizens to rationally assess arguments and ideas. Indeed, in a certain sense persuasion-control leaves our intellectual capacities untouched. For rather than directly damaging our intellectual capacities in any way, persuasion-control simply limits our ability to use those capacities to engage with certain arguments and ideas. In other words, it is the materials towards which we direct our intellectual capacities, rather than those capacities themselves, that the government is interfering with.

Might this kind of interference be considered objectionable? That is to say, might we consider our autonomy interfered with by the government's decision to curtail our access to (intellectually engage with) certain materials? One way of cashing-out this claim, is to frame this interference as a form of mental coercion. Nagel, for instance, characterises certain censorious interferences as "thought control".<sup>107</sup> David Strauss uses the terminology of mind-control, and describes censorious interferences as manipulative<sup>108</sup>. In a similar vein we might say that the

---

<sup>107</sup> p.97 of Nagel, *Personal rights and public space*'.

<sup>108</sup> Strauss, 'Persuasion, autonomy, and freedom of expression'.

way persuasion-control prevents citizens from developing certain kinds of ideas and intentions, through its suppression of supporting arguments. Just as say hypnosis or conditioning would be unacceptable methods of the government securing compliance with its laws and discouraging harmful behaviour, so the argument goes, so too persuasion-control should be considered beyond the pale.

The problem with this reading is that it exaggerates the potency of persuasion-control. Even if persuasion-control made it impossible to access any arguments in support of some belief or intentions – X – it nonetheless typically falls well short of preventing citizens from developing the belief or intention, X, for themselves. For instance, citizen will still be free to think up arguments in favour X all by themselves. Indeed, coming to affirm X based on such self-generated arguments may well count as an especially authentic use of one's autonomous capacities. Moreover, speech regulation does not tend to be anywhere near this sweeping. For instance, the government might decide to limit attempts to encourage others to join terrorist causes, judging these attempts to be the most dangerous forms of encouragement of terrorist acts, while still permitting a whole range of other arguments that attempt to understand, analyse, and even encourage empathy towards, the terrorist political causes in question.

These suggestions speak to the idea that the censorship of arguments in favour of X, still leaves some room for citizens to consider and deliberate upon arguments in favour of X, as well as come to affirm X. In some ways these suggestions speak to the power of our imaginative and intellectual capacities; even while living under a censorious government, citizens still retain the ability to reflect on the world around them, to formulate their own arguments, to be inspired by other topics and analogous debates that are permitted to be discussed, etc. None of this is to say that persuasion-control has no impact on our intellectual capacities; if it had no kind of impact upon our thinking, then advocates of persuasion-control would lose their main argument for supporting such regulation in the first place. Indeed, it isn't even to suggest that persuasion-control can never have some degree of coercive influence on citizens, or that it might constitute a worrying kind of influence. But framing persuasion-control as a form of mind-control roughly analogous to hypnotism or conditioning, strikes us as an unhelpful way of unpacking the threat to our autonomy posed by persuasion-control.

Here's a more subtle approach: we locate the fault of persuasion-control in the way that it constitutes a significant interference with our autonomous capacities, insofar as it makes it

harder for us to consider certain arguments, and by extension to engage with those arguments using our autonomous capacities. Let us be clear here, persuasion-control does not make it impossible to consider certain arguments in favour of X. Rather, the worry is that persuasion-control objectionably raises barriers to my accessing certain arguments.<sup>109</sup> An analogy will help clarify matters. In the case where Jack blackmails Jill to act a certain way, Jack's actions are objectionable even if they do not literally coerce Jill to act a certain way or annihilate the possibility of her acting otherwise. That is, the mere illegitimate raising of significant barriers is enough to render Jack's actions highly objectionable. In a similar vein, the government wrongs us when it raises significant barriers to our engaging with certain arguments and ideas, even if it does not technically make it impossible for us to engage with these arguments.

This way of understanding the worry with persuasion-control enjoys a certain *prima facie* plausibility.<sup>110</sup> Still, it doesn't quite fit the bill. First, as has been shown most notably by Brison, but also the likes of Paul Gomberg, and Easton, this understanding of Scanlon's account may in fact open the door to censorship, rather than simply closing it as Scanlon intends.<sup>111</sup> After all, limiting our access to certain arguments may result in a net-gain in the number and variety of arguments that are readily presented in the marketplace of ideas, just as removing certain weeds may make room for a more varied and healthy garden. Take the suggestion that restrictions on hate speech help secure a more civil, welcoming environment that encourages widespread participation in political discussion. Alexander Brown, for instance, refers to how targets of hate speech might plausibly feel less inclined to participate in public discourse, having been insulted, treated without dignity, and in extreme cases threatened.<sup>112</sup> In a slightly different vein, Andrew Reid suggests that hate speech can lead to certain political voices being

---

<sup>109</sup> Thanks to Kalle Grill for suggesting this reading.

<sup>110</sup> Jonathan Seglow (2016). "Hate speech, dignity and self-respect", *Ethical Theory and Moral Practice* 19 (5), pp. 1103-16. Seglow at least hints at this reading when he writes that Scanlon's objection to censorship rests on the idea that that "autonomous individuals have an interest in not having the free circulation of ideas *interfered* with by the state" (p.110).

<sup>111</sup> Susan Brison 'The autonomy defence of free speech'; Easton, 'Autonomy and the free speech principle'; Paul Gomberg (1994). "Autonomy and free expression", *Journal of Social Philosophy* 25 (2), pp. 97-104. See also Chapter 3 of Alexander Brown, 'Hate Speech Law: A Philosophical Examination'.

<sup>112</sup> See, for instance: 'Brown, 'Hate Speech Law: A Philosophical Examination', p.200; Alexander Brown (2017). "Hate speech laws, legitimacy, and precaution: a reply to James Weinstein", *Constitutional Commentary* 32 (3), pp.599-617, p.612

systematically disregarded by their audience. On top of these political voices being essentially denied a (reasonable) audience, this also discourages future political expression.<sup>113</sup>

Moreover, in addition to possibly increasing the number of arguments on offer for people to engage with, speech restrictions might also encourage better, more rational engagement with those arguments. Leiter points out that courts routinely limit what kind of argumentative appeals can be made in an attempt to encourage more rational, more informed, less biased deliberation from jurors.<sup>114</sup> Hearsay, for instance, might well be persuasive. But hearsay is sometimes prevented from being presented in a court, for fear that it might result in a misinformed jury. Perhaps certain persuasive but misleading appeals could be similarly censored, in a way that promotes our rational engagement with the issue – unsubstantiated and dangerous conspiracy theories spring to mind. In a similar vein, pre-trial motions often involve deciding whether a piece of ‘scientific evidence’ is in fact admissible in court. When the scientific evidence in question is judged to be in fact bad science, Leiter tells us, it may be dismissed on the grounds that such evidence would not in fact serve the court’s aims. Again, an analogous case can be made here for the autonomy-promoting effects of censoring junk science arguments and unfounded conspiracy theories. Leiter also points out that laypersons often find it difficult to assess whether the arguments being presented to them, are being delivered by genuine epistemic authorities. Perhaps restrictions on the testimony of non-experts, or at least restrictions on the way the testimony of such non-experts may be framed, might promote more rational deliberation on the arguments being presented.

All of this is to say that persuasion-control might in fact protect and promote the use of our autonomous capacities. For even if persuasion-control involves some interference with the range of arguments available for autonomous agents to assess, it may still result in a net benefit in the total numbers of arguments available. What’s more, persuasion-control may encourage more rational, informed engagement with these arguments. Significantly, it was considerations of these kind that led Scanlon himself to later disavow his 1972 argument. In a later paper Scanlon would suggest that caring about the intellectual autonomy of citizens, can come in different forms. Building on this insight, he suggested that his 1972 paper focused too narrowly

---

<sup>113</sup> Andrew Reid (2020). “Does regulating hate speech undermine democratic legitimacy? A cautious ‘no’” *Res Publica* 26, pp.181-99.

<sup>114</sup> Leiter, ‘The case against free speech’.

on merely protecting citizens from having their autonomy undermined by censors. As a result, the paper was insufficiently sensitive to the way censorship might sometimes promote our autonomy, as well as the threats to our intellectual autonomy that might a stringent free speech approach might contain.<sup>115</sup> Should we side with Scanlon on this matter? Should we abandon the idea that persuasion-control is unacceptable because of the direct threat to our autonomy that it poses?

There is a reply that can be made here by the proponent of the barrier-raising objection. They can argue that while the barrier-raising involved in persuasion-control can sometimes be all-things-considered justified, such barrier-raising is nonetheless always deeply problematic. An analogy will serve to clarify: just as stealing is always *pro tanto* wrong, even when it is necessary to secure certain beneficial consequences, interfering with the range of arguments available to agents always involves an unacceptable blow to the autonomy of citizens, even if this blow can be in some sense all-things-considered justified. If this conclusion is right, it would go a long way to vindicating the anti-censorship sentiments this thesis investigates. Still, for at least three reasons, it is unlikely that the barrier-raising argument can support even this more limited conclusion.

First, the barrier-raising argument struggles to explain why persuasion-control constitutes a distinctive, and especially problematic, threat to our autonomy. Consider the following cases: the government closes down a library because it has become covered in toxic waste, the government incarcerates an interesting but murderous political thinker, the government disperses a public debate after receiving a bomb threat. The point of Scanlon's critique is to single out persuasion-control as a distinctly objectionable case, different to other, more mundane government interference. But all of these cases, and many more, involve the government erecting barriers to what arguments can be heard. In a similar vein, the direct autonomy argument struggles to explain why government persuasion-control is any worse than the standard private market failure that place limitations on what ideas and arguments can be heard. These two points combine to present a challenge for the direct-autonomy approach to problematising persuasion-control. Many of the critics of persuasion-control, especially the principled critics that this thesis discusses, think that there is something especially

---

<sup>115</sup> Scanlon, 'Freedom of expression and categories of speech', p.534.

objectionable about persuasion-control, beyond the way it inconveniences citizens or makes it harder for them to hear certain views. Such censorious interferences are painted as violating our principled and deep commitments, rather than merely imposing material costs. For this reason, then, the direct autonomy approach is unlikely to fit the bill.

Second, as the likes of James Wilson and Sarah Conly have shown, governments are very much in the business of raising barriers and even coercing citizens.<sup>116</sup> Such interference can form part of an attractive political system in which reasonable limits are imposed on the freedom of citizens, in order to promote the wider long-term interests of society and its peoples. For instance, the government taxes its citizens, bans them from driving over a certain speed limit, bans them from owning certain kinds of weapons, refuses to enforce certain contracts, and demands that businesses pay their employees at least a set hourly wage. Similarly, governments sometime make it more difficult for citizens to eat unhealthy food, to go without proper healthcare, or to make rash health choices. It is plausible, then, to suppose that persuasion-control is not only analogous to many other forms of government interference, but that these interferences are just part and parcel of having a functioning government. If we do not wish for the scepticism towards persuasion-control being examined to collapse into an anarchist scepticism towards government interference in general, we will need to hear more specifically about why raising barriers to intellectual autonomy is a special case, and not simply business as usual.

Third, erecting barriers to freedom and autonomy is at least sometimes uncontroversial and comes at no moral cost. In particular, as Howard has argued, the government would seem to be generally fully justified in (proportionately) limiting the freedom of citizens to harm one another.<sup>117</sup> After all, it is not even controversial for the government to prevent citizens from committing murder and battery. Given that a broadly similar kind of harm-prevention rationale lies behind most cases of persuasion-control, we need to hear more from our censorship sceptic about why such a *prima facie* plausible justification does not render persuasion-control similarly uncontroversial.

---

<sup>116</sup> James Wilson (2011). "Why it's time to stop worrying about paternalism in health policy", *Public Health Ethics* 4 (3), pp.269-79; Sarah Conly (2013). "Coercive paternalism in health care: against freedom of choice", *Public Health Ethics* 6 (3), pp.241-45.

<sup>117</sup> Howard, 'Dangerous Speech'.

To end this section, let us consider what might be termed the ‘sovereignty version’ of the autonomy argument against persuasion-control. According to the sovereignty version, people should be sovereign with respect to the capabilities that constitute them as an autonomous agent. On this reading then, the problem with persuasion-control isn’t so much that it necessarily undermines the functioning of citizens’ autonomous capabilities, or that it constitutes a significant interference with the available arguments that we can engage with. Rather, it is that persuasion-control interferes in any way with capabilities that should be controlled exclusively by the citizens themselves. As with someone’s home and garden, it is the mere trespass by another upon this area that is unacceptable, even if nothing is damaged. This reading of Scanlon is encouraged by several passages in the 1972 article. Most notably, Scanlon writes that “to regard himself as autonomous in the sense I have in mind a person must see himself as sovereign in deciding what to believe and in weighing competing reasons for action”.<sup>118</sup> Even so, the reading fares no better than what has gone before. First, several of the objections canvassed above apply to this reading too. For instance, more needs to be said in order to substantiate the idea that persuasion-control is a special kind of interference with our sovereignty, such that it can be distinguished from more mundane government regulation.

Second, the sovereignty reading would appear to be premised on the notion that people are able to independently decide what to believe and weigh up their reasons for action, and that these abilities should be accorded a kind of sovereign status. The problem is that these abilities, such as they are, are always compromised to some extent, just by virtue of our cognitive limitations. Without our realising it, our decisions about what to believe and how to act are often influenced by internal biases – confirmation bias, framing effects, groupthink, self-serving bias and anchoring bias are all well-known examples – as well as external conditions and more generally the social environment in which we find ourselves.<sup>119</sup> Take Sarah Conly’s suggestion that:

---

<sup>118</sup> Scanlon, ‘A theory of freedom of expression’, p215.

<sup>119</sup> Confirmation bias, framing effects, groupthink, self-serving bias and anchoring bias are all well-known biases that impact upon our decision making in a way that would seem to undermine our sovereignty in these areas, see: Atsuo Murata, Tomoko Nakamura, and Waldemar Karwowski (2015) “Influence of cognitive biases in distorting decision making and leading to critical unfavorable incidents”, *Safety 1 (1)*, pp. 44-58; Andrea Caputo (2013) “A literature review of cognitive biases in negotiation processes”, *International Journal of Conflict Management* 24 (4), pp. 374-98; Robert Mark Simpson (2013). “Intellectual agency and responsibility for belief in free-speech theory”, *Legal Theory* 19 (3), pp. 307-30.

We are...unduly influenced by the particular description used in the presentation of our options (more likely to choose a medical procedure with a 20 percent chance of success than one described as having an 80 percent chance of failure); unduly prone to think that we ourselves are less likely than others to suffer misfortune, even of something entirely random, like lightning; prone to miscalculate the value of a thing depending upon whether we do or don't yet own it; prone to assuming things that have one superficial characteristic in common also have similarities throughout (commonly known as stereotyping).<sup>120</sup>

So far as we are never in complete control of decisions about what to believe, or in weighing up reasons for action, this blunts any objection to government actions that encroach upon our supposed sovereign powers in these respects. The demand for sovereignty, thus construed, looks implausibly idealistic from the start. It might be replied to this objection, that there is a key difference between our sovereignty being undermined by naturally occurring biases within ourselves, and our sovereignty being undermined by the deliberate decisions of other agents. This idea will be explored in more detail below, as will the more general thought that persuasion-control involves a distinctive kind of interference with our autonomy. For now it suffices to say that that while there may be a certain disanalogy here, it is not a disanalogy that can be cashed-out in terms of a diminishment of our sovereignty; one loss of sovereignty, is just as bad as another. If we wish to cash-out Scanlon's principled objection to persuasion-control, we will have to move beyond merely considering the direct impact of persuasion-control on the autonomy of citizens.

### §III. Motives and meaning

The previous section cast doubt on whether a direct autonomy-based argument can be used to justify the kind of principled anti-persuasion-control line that this thesis investigates. Chief among the challenges for such direct autonomy-based arguments, is to explain why persuasion-control constitutes a special threat to our autonomy, distinct from that involved in other kinds

---

<sup>120</sup> Conly, 'Coercive paternalism in health care: against freedom of choice', pp.21-22.

of governmental and private interferences in our lives. How might one go about trying to meet this challenge? Scanlon's comments on the normatively transformative nature of motivation when dealing with normative problems, may offer a solution.

What sometimes goes underappreciated by philosophical commentators on Scanlon, is that his critique of persuasion-control is not – cannot – be based on the generic way in which persuasion-control deprives citizens of information or thwarts their ability to consider various arguments and ideas.<sup>121</sup> One way of showing this is by reflecting on how (as we have seen) such generic autonomy-based arguments fail to establish the kind of distinctive objection to persuasion-control that Scanlon is looking for. Another is by reflecting on how in his 1972 article, Scanlon does not in fact object to censorship *simpliciter*, nor to depriving citizens of access to information. Indeed, the Millian Principle permits at least some cases, which Scanlon cites approvingly, of depriving citizens of information.<sup>122</sup> Rather, what Scanlon objects to is the act of government censorship *when it is based on certain motivations or justifications*. In particular he objects to governments depriving citizens of materials, *out of a concern for the way those materials might prove harmfully persuasive*. The crucial thought here seems to be that the presence of this kind of motivation changes the meaning of the government's act of censorship, such that it now poses a new kind of threat to our autonomy.

In many ways Scanlon's insight echos an analysis of moral concepts proposed in recent times by Kalle Grill, but which in fact has its roots in Kant's talk of maxims.<sup>123</sup> Grill has suggested that when we are analysing complicated moral concepts, oftentimes it is not enough that we merely analyse in isolation the action or reason behind that action. Rather, we need to analyse what he calls the *action-reason compound*. What this means is, we need to attend to any

---

<sup>121</sup> Scanlon, 'A theory of freedom of expression', p.209: "it is easier to say what the classic violations of freedom of expression have in common than it is to define the class of acts which is protected by that doctrine. What distinguishes these violations from innocent regulation of expression is not the character of the acts they interfere with but rather what they hope to achieve—for instance, the halting of the spread of heretical notions. This suggests that an important component of our intuitions about freedom of expression has to do not with the illegitimacy of certain restrictions but with the illegitimacy of certain justifications for restrictions. Very crudely, the intuition seems to be something like this: those justifications are illegitimate which appeal to the fact that it would be a bad thing if the view communicated by certain acts of expression were to become generally believed; justifications which are legitimate, though they may sometimes be overridden, are those that appeal to features of acts of expression (time, place, loudness) other than the views they communicate".

<sup>122</sup> Scanlon, 'A theory of freedom of expression', p.210.

<sup>123</sup> Kalle Grill (2007). "The normative core of paternalism", *Res Publica* 13 (4), pp. 441-581; Kant 'The Groundwork for a Metaphysics of Morals'.

moral problems involved in agents using *this* particular reason, to justify *that* particular action. The key idea here is simple enough: the meaning and normative implications of a given act, depend on the underlying reason; pushing someone down in order to save their life is very different to pushing someone down for your own amusement.<sup>124</sup>

How does this help shed light on the persuasion-control issue? The motivation that underpins persuasion-control may well be thought to transform an otherwise relatively unremarkable act of censorship.<sup>125</sup> In cases of persuasion-control, the government engages in censorship with the purpose of trying to steer citizens away from certain beliefs and intentions that it judges to be beyond the pale. Reflecting on both the action and motivation involved here, a philosopher like Seana Shiffrin might say that the government is attempting to ‘substitute’ in its judgement on the matters of what kind of beliefs and intentions citizens ought to possess, and substitute out (or even pre-empt) the contrary judgement of citizens.<sup>126</sup> Still, an even more straightforward way of characterising persuasion-control, is as an attempt by the government to ‘tell its citizens what to think’.<sup>127</sup>

Characterising persuasion-control in this way helps contextualise the complaint that persuasion-control conflicts in a special way with our intellectual autonomy. Consider the regulation of weapon ownership, battery, earnings, blackmail, bribery, threats, etc. Or, to focus on speech regulation that might limit the proliferation of ideas, consider the following cases again: the government closes a library covered in toxic waste; the government incarcerates an interesting but murderous political thinker; the government disperses an interesting political debate following a bomb threat. Like persuasion-control, these government regulations may well curb the material freedom and autonomy of citizens in various ways i.e. in the form of

---

<sup>124</sup> This point about action-reason compounds, or at least a very similar point, is sometimes made in the context of philosophical discussions of moral concepts like deception and manipulation, see Shlomo Sher (2011). “A framework for assessing immorally manipulative marketing tactics”, *Journal of Business Ethics* 102 (1), pp. 97-118.

<sup>125</sup> A note here: Grill cautions against simply taking the pro tanto bad of the action and adding it to the pro tanto bad of the reason. Instead Grill argues that rather than stacking up in a predictable manner like building blocks, it may be that some actions and reasons combine in unexpected ways to present us with distinctive normative problems

<sup>126</sup> Shiffrin, ‘Paternalism, unconscionability doctrine, and accommodation’.

<sup>127</sup> The phrase ‘telling us what to think’, as used here, is a term of art. Presumably the government can permissibly tell us what to think in the sense of e.g. advising us not to act a certain way, or presenting arguments in favour of certain behaviour. But the government, so the thought goes, should not ‘tell us what to think’ in the sense of attempting to interfere with the proliferation of ideas on the grounds that such interference will promote ‘better’ beliefs and intentions. Perhaps, then, this way of framing the worry with persuasion-control could be alternatively summed up as: the government should not forcibly tell us what to think.

limiting how citizens may act, or in the form of limiting what ideas and arguments citizens have access to. But only persuasion-control, insofar as it is motivated by a desire to promote certain beliefs and intentions over others, amounts to an attempt to tell citizens what to think. The thought here is that while the government may interfere in our lives and sometimes even engage in coercion, the government should nonetheless refrain from telling its citizens what to think. Or, to put it another way: while the government may use coercion to, say, restrain us from assaulting others, it should not be in the business of interfering with what persuasive appeals may be made, in an attempt to prevent citizens from ever arriving at the intention of performing assault in the first place.

#### §IV. Democratic values

It is tempting to leave the argument here. On the one hand it might seem like we have a perfectly viable account of the way persuasion-control involves a distinctive interference with our autonomy. Certainly it is helpful to characterise persuasion-control as a kind of substitution of intellectual autonomy where the government tries to forcibly tell its citizens what to think; such a description helps unpack the concept of persuasion-control, and will serve as a useful description throughout this thesis. On the other hand, we have still yet to explain why the government should not attempt to forcibly tell citizens what to think. When it comes to explaining the principled worry with persuasion-control, this re-description rather kicks the can down the road. Moreover, direct-autonomy accounts fail to shed light on the issue. After all, while there may be something distinctive in the character of persuasion-control – something unique about the attempt to tell others what to think – note that this doesn't correspond to a more severe impairment of our intellectual capacities.

What are we to do next, then? As we saw earlier, Scanlon invokes the value of democracy and government legitimacy in order to problematise persuasion-control. Do Scanlon's suggestions help cash-out a principled objection to persuasion-control? Do his comments help frame governmental attempts to tell citizens what to think, as conflicting with our democratic values? Perhaps. In truth, Scanlon's comments on democracy in his 1972 article are rather sparse. What comments we do get have a rather table-thumping quality to them. Scanlon suggests that democratic governments are committed to treating citizens as intellectually autonomous – virtually having a side-constraint duty to refrain from telling

citizens what to think. The key underlying thought behind Scanlon's account is that democratic institutions are built *for* intellectually autonomous citizens. That is to say, the key warrant of a democratic government is to maintain and protect the intellectual autonomy of citizens, rather than promote overall intellectual autonomy, or make efficient trade-offs. As such, while persuasion-control may help secure real benefits, it does so in a way that democratic governments have no warrant to do. After all, the benefits accrued by persuasion-control depend on the government riding roughshod over the intellectual autonomy of citizens via telling them what to think. The benefits in question may be real, but they are benefits that a democratic government cannot appeal to in justifying its actions. Or, to put it another way, democratic governments may only appeal to benefits of a certain kind when justifying their actions, but these benefits are by definition absent in the persuasion-control case.

#### §V. Unsupported foundations

Where does all of this leave us? Having suggested that direct autonomy-based arguments against persuasion-control do not work, we considered Scanlon's focus on motivations and whether persuasion-control can be helpfully framed as an attempt by the government to tell us what to think. Moreover, we have considered how Scanlon's comments on democratic legitimacy might help establish an objection to this persuasion-control. We draw this chapter to a close by considering two powerful objections to Scanlon's account. The first objection draws on a critical insight offered by Howard, while the second builds on a suggestion made by Robert Amdur. Ultimately it is argued that while we may be able to defend Scanlon from these objections, these defences contain important explanatory gaps. In the end, Scanlon's rather brief account of democratic legitimacy is ruled to be too brute, with key ideas still left to be unpacked. The rest of the thesis is dedicated to the task of unpacking these ideas, and considering more detailed accounts of how our democratic values conflict with persuasion-control.

The first objection runs with a subtle point made by Howard in his discussion of dangerous speech.<sup>128</sup> Howard writes that:

---

<sup>128</sup> Howard 'Dangerous speech'.

While citizens' moral faculties are working satisfactorily in cases of reasonable disagreement—when there are good arguments on both sides of a debate—the same cannot be said of citizens who are defending incontrovertible injustice. Indeed, if the impetus for respecting citizens' moral agency flows from our recognition of them as agents of justice...then we rightly view them as duty-bound to refrain from activities that are morally incompatible with these duties. A moral prerogative to incite incontrovertible injustice is incompatible with the obligations of democratic citizenship.<sup>129</sup>

A couple caveats are worth mentioning here. The comments above relate to the autonomy of speakers, and are just one step in a larger, more complicated argument presented by Howard. Still, we can distil a key insight here: it is one thing to talk about the democratic importance of citizens being intellectually autonomous, but our democratic values also tell against citizens being free to use this intellectual autonomy to engage in, and be engaged by, speech that endangers others. Here the idea isn't that we should sacrifice our democratic values in order to promote the well-being of others, or that our democratic values can be permissibly undermined in order to promote other values. It is rather that the very democratic commitments that generally tell in favour of protecting our intellectual autonomy, in fact also tell against allowing this freedom in the case of dangerous speech. It is, in other words, our democratic values that call for such sacrifices. After all, there is more to a democracy than citizens being free to say and hear what they like. Citizens must also conduct themselves in ways that are duly morally responsible, and that respect the claims of their fellow citizens.

This is a sensible objection that reminds us how our commitment to democracy comes with duties, as well as privileges. Indeed, while democratic governments may generally want to treat what ideas and arguments citizens have access to as a private matter, this plausibly ceases to be a private matter when innocent citizens are thereby exposed to (risk of) harm. Still, it may well be possible for Scanlon to defend himself from this objection. In particular he could argue that while our commitment to democracy does indeed impose certain responsibilities and duties on us, none of these responsibilities or duties ought to entail the government forcibly

---

<sup>129</sup> Howard 'Dangerous speech', p.232.

telling us what to think. The idea here is that intellectual autonomy, far from being just another aspect of our democratic vision, lies at the very foundation of our commitment to democracy – it is a value that precedes and pre-empts other democratic values. To reiterate a point made earlier: the suggestion is that while a well-functioning democratic government should do its absolute best to promote the interests and protect the rights of its citizens, there are nonetheless limits on how it may go about doing this. One of those limits is that it must refrain from telling citizens what to think. For the time being, this will serve to at least partially defend Scanlon from the first objection.

Still, Scanlon isn't out of the woods yet. The second objection suggests that, contra Scanlon, we can take the intellectual autonomy of citizens to be centrally valuable to our democracy, while still permitting limited interference with the intellectual autonomy of citizens. In particular it is argued that interferences with our intellectual autonomy can be justified within a democracy, whenever intellectually autonomous agents who are suitably rational and concerned with the interests of others as well as themselves, would agree to those interferences. The contractualist thought here, proposed by the likes of Amdur, and touched upon by the likes of Brison and Easton, is that intellectually autonomous agents can use their intellectual capacities, to see that those very same intellectual capacities should sometimes be curtailed – i.e. that the government may justifiably 'tell them what to think' in the form of persuasion-control.<sup>130</sup> Brown makes a similar point about how 'free and equal' agents might agree to limits on hate speech in order to protect their civil dignity.<sup>131</sup> The point here isn't so much that the losses of autonomy are justified by the potential gains in autonomy – although that is part of the story. Rather, it is that because of this consideration, citizens might use their intellectual autonomy to embrace interferences with their autonomy. In this case, these interferences are reconceptualised as compatible with respecting the intellectual autonomy of citizens.

How might Scanlon respond to this objection? Once again, with a limited range of options, he might choose to double-down on his point that respecting the intellectual autonomy of citizens constitutes a foundational, constitutive condition of a functioning

---

<sup>130</sup> Robert Amdur (1980). "Scanlon on freedom of expression", *Philosophy & Public Affairs* 9 (3), pp. 287-300; Susan Brison, 'The autonomy defence of free speech'; Susan Easton, 'Autonomy and the free speech principle'.

<sup>131</sup> Brown, 'Hate speech laws, legitimacy, and precaution: a reply to James Weinstein'.

democracy. According to this reply, respecting citizens as intellectually autonomous is not just another democratic commitment to be placed along others; it is rather the foundational democratic commitment. Another way to put this point, is that the arguments and ideas that citizens rationally engage with, must be considered a private matter, beyond the reach of a democratic government's legitimate authority. Granted, so the thought goes, the government has other democratic duties (e.g. promoting equality, respecting the rights of citizens), and it should pursue those duties. But not when this pursuit involves trampling over its more fundamental commitment to treating citizens as intellectually responsible agents that think for themselves.

We are almost at the close of this chapter, and so it is notable that we have arrived at a tentative defence of the Scanlonian objection to persuasion-control. On this defence, Scanlon's anxiety can be defended even from sensible and well-motivated objections, but only by construing the democratic state as having a foundational commitment to treating citizens as autonomous agents that should never be told what to think. Here we arrive at the main problem with Scanlon's account, and indeed the problem that has been lurking in the background of this chapter the whole time – the bruteness problem. How can we construct a principled and deep objection to persuasion-control, without simply brutally assuming that it is beyond the pale? In Scanlon's case, he somewhat brutally assumes that the government has a foundational and near side-constraint commitment to treating its citizens as intellectually autonomous. Our democratic values preclude the government telling us what to think, so the suggestion goes, even when those very same values, not to mention other values such as promoting welfare, also help explain why it might be important to tell citizens what to think.

The problem with this, is it leaves us without a substantive argument as to why we should accept this democratic vision in the first place. Indeed, Scanlon does little to even contextualise or explain the attraction of this democratic vision. To put it another way, Scanlon's account lacks the kind of fine-grained analysis of our democratic values, the status of citizens within a democracy, the role of the government, etc, that would help justify his objection to persuasion-control. But more than this, Scanlon's brute assumptions about democratic legitimacy are also 'normatively expensive' – conflicting with some of our normative intuitions, and thus standing in special need of justification. Drawing together several of the insights gleaned from our discussion, this chapter ends by commenting on what's driving up the price here.

The Scanlonian approach to persuasion-control will strike many as hyperbolic; an unreasonable response to reasonable government action. In particular it can be argued that, in contrast to Scanlon's rather alarming characterisation, the best kind of persuasion-control is in fact 'justified', 'piecemeal' and 'light of touch'. Such persuasion-control is 'justified' insofar as it forms part of an effective strategy of changing attitudes in society, in a way that will predictably lead to less dangerous beliefs and intentions, and will ultimately help prevent serious harms from taking place. It is piecemeal insofar as persuasion-control is only an attempt to curb our *harmful* beliefs and intentions. Persuasion-control, after all, does not interfere with the beliefs and intention of citizens more generally, or attempt to push citizens towards simply 'better' beliefs and intentions. On the contrary, it is an effort to steer citizens away from just those beliefs and intentions that might prove harmful and likely damaging towards democratic harmony.

Finally, the best kind of persuasion-control is light of touch insofar as it imposes only relatively low barriers to the prohibited beliefs and intentions in question. Persuasion-control, for instance, falls well short of brainwashing or conditioning or other such activities that might constitute an egregious assault on our intellectual autonomy. Granted, persuasion-control is supposed to make it harder for us to access certain ideas and arguments. But often times persuasion-control only applies to expression in public spaces, and even when it applies to all spaces, the point of persuasion-control is merely to make it harder or less likely – rather than impossible – that those arguments will be considered. In some ways, there is an analogy to be drawn here with nudging. Roughly put, governmental nudges make it easier for citizens to make good choices, or at least harder for them to make bad ones. Nudges achieve this non-coercively by, for instance, making the good choices the default option, attaching more paperwork to bad choices, highlighting (sometimes in a very literal sense) good options, and drawing attention to the dangers involved in bad ones.<sup>132</sup> When these measures are imposed on thousands of people at a time, they result in a predictable increase in the number of good choices being made. Yet no single individual is coerced or has their autonomy interfered with

---

<sup>132</sup> Jennifer Blumenthal-Barby (2013). "Choice architecture: a mechanism for improving decisions while preserving liberty", in *Paternalism: Theory and Practice* (Cambridge: Cambridge University Press), ed. Christian Cowons and Michael Weber, pp.178-97; Andreas Schmidt and Bart Engelen (2000). "The ethics of nudging: An overview", *Philosophy Compass* 15 (4), pp.1-13; Richard Thaler and Cass Sunstein (2008). *Nudge: Improving Decisions About Health, Wealth, And Happiness* (London: Yale University Press).

in a significant way. The most normatively attractive cases of persuasion-control similarly aim to strike this balance. In sum, Scanlon's brute assumptions about democratic legitimacy not only stand in need of further development, this development also proves to be essential if we are to problematise censorship that, for the reasons discussed above, will strike many as utterly reasonable.

### Conclusion

What lessons can we take from this chapter? Well for one, we have confirmed that a direct autonomy-based argument is unlikely to substantiate the kind of principled censorship scepticism we are investigating in this thesis. That is, whatever the deep, principled worry is that the likes of Scanlon and Nagel have with persuasion-control, this worry goes beyond the direct impact that persuasion-control has on our autonomous capacities. Of course, this doesn't imply that we should jettison talk of autonomy altogether. It may be useful, for instance, to characterise persuasion-control as involving an attempt by the government to tell its citizens what to think. But if this is to establish a principled argument against persuasion-control, more needs to be said. Evoking the value of democracy is a promising option. Although more fine-grained analysis will be needed than, say, Scanlon provides.

The focus of this thesis now becomes whether we can provide this missing analysis. Can we cash-out the connection between our democratic values and intellectual autonomy in a philosophically satisfying way? Can we explain precisely how our democratic values establish a principled and powerful objection to persuasion-control? The next chapter examines whether our democratic vision ought to be committed to a certain conception of 'the citizen', such that persuasion-control ought to be rejected. In particular the chapter considers whether persuasion-control involves a lack of respect or recognition for the citizens who are told what to think.

## Chapter 3

### Recognising people; recognising flaws

The government bans the praise of terrorist attacks because it is worried that such praise will inspire further terrorist acts; in a bid to reduce the number of murders in society the government makes it illegal to publicly advocate murder; the government censors arguments that tell in favour of holding racist, sexist and otherwise discriminatory views, for fear that the dissemination of such views will promote discrimination and fortify pre-existing identity-based prejudices; in an effort to get more parents vaccinating their children, the government introduces restrictions on the publication of anti-vaccination propaganda.

The cases above all fall within the category of what has been termed ‘persuasion-control’; a subset of censorship that involves the government limiting our access to ‘corrupting’ arguments and ideas that it legitimately fears might persuade citizens to develop harmful beliefs and intentions. In this chapter we continue to explore the scepticism many register towards persuasion-control, by asking whether persuasion-control might amount to an insulting or degrading interference with citizens.<sup>133</sup> Several subtly different versions of this worry are explored, but what unites each of them is the idea that our democratic institutions ought to be

---

<sup>133</sup> In this sense, the arguments canvassed in this chapter are ‘listener-based’, insofar as they construct arguments against persuasion-control that build upon the claims of those would-be listeners who are affected by persuasion-control.

committed to a certain vision of its citizens, and that this in turn ought to inform the government's actions.<sup>134</sup> When it comes to persuasion-control, the worry is that this kind of interference involves a failure to properly recognise the intellectual capacities of citizens.<sup>135</sup> By contrast, so the thought goes, an alternative political scheme that prohibits persuasion-control, or permits it only in exceptional cases, does a better job of appropriately conceptualising citizens as independent, rational, morally responsible agents.

Throughout this chapter I assume that insulting regulation involves a cost to the government's political legitimacy. The idea here is that insulting regulation is not just wrongful or objectionable, but in fact lies in special conflict with our democratic vision of how the government ought to operate. Why might one think this? We will highlight two main ideas. On the one hand, many theories of governmental (democratic) legitimacy take consent to be necessary for political legitimacy. This comes in various forms: some theories state that what matters is that actual citizens have, or at least could have, consented to the government's conduct. Other theories hold that what matters is that idealised citizens could hypothetically consent to the government's conduct.<sup>136</sup> However one takes consent to ground political legitimacy, one is likely to judge insulting regulation to constitute a threat to the government's legitimacy. The idea here is that citizens have not, or would not consent, to insulting regulation. Or at the very least, such insulting regulation would stand in need of special justification if citizens (real or hypothetical) are to consent to it. On the other hand, there is the idea that a democratic political community must involve moral equals. While some citizens may hold a position in government and thus wield political authority over others, all citizens must be treated as moral equals and granted e.g. an equal, minimal kind of respect and

---

<sup>134</sup> This vision of citizens may be based on a realistic assessment of the capabilities of the average mentally competent citizen, or else may be an idealised construction. For the former, see Dworkin, '*Freedom's Law: The Moral Reading of the American Constitution*'. For the latter, see Rawls 'Kantian constructivism in moral theory; Rawls, '*A Theory of Justice*'.

<sup>135</sup> To clarify, the suggestion here is not that the insulting or demeaning quality of persuasion-control necessarily renders this kind of intervention impermissible, although this may well sometimes be the case. Rather, the more subtle suggestion is that even when compelling harm-based reasons ultimately tell in favour of persuasion-control, such persuasion-control may nonetheless be objectionably insulting in a way that results in a deficit in legitimacy. So construed, even the best kind of persuasion-control necessarily involves a significant political cost that leaves us with "something to regret".

<sup>136</sup> Amanda Greene (2016). "Consent and political legitimacy", in *Oxford Studies in Political Philosophy* (Oxford: Oxford University Press), ed. David Sobel, Peter Vallentyne, and Steven Wall, pp.71-97.

recognition. In practice, this requires that citizens not be insulted or humiliated unnecessarily; that they should not have to grovel before their government in order to be treated decently, or be compelled to flatter or pander government officials.<sup>137</sup> Rather than taking ‘democracy’ to be simply synonymous with a community of moral equals’, it is sometimes supposed that the ideal of democracy in fact proceeds from the more fundamental fact of moral equality between persons. On this line of thought, the concept of a democratic community and the related duties such a community involves, emerges from having to maintain moral equality between all citizens, even while some citizens are in a position of political authority over others.<sup>138</sup>

The chapter is structured as follows: §I explores the idea that persuasion-control is insulting qua involving a negative appraisal of the citizens being interfered with. The key idea here is that persuasion-control involves a lack of what Stephen Darwall terms ‘appraisal respect’, insofar as the government is suggesting that citizens cannot be trusted to manage their own beliefs and intentions.<sup>139</sup> Inspired by Nagel, §II explores the idea that persuasion-control may diminish the political status of citizens. Particular attention is paid to the worry that persuasion-control diminishes the status of citizens as independent, thinking beings. Finally, §III explores the suggestion that persuasion-control is incompatible with a full appreciation of the thinking nature of citizens, and thus involves a lack of what Darwall would term ‘recognition respect’.

Ultimately it is suggested that none of these accounts vindicates the view that persuasion-control is deeply insulting towards autonomous agents. On the contrary, it is argued that persuasion-control may in fact stem from a rather richer appreciation of the diverse range of capabilities, needs, vulnerabilities, and responsibilities of autonomous agents, as well as the need for co-operation between citizens if flourishing is going to be achieved on any kind of large scale. Far from being insulting or demeaning, it is argued that this is an attractive, grounded conception of citizens with which to underwrite our democratic vision.

## §I. Appraisal respect

---

<sup>137</sup> Judith Resnik and Julie Chi-hye Suk (2003). “Adding insult to injury: questioning the role of dignity in conceptions of sovereignty”, *Stanford Law Review* 55 (5), pp.1921-62.

<sup>138</sup> Allen Buchanan (2002). "Political legitimacy and democracy", *Ethics* 112 (4), pp.689-719.

<sup>139</sup> Stephen Darwall (1977). “Two kinds of respect”, *Ethics* 88 (1), pp. 36-49.

Perhaps the most straightforward way in which some form of conduct or verbal expression can be insulting, is in involving a negative assessment of another agent. The relatively simple idea here is that we insult others when we make negative appraisals of their actions, character, capacities, etc. For instance, I might insult my neighbour by suggesting that his moral character is lacking in some way, or that his athletic abilities are sub-par. One way of framing such insults, is to say that they involve a lack of what Stephen Darwall calls ‘appraisal respect’.<sup>140</sup> In other words, I insult and show a lack of respect for another person, when I appraise their actions, character, capabilities, etc, and find them wanting. Can this help shed light on the suggestion that persuasion-control is insulting?

It is probably fair, albeit a little blunt, to say that the government that engages in persuasion-control finds their citizens wanting in some respect. After all, the government that engages in persuasion-control does so because it fears that at least some citizens, left to their own devices, and left free to hear all available persuasive appeals, will end up developing harmful beliefs and intentions that place others in danger. In this way, the government judges that citizens cannot be altogether trusted to manage their own beliefs and intentions by themselves, and would benefit from state interference. What’s more, it is easy to see why some might find this insinuation to be insulting; certainly, many take pride in being able to decide for themselves what is right and wrong, as well as more generally what kind of beliefs and intentions they ought to have.

At this point it will be useful to quickly canvass and reject a potential objection a reader might have. The government, so the objection goes, frequently and seemingly without much controversy suggests that its citizens lack competency. After all, traffic safety laws, contract laws, food packaging laws, and almost any other law one cares to mention, all imply that citizens sometimes need help from the government, and would likely come up short if left to their own devices. Along these lines, persuasion-control is no more insulting than the most mundane and uncontroversial forms of government regulation.

In response, let us offer that we should be wary of trying to draw an analogy between persuasion-control and more run-of-the-mill regulation cases. To see why, we need only to remind ourselves that agents take more pride in managing certain areas of their lives compared

---

<sup>140</sup> Darwall, ‘Two kinds of respect’.

to others. For instance, most agents are more likely to be insulted by the suggestion that they cannot adequately take care of their own children, than they are by the suggestion that they cannot drive safely without the government's help. Similarly, it is more insulting for someone to suggest that you cannot look after your own health, than for them to suggest that you sometimes come up short reading the fine print on a contract. Along these lines, those who take persuasion-control to betray a special lack of appraisal respect for citizens, will likely claim that judging for one's self what persuasive appeals are effective, and deciding for one's self which beliefs and intentions to develop in light of these appeals, is an especially important part of what it means to operate as a mature and responsible agent. Insinuations of incompetence and attempts to take over the management of this area, are thus revealed to be especially insulting in a way that distinguishes persuasion-control from more mundane government interferences with our lives.

This point will serve to preserve the suggestion that persuasion-control is at least *prima facie* distinctively insulting to autonomous agents. What distinguishes persuasion-control as a special case of insulting government interference, so the argument goes, is that it involves the government suggesting that citizens cannot be trusted to manage an area of their lives so fundamental to what it means to function as a responsible and autonomous agent, as deciding for one's self what to think in light of available arguments. Have we, then, arrived at a persuasive account of why persuasion-control is so objectionable? Should we conclude that what marks out persuasion-control as a troubling and delegitimising form of government interference, is the way it delivers a distinctive insult to citizens? The answer is 'no'; despite the pride many take in deciding for themselves what to think, there is nothing objectionably insulting about suggesting that we would benefit from state intervention from time to time. In particular, three points are worth noting.

First, whether the suggestion that John is falling short in some endeavour counts as insulting, depends on the challenges involved with that endeavour, and whether we can reasonably expect errors to be made in the pursuit of that endeavour.<sup>141</sup> Consider the case where John's PhD supervisor reflects upon the work he has produced, and points to places where that work could be improved. Even granting that the supervisor is making a negative

---

<sup>141</sup> Or at least, amongst other things. For it may be that the necessity of the suggestion, or the way in which the suggestion is made, also contribute to whether the suggestion counts as insulting.

appraisal of John and his work, this negative appraisal needn't strike us as insulting so long as it takes place in the context of a recognition of the challenges and expected difficulties involved in completing a PhD. If John's supervisor knows that producing a PhD thesis is difficult, and that occasionally making mistakes is just part and parcel of completing a PhD, then there needn't be anything wounding or insulting when he points to places where John's work might be improved.

A similar story scans on to the case of governmental persuasion-control. The government might view the development of misguided (and even harmful) beliefs and intentions as just part and parcel of human reasoning, and to be expected when we have a free-flow of different arguments and ideas. Persuasion-control then, rather than stemming from a supercilious or insulting vision of citizens as incompetent, might instead be rooted in a view of about the difficulties involved in coming to develop the right (or at least harmless) kind of beliefs and intentions, and the imperfect nature of human reasoning.<sup>142</sup>

For instance, no citizen can claim to be a perfect reasoner or a flawlessly rational creature. In particular, a good deal of research has gone into showing how our decisions about what to believe and how to act are often, without our realising it, influenced by a number of biases and prejudices outside of our control; confirmation bias, framing effects, groupthink, projection bias, self-serving bias, and anchoring bias, for instance, are all well-known biases that impact upon our decision making in a way that would seem to undermine the rationality and control we have in these areas.<sup>143</sup>

Moreover, sometimes the nuances and technical difficulties associated with a subject, can lead us to make (harmful) errors about what to believe and how to act. The point is made clear when we reflect on disputes concerning immigration, the connections between religious groups and terrorism, climate change, vaccinations, and the threat posed by our political adversaries. All of these topics involve various complexities, and disputes often turn on the more technical aspects of these topics, disagreements about what the empirical data tells us, and to what extent

---

<sup>142</sup> This point is inspired by a suggestion Peter de Marneffe makes in: (2006). "Avoiding paternalism", *Philosophy & Public Affairs* 34 (1), pp. 68-94.

<sup>143</sup> Caputo, 'A literature review of cognitive biases in negotiation processes', *International Journal of Conflict Management*, Murata et al 'Influence of cognitive biases in distorting decision making and leading to critical unfavorable incidents', Matthew Rabin (2003). "Projection bias in predicting utility", *The Quarterly Journal of Economics* 118 (4), pp.1209-48.

we can rely on ‘experts’ to help fill-in gaps in our own understanding. What’s more, charismatic speakers are often adept at exploiting these complexities and our own lack of expertise. In short, then, even ostensibly reasonable citizens, reflecting upon the various arguments and ideas they are exposed to, can come to mistaken and even harmful conclusions on these topics.

Even leaving aside the issue of our reasoning biases, as well as some of the subject-based difficulties that can affect our reasoning, arriving at the right beliefs and intentions can be made considerably more difficult simply by the emotional context in which we find ourselves. Consider recent calls to censor particularly exploitative kinds of so called ‘seed faith’ appeals. Seed faith appeals involve religious leaders encouraging their followers to donate (sometimes very large) sums of money to the Church, while assuring their followers that these donations will be rewarded by God in this lifetime. Sometimes these appeals prey on highly vulnerable people. For instance, leaders have convinced seriously ill believers to donate thousands of dollars to the Church rather than spending that money on medical treatment, convincing them that they stand a better chance of being cured by God than by a doctor. Calls to ban these kinds of seed money appeals are often dismissed as insulting. But there’s nothing insulting about recognising that those suffering from deadly medical conditions might be particularly open to exploitation, or that those in desperate situations might be especially prone to making errors in their reasoning. Granted, such interferences suggest that people are susceptible to making mistakes when their backs are against the wall. But there is no shame in being vulnerable in this way; or at least, there shouldn’t be.

These points all tell in favour of a relatively simple thesis: there needn’t be anything insulting or shaming in the government’s suggestion that its citizens occasionally need help developing the right kind of beliefs and intentions. Undoubtedly the government that engages in persuasion-control is implying that some of its citizens are in danger of erring in some way. But to err is human, and there is nothing insulting about being told you are human.<sup>144</sup>

---

<sup>144</sup> There is a related worry here about whether we should trust the government to make better judgements than us about what to think. The worry can be expressed as follows: ‘granted, we may not always adequately manage the matter of deciding what thoughts to develop, but why think the government’s micro-management will do any better?’. In response to this, note that this is not so much as a worry about the insultiness of governmental persuasion-control, as it is a scepticism about how trustworthy the government is, or how effective its interventions are likely to be. Moreover, while these are sensible concerns to have, they do not tell decisively against persuasion-

One might worry that this line of reasoning risks proving too much. In particular a reader might worry that we are suggesting that governmental persuasion-control never involves the government making a negative appraisal in an insulting way. In response, let us say that governmental persuasion-control certainly *can* involve an insultingly negative appraisal. For instance, if the government believed its citizens needed help arriving at the right kind of thoughts on account of the fact that they are exceptionally stupid, or morally degenerate, then this would surely count as an insultingly negative appraisal. The key point being made above isn't that the negative appraisal involved in governmental persuasion-control can never be insulting. Rather, it is that this negative appraisal is not insulting when it stems from a perfectly reasonable and sympathetic awareness of the difficulties involved in forming beliefs and intentions.

Moving on to the second objection; even if we set aside this first objection and insist that persuasion-control involves an insultingly negative appraisal, a lot rides on whether this insulting view is warranted or not. David Enoch argues that whether an insult rises above just being hurtful and becomes *pro tanto* morally wrong, depends largely on whether it has sufficient epistemic warrant.<sup>145</sup> If all the evidence suggests that I intend to steal your birthday cake, then it is not objectionable for you to form the insulting (but fair) judgement that I am a thief and to decide to pre-emptively hide your birthday cake. Indeed, in this case arriving at some other less insulting conclusion about my likely behaviour would seem to require that you ignore the evidence and adopt an irrational belief. Similarly, it might be *prima facie* insulting for a judge to deny bail to a defendant awaiting trial because they believe them to be a legitimate flight risk. Still, assuming that the judge's decision is based on a reasonable assessment of the evidence they have before them, this decision ought to strike us as justified and unobjectionable.

How does all of this this apply to the case of persuasion-control? It would seem that whether the suggestion of incompetence involved in governmental persuasion-control is in fact objectionable, largely depends on whether this suggestion is epistemically warranted or not. With this in mind, if the government has done the research and assessed the available evidence, and upon doing so concluded that it has compelling reason to think that some

---

control per se, or make sense of the principled worry many register towards persuasion-control. What these concerns leave unexplained, in other words, is why we might object even to effective persuasion-control.

<sup>145</sup> David Enoch (2016) "What's wrong with paternalism: autonomy, belief, and action", *Proceedings of the Aristotelian Society* 116 (1), pp.21-48.

citizens will develop harmful views without its intervention, then the subsequent decision to engage in persuasion-control in an effort to prevent these harms need not strike us as objectionably insulting. On the contrary, it should strike us as a reasonable decision to make in light of the relevant evidence.<sup>146</sup> As David Enoch might put it, the government's negative appraisal in this case might not be to the liking of everyone, but it isn't objectionable (in the sense under current consideration) for the government to simply follow sensible epistemic norms. Indeed, forming a less insulting appraisal in the case described would qualify as epistemically peculiar.<sup>147</sup>

Of course, even epistemically warranted insults might still be objectionable if they are needlessly hurtful or simply designed to cause offense. Enoch's account appears to overlook that there is something cruel in my turning to the overweight stranger sitting next to me on the train and taunting him for being 'fat', even if this judgement is true. Passing this kind of negative appraisal, at least in such a public way, seems unnecessarily callous and designed only to hurt the feelings of my fellow passenger. But this thought, as right as it is, in no way implies that governmental persuasion-control is therefore insulting. For while governmental persuasion-control involves the government passing a negative judgement concerning the ability of (some of its) citizens to alight upon the right kind of beliefs and intentions, this negative judgement need not stem from any vindictive or cruel intentions. On the contrary, the government that passes this judgement and uses it as the basis for censoring certain persuasive appeals, may well be acting from a perfectly appropriate desire to prevent innocent agents from being harmed.

Third, even leaving these issues aside, it seems doubtful that what makes governmental persuasion-control so problematic is really the mere insulting suggestion of incompetence. Even conceding, *arguendo*, that persuasion-control involves an insulting attitude on the government's part, this would still only put persuasion-control on a par with a public service announcement (PSA) where the government declared that its citizens are lamentably incompetent. But this rather undersells the supposed wrong of persuasion-control. To the extent then that critics of persuasion-control wish to classify persuasion-control as an especially worrying case of government interference, it seems that we will have to move beyond simply examining the

---

<sup>146</sup> See Richard Arneson (2015). "Nudge and shove", *Social Theory and Practice* 41 (4), pp.668-91.

<sup>147</sup> David Enoch, 'What's wrong with paternalism: autonomy, belief, and action'.

insulting message that lies behind this intervention. In other words, if we wish to make sense of the deep anxiety felt by critics of governmental persuasion-control, we will need to do better than merely conceptualising governmental persuasion-control as a kind of mean-spirited announcement or nasty message.

## §II. Nagel and status

How else might persuasion-control be deemed to be insulting? One interesting suggestion comes from Thomas Nagel in his 1995 essay '*Personal rights and public space*'.<sup>148</sup> In the essay Nagel touches upon a number of critical insights concerning rights, freedom of expression, equality, and the role of sex in the public sphere. Still, his central aim is to substantiate what he sees as a widely held worry with content-based government (including what has been termed persuasion-control). In particular Nagel takes aim at attempts by the government to curb speech that it judges likely to reinforce sexist, racist, homophobic, and other kinds of identity-based prejudices.<sup>149</sup> Drawing on the work of Frances Kamm and Warren Quinn, Nagel suggests that as a matter of moral status, citizens should be exempt from these kinds of censorious interference.<sup>150</sup> How does this argument work?

At the foundation of the Nagelian account, lies the thesis that the inherent moral status of persons determines what kinds of freedoms those persons ought to enjoy, and what kind of interferences with those persons are permissible.<sup>151</sup> When it comes to assessing the government's interferences with our lives, Nagel is interested in how the moral status of citizens determines what kind of burdens the government may impose upon them, and what kind of justificatory reasons they may appeal to when imposing these burdens.<sup>152</sup> Nagel's core

---

<sup>148</sup> Nagel, '*Personal rights and public space*'.

<sup>149</sup> Nagel, '*Personal rights and public space*', p.96.

<sup>150</sup> Frances Kamm (1996). *Morality, Mortality: Volume II: Rights, Duties, and Status* (Oxford: Oxford University Press); Warren Quinn (1989). "Actions, intentions, and consequences: the doctrine of doing and allowing", *The Philosophical Review* 98 (3), pp.287-312.

<sup>151</sup> On p.85 Nagel writes that: "Moral status, as conferred by moral rights, is formally analogous to legal status, as conferred by legal rights, except that it is not contingent on social practices. It is a universal normative condition, consisting of what is permitted to be done to persons, what persons are permitted to do, what sorts of justifications are required for preventing them from doing what they want, and so forth".

<sup>152</sup> On p.94, for instance, Nagel asks: "What kind of force may be used against people, and for what reasons?". Note that this closely resembles Grill's analysis of so called 'action-reason compounds' discussed in chapter 2. See Grill, '*The normative core of paternalism*'.

idea here is simple enough; beings that enjoy a higher moral status have powerful claims against certain kinds of interferences, even though these interferences might be acceptable when dealing with beings of a lower moral status. For instance, perhaps certain kinds of animals can be permissibly killed for food or hunted for sport, but human beings – possessing an elevated moral status – cannot be used and abused in this way. Or consider the case where we must decide whether to kill one person in order to save five (different) people from being killed. On a straightforward consequentialist analysis of this case, we ought to kill the one in order to save the many; after all, why protect just one person from being killed, when you can protect five? But Nagel demurs. Those who hold that persons may not be killed in order to prevent five further killings, view persons as enjoying the kind of elevated moral status that means they may not be compelled to sacrifice their life in the name of the greater good. In other words, Nagel tells us, they view persons as enjoying an inherent ‘inviolability’, possessed by higher moral beings.

It is helpful at this point to tease out just what is at stake in Nagel’s suggestions here. When Nagel – and Kamm before him – say that persons are ‘inviolable’, they of course do not mean that persons are physically invulnerable or cannot be compelled to make these sacrifices. Rather, what they mean is that persons, unlike say livestock, are just the kinds of beings that should not be forced to sacrifice life and limb in the name of promoting the greater good; that persons enjoy an altogether higher moral status that grants them normative immunity from making these sacrifices. Nagel’s suggestions here speak to an almost sublime, quasi-religious understanding of persons, as belonging to an order of moral significance that reaches beyond the material confines of this world, and which grounds a seemingly undefeatable moral claim against being sacrificed in this manner.

These reflections upon the moral status of persons, Nagel tells us, ought to inform how the government treats its citizens and what kind of interferences with those citizens it is willing to engage in. In particular the government ought to grant citizens the kind of political status – that is, the kinds of political rights and liberties – that they are owed as beings possessing an elevated moral status. When it comes to the government that allows its citizens to be sacrificed, the problem isn’t so much that this deals a material blow to the freedoms of its citizens, although this might well be the case. Rather, what has centrally gone wrong is that the government has failed to fully recognise the elevated moral status of the persons they rule;

bestowing upon those citizens a demeaning political status that fails to treat them as the inviolable moral beings they in fact are. As Nagel puts it:

What is good about the public recognition of such a status is that it gives people the sense that their inviolability is appropriately recognized. Naturally they're gratified by this, but the gratification is due to recognition of the value of the status, rather than the opposite-i.e., the status does not get its value from the gratification it produces...It may be that we get the full value of inviolability only if we are aware of it and it is recognized by others, but the awareness and the recognition must be of something real. <sup>153</sup>

One final thought on the foundations of the Nagelian account: note that on this picture, allowing the government to interfere with and abuse one person, implicates the political status of all similarly placed persons within that society. For instance, in a society where a lottery is run and those whose numbers are selected are killed for sport by the government, the political status of all persons would be diminished. For even those lucky enough to never have their number picked, would nonetheless still belong to a group that is liable to be abused and sacrificed in this way. Nagel's message is that our political status is determined not merely by what the government does to us, but by what the government may do to us.<sup>154</sup> Indeed, so understood, even if the government never in fact killed anyone for sport, it might still plausibly undermine the political status of its citizens by merely indicating that under certain circumstances (say an extreme lack of entertainment in society), it would be willing to do so. For in this society, this mere willingness on the government's part to sacrifice citizens for sport, indicates that citizens possess the political status of beings that can be used and abused.

---

<sup>153</sup> Nagel, 'Personal rights and public space', p.93.

<sup>154</sup> "It is true that a right may sometimes forbid us to do something that would minimize its violation-as when we are forbidden to kill one innocent person even to prevent two other innocents from being killed. But the alternative possibility differs from this one not only in the numbers of innocents killed. If there is no such right, and it is permissible to kill the one to save the two, that implies a profound difference in the status of everyone not only of the one who is killed. For in the absence of such a right, no one is inviolable: Anyone may be killed if that would serve to minimize the number of killings. This difference of status holds true of everyone whether or not the situation will ever arise for him". Nagel, 'Personal rights and public space', p.90.

So that's a rough sketch of the foundation of Nagel's account. But what has this got to do with censorship and persuasion-control? The answer is that having reflected upon the elevated moral status of persons and the kinds of claims that this establishes, Nagel now shifts his focus towards our status as independent, thinking beings and what kinds of claims this establishes. While our general moral status grants us inviolability with regards to the sacrifice of life, Nagel argues that our status as independent thinking beings grants us a kind of intellectual inviolability that rules out persuasion-control.

That the expression of what one thinks and feels should be overwhelmingly one's own business, subject to restriction only when clearly necessary to prevent serious harms distinct from the expression itself, is a condition of being an independent thinking being. It is a form of moral recognition that you have a mind of your own...The sovereignty of each person's reason over his own beliefs and values requires that he be permitted to express them, expose them to the reactions of others, and defend them against objections. It also requires that he not be protected against exposure to views or arguments that might influence him in ways others deem pernicious, but that he have the responsibility to make up his own mind about whether to accept or reject them.<sup>155</sup>

For Nagel, the problem with persuasion-control is a problem of recognition and status. Like Scanlon, he is concerned with the capacity persons possess to judge for themselves the persuasiveness of various arguments for and against different beliefs and intentions. The government that refuses to engage in persuasion-control and establishes a stringent legal right against such content-based censorship, treats its citizens in line with their (purported) metaphysical status as independent, responsible, thinking beings. By steadfastly ruling out persuasion-control even in those cases where it might prevent harms or benefit the populous, the government recognises that its citizens possess, by nature, a powerful claim against other telling them what to think. Again, there is the hint of the sublime present in this conception of

---

<sup>155</sup> Nagel, 'Personal rights and public space', p.96.

persons as having a deep-seated claim to sovereignty over their own minds, even when material considerations tell against it.

By contrast, the political system that permits the government to engage in persuasion-control, bestows on citizens a rather lower political status. By permitting persuasion-control, each citizen is thereby relegated to the lower echelon of creature that only sometimes gets to be sovereign over their own minds; forcibly stripped of sovereignty in this area when the material considerations call for it. The implications of all of this are subtle, and worth drawing out with care: even if we agree, *arguendo*, that persuasion-control is sometimes justified in order to prevent serious harms from befalling innocent persons (and to be clear, Nagel is reluctant to even admit this much), the Nagelian picture still holds that permitting persuasion-control nonetheless comes at the significant cost of demeaning our political status within that society. A democratic system where the government routinely engages in persuasion-control may well be a safer place to reside. But we who reside in this society will no longer be quite the same sublime thinking creatures we sometimes imagine ourselves to be.

Before canvassing some worries with all of this, let us say that there is much to like about the Nagelian picture. Nagel begins by suggesting that many, including himself, register a deep intuitive unease with censorious interferences. His account is in an attempt to unpack and situate these intuitions. Notice however, that Nagel deploys a striking argumentative strategy here, reflecting not so much on the interference itself (e.g. its harmful or disruptive qualities), but rather on the nature of the persons being interfered with. Of course, Nagel's analysis of persons has implications for how we understand things like the forcible sacrifice of life in the name of saving others, and persuasion-control. But his starting point is always the moral qualities of the persons being interfered with in these cases.

Furthermore, while it remains to be seen whether Nagel's suggestions about status provide quite the right way of understanding persuasion-control in particular, we grant that these reflections on moral status ought to have at least some place in our analysis of political systems. For instance, these reflections on moral status go a long way to explicating the steadfast opposition many have toward permitting torture in the name of national security. The awful dehumanisation involved in willfully torturing someone – the humiliation and pain inflicted – strike many commentators as not just a case of terrible harm, but also a fundamental failure to recognise the moral status of the victims involved. It is for this reason that so many of

us find torture abhorrent and unacceptable, even in those cases where it might secure an overall net benefit for society.

Still, we wish to discuss two fundamental problems with Nagel's account. The first objection involves taking a closer look at the mechanics of Nagel's objection persuasion-control. It is argued that Nagel's key argumentative move – that persuasion-control is bad because it bestows on citizens a violable political status – can be interpreted in two subtly different ways. Unfortunately for Nagel, neither is promising. The second objection pushes the thought that while interfering with the intellectual autonomy of persons is generally status-diminishing, it is not so when it is necessary to prevent harm.

That first objection, then: the key move in Nagel's argument against permitting persuasion-control, is his suggestion that persuasion-control objectionably bestows on citizens the political status of intellectually violable beings. But why is this supposed to be objectionable? How precisely should we understand Nagel's complaint here? The 1995 paper touches upon at least two different ways of reading Nagel. The first reading construes Nagel as objecting to the way persuasion-control bestows on citizens a lower political status than they might otherwise have enjoyed. That is to say, given that citizens might have otherwise enjoyed the political status (and concomitant liberties and legal rights) of intellectually inviolable beings, we have reason to lament the government's decision to permit persuasion-control and bestow on citizens the lower political status of beings that may have their intellectual autonomy undermined. On the first reading then, it is the loss of status involved with permitting persuasion-control that is objectionable.

The main problem with this first reading is that, even granting that permitting persuasion-control changes our political status, it is doubtful that this change involves a significant loss of status, if indeed it involves a loss at all. Three points in particular are worth emphasising. First, if there is a loss of status involved in the move from intellectual inviolability to intellectual violability, it is likely a subtle one. The government that engages in precisely worded, narrowly framed speech regulation, need not suppose that their citizens may have their intellectual autonomy undermined willy-nilly. On the contrary, they may hold that citizens generally have a strong claim to exercise their intellectual autonomy as they see fit. It is just that this government also supposes that there are certain select cases where the harms involved are such that interferences with the intellectual autonomy of citizens, while regrettable, are

nonetheless all-things-considered justified. Whatever else we might say about this change then, it is a subtle one, and may not be able to support the powerful objection to persuasion-control that the likes of Nagel wish to establish.

Of course, Nagel might reply that when it comes to losing inviolability, there are no subtle changes. Here the thought is that losing inviolability always involves a substantial loss. But this kind of argument works best when made in the context of our having an inviolable claim against torture, or an inviolable claim against being made to sacrifice our life for others. It functions less well when made in the context of whether we can sometimes have our access to dangerous persuasive arguments partly blocked. Construing agents as violable in this way does not seem to strike the same demonstrable blow to the status of citizens, as say permitting torture might. Indeed, we can strengthen this point by reflecting on how even the likes of Nagel will admit of some acceptable cases of interference with our intellectual autonomy e.g. in the form of suppressing incitement to imminent violence.

Second, while permitting persuasion-control may subtly change our political status, it is arguable that this is a lateral, rather than downward, move. In his response to Kamm's suggestion that the status of all persons would be degraded were the government to permit persons to be killed in order to save the lives of others, Shelly Kagan argued that a decreased level of inviolability simultaneously secures for each agent an increased level of 'saveability'.<sup>156</sup> In this sense, so the response goes, permitting persons to be sacrificed does not necessarily diminish or degrade the status of citizens. For whatever loss of status is experienced as a result of being treated as a being that can sometimes be sacrificed, is made-up for by being treated as a citizen with an increased claim to being saved.

An analogous argument can be made concerning persuasion-control. Permitting persuasion-control may well treat citizens as intellectually violable. But this decreased inviolability secures an increase in one's protection from the state, and the claims one has against others engaging with arguments and ideas that might endanger you. The point here is not a consequentialist one to do with balancing the benefits and burdens of persuasion-control. It is that bestowing on citizens powerful legal claims against others endangering them, speaks to a certain way of valuing those citizens that enhances their status within the political

---

<sup>156</sup> Shelly Kagan (1991). "Replies to my critics", *Philosophy and Phenomenological Research* 51 (4), pp.919-28.

community. It may well be, then, that persuasion-control both diminishes and enhances our political status in various respects.

Third and perhaps most importantly, the very aspect of persuasion-control that Nagel sees as status-diminishing – the fact that it forces citizens to sacrifice their intellectual autonomy for others – may in fact itself be status enhancing. The government that engages in persuasion-control would seem to view its citizens as liable to take on burdens and sacrifices for one another, including sometimes having their intellectual autonomy interfered with by the government. The impression one gets from Nagel, is that this can only be a bad thing; that being violable functions as a liability, and that as such, to view an agent as violable is to view them as less morally valuable or less perfect than an inviolable agent.<sup>157</sup> But Nagel is mistaken.

There may in fact be something ennobling about asking that we occasionally take on burdens for our fellow citizens. Such a demand conceptualises citizens as partly responsible for one another's welfare; partaking in a co-operative political system that emphasises fraternity, solidarity, and limited duties to assistance. Of course, citizens may not be cheered when the government asks them to make sacrifices for others, such as asking that they not express or listen to certain dangerous kinds of persuasive appeal. But the key question isn't whether holding citizens as duty-bound in this way will cause cheer, but rather whether it is status-diminishing or not. Along these lines, there is nothing demeaning or degrading, and indeed there may even be something rather elevating, about viewing citizens as bearing responsibilities and duties towards one another. Fraternity may not be to everybody's liking, but it is far from a degrading view of persons.

Let us move on to the second way of understanding the key move in Nagel's argument. On this second reading, the problem with persuasion-control is not only that it bestows on citizens a lower political (i.e. intellectually violable) status than they might have otherwise enjoyed, but that this lower status is unfitting. Persons, so the thought goes, naturally possess a special kind of moral status that makes them intellectually inviolable. As such, the government

---

<sup>157</sup> Quinn also appears to share some of Nagel's worries with demanding that agents make sacrifices for one another. Quinn cashes out his answer in terms of the Kantian notion of ends: "the doctrine reflects a Kantian ideal of human community and interaction. Each person is to be treated, so far as possible, as existing only for purposes that he can share. This ideal is given one natural expression in the language of rights. People have a strong prima facie right not to be sacrificed in strategic roles over which they have no say. They have a right not to be pressed, in apparent violation of their prior rights, into the service of other people's purposes". Quinn, 'Actions, intentions, and consequences: the doctrine of doing and allowing', p.350-1.

fails to properly recognise its citizens and their true moral status, when it bestows upon political rights that leave them intellectually violable. It fails, in other words, to grant citizens the kind of elevated political status that is appropriate and right for beings like us.

Exegetically speaking, this second reading better reflects Nagel's settled thoughts on the topic of censorship and persuasion-control, and indeed echoes the argumentative strategies we find in Nagel's chief inspirations (Kamm and Quinn).<sup>158</sup> The main problem with this second reading, however, is that it boils down to a rather brute claim about the underlying inviolable status of persons. Nagel's account began as an attempt to contextualise and unpack a certain kind of intuitive worry many (including Nagel) register with content-based censorship like persuasion-control. Rather than leave the argument at the level of intuition, however, Nagel suggested that we can productively unpack and even help justify this intuitive response, by interpreting it as a worry about status and recognition. The problem is, at this point in Nagel's argument we now find ourselves with the more or less brute claim that persons just are the kinds of beings that are intellectually inviolable and enjoy a powerful moral claim against interferences with their intellectual autonomy. Indeed, Nagel at least, provides no further independent argument or justification in favour of this brute claim. As such, those of us who do not find ourselves naturally drawn to this striking vision of citizens as naturally intellectually inviolable, will likely find Nagel's account unappealing.

Furthermore, while some of Nagel's chief inspirations – most notably Kamm – also propose arguments that terminate in a similar kind of brute claim, it is important to note some of the key differences in the arguments presented. Kamm's reflection on status and recognition are designed to help explain why persons should not be forced to sacrifice life and limb for the greater good. To Kamm's credit, it is likely that many readers will agree with her that persons *just are* the kind of special beings that should not be forced to sacrifice their lives in the name of saving others; that forcing agents to make this kind of radical sacrifice, would be fundamentally dehumanising and wrong. By contrast, Nagel's argument concerns whether agents should have limitations placed on what kind of arguments they have access to. The change in context here is important. For while there may be something dehumanising and

---

<sup>158</sup> See, for instance: Warren Quinn (1994) *Morality and Action* (Cambridge: Cambridge University Press). On p. 173 Quinn writes: "it is not that we think it fitting to ascribe rights because we think it a good thing that rights be respected. Rather we think respect for rights a good thing precisely because we think people actually have them- and, if our account is correct, that they have them because it is fitting that they should".

degrading in having to sacrifice one's life for the greater good, we should be wary of going so far in the other direction as to hold that citizens should never take on burdens for one another, or that any kind of sacrifice is degrading.

Finally, it is helpful to bear in mind that, as discussed earlier, those who hold that persuasion-control is sometimes permissible, need not conceptualise citizens as having no claims at all against having their intellectual autonomy interfered with. Similarly, they need not conceptualise citizens in a way that is demeaning or *less than* intellectually violable agents. Indeed, there may be something rather ennobling about the conception of citizens that demands citizens sometimes sacrifice their intellectual autonomy for one another. This is important, because it means that those like Nagel who find the intellectually violable conception of citizens to be unfitting or intuitively unappealing, cannot simply claim that this conception is dismal or unattractive. On the contrary, a more fine-grained analysis is required to explain why this move from intellectual inviolability to violability is so demeaning. Unfortunately, however we read it, Nagel's account does not provide such an analysis.

Let us move now to the second objection: the key claim underpinning the Nagelian picture, is that permitting persuasion-control necessarily involves tarnishing the political status of citizens as independent, thinking beings. However, even if we were to grant that persuasion-control *can* sometimes tarnish citizens' political status, we should be wary of assuming that it *always* does so. In particular, we should be wary of assuming that it does so in cases where persuasion-control is designed to simply prevent citizens from harming others; something which they have no right to do so in the first place.

It is uncontroversial to suppose that, generally speaking, citizens have no claim to harm other citizens. Because of this, government interferences with citizens that might otherwise be objectionable and status-diminishing, are not so when they merely prevent agents from harming others. A simple example might be breaking someone's arm. In most contexts this is an outrageous interference, and one might well conclude that generally permitting citizens to be interfered with in this way would degrade their moral status and treat them as the mere plaything of the person breaking their arm. But matters are different when the 'victim' having their arm broken is intending to commit bloody murder, and breaking her arm is necessary to stop a murder taking place. In this case, the interference seems to involve no moral cost whatsoever. Why is this? A straightforward explanation would be that the government's actions

in this case merely prevent (assuming other background conditions have been met e.g. the interference is proportionate, significantly less costly but equally effective counter-measures are not available) a citizen from doing what they have no right to do in the first place (i.e. harm others).<sup>159</sup>

All of this casts a degree of doubt on the Nagelian picture. For if breaking someone's arm can be done without undermining their moral status – so long as this interference is an effective way of preventing them from harming others – then why not say the same thing about persuasion-control? Why not conclude that while interfering with the intellectual autonomy of citizens normally constitutes an assault on their status as thinking beings, this interference involves no such assault when it is necessary to prevent them from harming others? We can put this problem in the form of a dilemma for Nagel: either Nagel implausibly insists that the police cannot break someone's arm in order to prevent them from committing a harm without degrading that moral person's status, or else he concedes that this kind of interference is compatible with leaving the person's moral status untouched, but then it becomes unclear why we should not draw a similar conclusion about governmental persuasion-control. Neither option is promising.

In response to this objection, Nagel might concede that there is no moral cost involved when persuasion-control simply involves interfering with agents who would otherwise harm others. After all, such 'irresponsible thinkers' are merely interfered with in a way that stops them from doing what they have no claim to do in the first place (harm others). All the same, Nagel might respond, persuasion-control generally also limits the freedom of citizens who, if allowed to hear all the arguments, would not be moved to develop harmful intentions or beliefs or indeed cause any harm whatsoever. We are talking here about 'responsible thinkers' who, left free to hear discriminatory arguments and anti-science ideas, would reject them as the nonsense they in fact are. It is an unfortunate reality of large-scale governance that when the

---

<sup>159</sup> Although this point was devised without prior knowledge of the following article, it should be noted that it echoes one of the key insights of: Howard, 'Dangerous speech'. One particularly interesting point Howard makes, which helps strengthen this thesis' argument, is as follows: while an absence of governmental interference may well benefit John, whether he has a legitimate complaint against the government deciding to interfere and take away these benefits, rests on whether he has a moral claim to these benefits. With all of this in mind, Howard's point is that citizens have no moral claim to the benefits they accrue from harming others. That said, one key difference between this thesis and Howard' arguments, is that while Howard focuses on the benefits speakers might gain from an absence of government interference and the claims such speakers have, this thesis focuses merely on the benefits and claims of would-be listeners.

government censors persuasive appeals that it fears might inspire certain ‘irresponsible thinkers’ to behave in harmful ways, it likely also prevents responsible citizens who have done nothing wrong from accessing these persuasive appeals. In the end, the reply pushes the thought that while persuasion-control may not damage the moral status of those irresponsible thinkers who merit being interfered with, it does damage the moral status of responsible thinkers who are burdened and have done nothing to deserve this.

This reply is well taken. In response we offer the following tentative suggestion: interferences with our freedom do not degrade our political status when they form part of a proportionate and effective harm prevention strategy, even when these restrictions are imposed on agents (e.g. ‘responsible thinkers’) who themselves have done nothing wrong. As a matter of general intuition, for instance, most of us do not find it necessarily degrading to endure moderately time-consuming and inconvenient security measures, so long as such measures are an effective way of promoting our safety and the safety of others. Consider pat-downs and body scans at airports, driving tests, criminal background checks, and restrictions on the ownership of dangerous weapons. All of these safety measures impose moderate burdens on innocent agents, yet most of us tend to take these inconveniences in our stride, even finding them mundane.

We could leave this point at the level of intuition. Instead, we offer an explanation: that in a democracy, citizens have certain responsibilities to one another, and that this is part of what it means to operate as a citizen in a co-operative, democratic political system. One such responsibility is, of course, the responsibility not to personally harm others. But citizens should also comply with just government schemes, refrain from obstructing state officials from doing their job, and offer their assistance to those who they can save with the minimum of effort.<sup>160</sup>

---

<sup>160</sup> Especially when we also benefit from these same government protections and schemes, see: John Rawls (1964). “Legal obligation and the duty of fair play”, in *Law and Philosophy: A Symposium* (New York: New York University Press), ed. Sidney Hook, pp.3-18. On pp.9-10 Rawls writes: “Suppose there is a mutually beneficial and just scheme of social cooperation, and that the advantages it yields can only be obtained if everyone, or nearly everyone, cooperates. Suppose further that cooperation requires a certain sacrifice from each person, or at least involves a certain restriction... Suppose finally that the benefits produced by cooperation are, up to a certain point, free: that is, the scheme of cooperation is unstable in the sense that if any one person knows that all (or nearly all) of the others will continue to do their part, he will still be able to share a gain from the scheme even if he does not do his part. Under these conditions a person who has accepted the benefits of the scheme is bound by a duty of fair play to do his part and not to take advantage of the free benefits by not cooperating”.

All of this speaks to a responsibility to comply with government's effort to protect others, at least when the costs are low.<sup>161</sup>

Nagel could respond by doubling-down and insisting that the costs that go along with persuasion-control are not so low. To this, let us offer the following three points. First, while at this juncture in our discussion Nagel may well double-down on his anti-censorship sentiments, there is little or nothing he can offer in way of further justification. Those of us, then, who do not share his intuitive unease with what we characterise as relatively mild, well-motivated, effective interference with dangerous expression, are unlikely to be won over by this move. Second, while Nagel might argue that permitting persuasion-control is costly on account of the way it bestows on citizens a lowly political status, we have already argued above that Nagel is mistaken in this respect, and that persuasion-control may even enhance our political status in various ways. Third, were 'responsible thinkers' to refuse to co-operative with the government's effort to protect others from the harmful behaviour of irresponsible thinkers, they would thereby endanger their fellow citizens, or at least become obstructions to the safety of their fellow citizens. In this way, these responsible thinkers would become appropriate targets for government interference.

The key point to draw from our discussion of Nagel is that even innocent citizens (e.g. responsible thinkers) may well be legitimately interfered with by the government. Granted, there will be limits on what demands the government can make of its innocent citizens. Torture and forcible hypnosis, for instance, may well be so dehumanising as to be unacceptable even when they serve a noble purpose. But without further argumentation, there is no reason to assume that the censorship of dangerous speech is in the same ballpark as these kinds of radical interference. On the contrary, persuasion-control may well be just the kind of effective and proportionate interference that, far from diminishing the status of the agents involved, in fact reaffirms their status as morally responsible thinking creatures who are partaking in a co-operative democratic system in which they must occasionally make sacrifices for one another.

### §III. Recognition respect

---

<sup>161</sup> Peter Singer (1972). "Famine, affluence, and morality", *Philosophy & Public Affairs* 1 (3), pp.229-43.

To end this chapter, let us explore one final way in which governmental persuasion-control might be thought to deliver a special kind of insult. Perhaps persuasion-control is insulting insofar as it involves a failure to properly recognise that its citizens are thinking beings. Or as Darwall might put, persuasion-control involves a lack of ‘recognition respect’ on the government’s part for their citizens and their fundamental thinking capacities. Of course, this is close to the Nagelian worry canvassed about how persuasion-control might bestow on citizens an unfittingly low political status. But the worry here is not so much about status, as it is about to what extent a government can simultaneously interfere with the intellectual capacities of its citizens in as direct a way as is involved in persuasion-control, while still showing due appreciation for those citizens and their intellectual capacities.

Unlike appraisal respect, recognition respect does not involve appreciating some achievement or excellence of character on a person’s part.<sup>162</sup> Instead, it primarily involves giving a person the due consideration and respect that is owed to them simply in light of their being a person.<sup>163</sup> In other words, recognition respect involves appreciating people, not because of their achievements, but in virtue of their human capacities and qualities. It is common to hear government who engage in fascistic policies and radical, right-violating behaviours, accused of lacking ‘recognition respect’ for their citizens. Such a government, so the thought goes, fails to recognise the basic human capacities of its citizens, and how these capacities ought to inform how the government treats these citizens. But we need not reserve this objection simply for such extreme cases. Jonathan Quong, for instance, has argued that paternalism involves a failure to recognise the nature and capacities of those being paternalized.<sup>164</sup> Can a similar argument be constructed in order to problematise persuasion-control?

A relatively straightforward version of such an argument, goes as follows: persuasion-control involves a failure on the government’s part to recognise the fundamental capacity for moral assessment that its citizens possess, and the importance of their exercising this capacity

---

<sup>162</sup> Darwall, ‘Two kinds of respect’.

<sup>163</sup> We say ‘primarily’, because Darwall also thinks that recognition respect can be granted by responding appropriately to someone’s ‘presented self’. We will say no more about this aspect of recognition respect, however, as it is not pertinent to our discussion of persuasion-control.

<sup>164</sup> Jonathan Quong (2010). *Liberalism Without Perfectionism* (Oxford: Oxford University Press). Echoing Rawls, Quong argues that agents have two crucial moral powers, the second of which is the “capacity to form, revise, and rationally pursue [one’s] conception of the good” (p.2). The problem with paternalism, so Quong suggests, is that it treats agents as though they lack this second Rawlsian power; it treats them as though they lack this capacity.

free from outside interference. In this way, persuasion-control involves a lack of recognition-respect insofar as it overlooks the fact that citizens are thinking beings who are capable of arriving at their conclusions about what to think. In sum, the government that engages in persuasion-control may have the best of intentions and aim only to prevent innocent persons being harmed. But its pursuit of these aims betrays a lack of recognition and respect for the fundamental intellectual capacities of citizens. Indeed, this species of argument underpins Ronald Dworkin's widely cited suggestion that the government "insults its citizens...when it decrees that they cannot be trusted to hear opinions that might persuade them to dangerous or offensive convictions".<sup>165</sup>

This straightforward version of the recognition respect worry is unlikely to win many admirers. Its key claim is that persuasion-control involves a failure on the government's part to recognise that its citizens, including those it is subjecting to persuasion-control, are thinking beings that possess the capacity to assess for themselves what kind of beliefs and intentions they ought to develop. However, not only is persuasion-control compatible with the government recognising that its citizens are thinking beings, such recognition is in fact necessary for engaging in persuasion-control in the first place. After all, the reason the government engages in persuasion-control is that it is worried that citizens, left to their own devices, will be exposed to arguments and ideas that persuade them to develop harmful thoughts. This government, then, is fully aware that its citizens are capable of coming to their own conclusions about what kind of thoughts are worth having. Indeed, that's the whole problem! It is the fact that citizens have this kind of intellectual power and may use it unwisely, that explains why intervention is necessary.

In reply, one might argue that all the above really shows, is that persuasion-control involves a formal recognition of the fact that citizens have certain intellectual capacities. As such, a more sophisticated version of the recognition respect worry, pushes the thought that true recognition involves more than this. The sadistic murderer who takes special delight in slowly extinguishing the sentience of his victims, may well formally acknowledge the humanity of those he kills. Indeed, this kind of formal recognition is part of his sadistic motivation for killing (he enjoys seeing his victims' humanity extinguished!). Yet at the same time he fails to

---

<sup>165</sup> Dworkin, '*Freedom's Law: The Moral Reading of the American Constitution*', p.200.

fully recognise the moral significance of their humanity and how it is supposed to modify his behaviour. In a similar vein, while persuasion-control may well be compatible with a formal recognition of the thinking nature of citizens, it is incompatible with a fuller, richer appreciation of their thinking nature.

This is an interesting and challenging objection to governmental persuasion-control. Still, we offer four replies. The first and second of these replies point to relatively minor concerns with this objection, while the third and fourth replies canvass deeper worries. That first reply, then: the objection is vulnerable to the kind of bruteness worry we canvassed earlier when discussing Nagel. After all, the objection more or less assumes that the only proper way to appreciate an agent's intellectual capacities, is to steadfastly refuse to interfere with her intellectual autonomy. However, this is a more or less brute assumption for which we have no further independent justification.

Second, depending on how one understands what it means to recognise and appreciate our intellectual capacities, promoting our intellectual capacities in the long-run may sometimes involve interfering with those same capacities in the short-run. In some ways this is a straightforward idea; we all know, for instance, that promoting a patient's long-term health may involve giving him medicines that make him unwell in the short-term. A similar point arguably applies when it comes to our intellectual capacities. For instance, cults often encourage their followers to stop engaging with arguments and ideas that are contrary to the cult's key values. In such a case, censorship may well promote the long-term use of the follower's intellectual capacities. A different kind of case involves persuasive appeals that help promote an environment that is hostile to certain marginalised groups. This point will be explored in more detail in chapter 5, but for now it suffices to say that these kinds of environments appear to deter members of marginalised groups both from expressing themselves in public, and from engaging with popular arguments and ideas. In such a case an appreciation for the value of agents utilising their intellectual capacities may in fact establish a case in favour of governmental persuasion-control.

Third, there are several ways in which the decision procedure that lies behind persuasion-control can explicitly incorporate an appreciation for the importance of the thinking nature of citizens. For instance, consider the following plausible scenario involving the government's decision procedure regarding two different kinds of intervention: on the one hand, the

government decides that it is worth engaging in persuasion-control so long as this intervention prevents 1000 units of harm. The government acknowledges that this kind of persuasion-control is a significant blow to citizens' moral autonomy and nature as thinking beings, but also posits that preventing 1000 units of harm is important enough to justify this kind of blow. On the other hand, the government decides that it is worth engaging in other kinds of censorship (i.e. non-content-based censorship) so long as this intervention prevents merely 500 units of harm. Thus, the government would be willing to say shut down a library where there has been a spill of a toxic substance, or cancel a political demonstration following a bomb scare, so long as this will prevent 500 units of harm.

What turns on the fact that the government is employing two different decision procedures here? The fact that the government has installed a higher bar for engaging in persuasion-control, compared to other, content-neutral forms of censorship, tells us that the government thinks there is something extra problematic about persuasion-control. It may be that the government places great value on the thinking nature of its citizens and is therefore loath to engage in a form of censorship that involves 'telling citizens what to think'. This is significant, for suddenly it looks like the government's decision to engage in persuasion-control is in fact compatible with an appreciation for the thinking nature of citizens. Indeed, it may be that the government cares deeply about the intellectual independence and thinking nature of its citizens, which is precisely why it sets such a high bar for justifying this kind of interference.

Indeed, one need not even compare the government's decision procedure concerning persuasion-control with other forms of interference, to see this point. One can simply consider the government's decision procedure concerning persuasion-control in isolation. That the government is willing to sometimes engage in persuasion-control in order to prevent 1000 units of harm (or some other morally sub-optimal outcome, if one prefers not to focus exclusively on harm) does not necessarily mean that it fails to recognise that we have a capacity for moral autonomy, or that exercising this capacity is very important. The fact that it is only willing to engage in persuasion-control once doing so will prevent 1000 units of harm, speaks to a recognition of the costs of persuasion-control and the importance of the autonomy of citizens. Persuasion-control then, need not imply a failure to recognise the thinking nature of citizens. It just implies that respecting the intellectual sovereignty of citizens does not trump all other duties the government has.

Let us move now to the second major worry. At this point in our discussion, the key question we need to ask ourselves is whether the government that engages in persuasion-control, may simultaneously appreciate the intellectual capacities of their citizens? We might also ask more generally: how does such a government view its citizens, and is this vision appropriately appreciative? Let us end this chapter with a sketch of the conception of citizens and their place in a democratic political community, that might underpin a government's decision to engage in persuasion-control. The sketch draws together several critical insights touched upon elsewhere in this chapter, and has one main aim: to demonstrate that a government that engages in persuasion-control, may nonetheless be committed to an appropriately appreciative and attractive vision of its citizens.

First, the government recognises that the capacity citizens possess for intellectual sovereignty (i.e. the ability to assess arguments and ideas, and form beliefs and intentions in light of this assessment) is to some extent flawed. That is to say, the government reasonably views their citizens as liable to discharge their intellectual capacities in ways that may be unwise, affected by bias, self-defeating, liable to be a cause for regret in the future, etc., and thus conclude that citizens are liable to at least sometimes arrive at imperfect beliefs and intentions.

Second, the government appreciates that citizens have a deep interest in exercising their intellectual autonomy. But it also rejects the simplified vision of citizens as merely intellectual beings, whose only or predominant interest is in enjoying intellectual inviolability. Instead, the government embraces a more holistic vision of citizens as having a range of interests and capacities, some of which are rational, but others of which may be more accurately characterized as emotional, social, relational, physical, etc. For instance, citizens have interests and capacities that relate to how they form their beliefs and intentions, what kind of goals they pursue, whether they are able to pursue those goals, whether they are happy, how they view themselves, whether they are able to form meaningful relationships with others, whether they feel respected, whether they feel safe, whether they are healthy, whether they are in pain, and so on.

Third, whether the interests of citizens are met, and whether the capacities of citizens are developed and perfected, depends in no small part on their environment. For instance, citizens benefit greatly from being able to express themselves, as well as being able to engage with a wide variety of ideas and arguments. But co-operation is required if the interests and needs of

citizens (including their interests in exchanging arguments and ideas) are to be met on any kind of large scale. Most germane for our purposes, is that the expression of others can be both directly harmful, as well as indirectly harmful (as when, for instance, it inspires listeners to harass, discriminate, attack, etc.).

Fourth, the government views citizens as partaking in a community of moral equals with whom their claims may compete, and therefore holds that the competing claims of agents must be weighed against one another by the government that acts as a fair arbiter. Again, most germane for our purposes is that exposure to certain kinds of arguments and persuasive appeals, may inspire beliefs and actions that strike a blow against the interests of other citizens. As such, the government recognises that imposing certain limits on expression (including what kind of persuasive appeals can be made, and in what circumstances) will help protect the interests of certain citizens.

Fifth, the government views citizens as having responsibilities and duties towards one another. This includes not only having duties to personally refrain from harming others, but also co-operating with government efforts to help prevent harm and meet the needs of their fellow citizens. Of course, there is a limit on the kinds of burdens citizens can be expected to take on in the service of their fellow citizens. All the same, there will be occasions when citizens will be expected to shoulder moderate burdens for one another. This may involve, for instance, so called ‘responsible thinkers’ being denied access to certain persuasive appeals, simply the wide exposure of these persuasive appeals might move less responsible agents to perform harmful actions.

This vision of citizens and of their role in the political community is neither insulting nor unappreciative. Granted, this vision does involve holding that citizens are flawed reasoners, and that they are sometimes liable to harm one another as a result of exposure to certain arguments. But this needn’t strike us as insulting, nor a failure of recognition. On the contrary, it stems from a deeper understanding of our intellectual capacities and how they might be used. Another striking element of this vision, is that citizens are described as rather more physically, socially and emotionally complex than the solitary, thinking creature one sometimes finds in philosophical literature. Adding this kind of nuance to our vision will introduce certain complexities to our analysis of persuasion-control, some of which will be explored in much more detail in chapter 5. For now, let us emphasise simply that this more complex

description of the citizen is far from demeaning or unappreciative, and in fact may stem from a richer appreciation of the varied capacities and interests of citizens.

Finally, this vision describes citizens as having certain responsibilities and duties to one another. Citizens, so the thought goes, should sometimes shoulder burdens for their fellow citizens, even when they themselves have personally harmed no-one. Granted, requiring citizens to take on burdens in this way may not always be to the liking of all involved. But there is nothing demeaning or unappreciative in this vision – it is not demeaning to suggest that citizens have duties to one another. In the end then, persuasion-control may simply stem from an appreciation of the need for co-operation in our political system, as well as the moral importance of meeting the needs of citizens.

### Conclusion

This chapter has examined the various ways in which we can recognise and appreciate persons, and the related question of how citizens ought to be conceptualised in a democratic project. As we have seen, it is common to hear governmental restrictions on speech (especially persuasion-control) being accused of insulting citizens, or else failing to properly recognise their intellectual capacities. However, upon closer examination, these complaints fall flat. In the end it may be that persuasion-control proceeds from a suitably grounded and holistic appreciation of the varying interests, capacities, limitations, and vulnerabilities of citizens, and how these citizens cannot thrive without co-operation. In the next chapter we examine a subtly different way of sketching the conflict between persuasion-control and our democratic project. On this approach persuasion-control conflicts with our conception of the government and how it should conduct itself in a democracy.

## Chapter 4

### Sympathy for the devil

Over the past ten years and across several articles, James Weinstein has defended an account of so called ‘participatory democracy’ that forcefully problematises hate speech regulation. Weinstein argues that a good deal of hate speech regulation is viewpoint discriminatory, insofar as it illegitimately penalises citizens based on the content of their views. This kind of viewpoint discrimination, or so it is suggested, comes at a significant democratic cost. Weinstein argues that viewpoint discrimination undermines the right of each citizen to participate as a political equal in public discourse. In so doing, it allegedly impairs the legitimacy of what he terms ‘downstream laws’; laws, such as those prohibiting racial discrimination, which citizens may no longer be able to fully express their opinions on, due to viewpoint discriminatory speech regulations.

This chapter investigates whether Weinstein’s discussion of hate speech regulation might help us problematise persuasion-control. Of course, persuasion-control is not the same thing as hate speech regulation.<sup>166</sup> All the same, the chapter takes inspiration from Weinstein’s work in order to consider whether persuasion-control is objectionably viewpoint discriminatory. Several versions of the argument that persuasion-control is viewpoint discriminatory, are canvassed. Each of these arguments draws upon the idea that in a functioning democracy, the government

---

<sup>166</sup> How precisely we should define hate speech turns out to be tricky, philosophically subtle issue. Alexander Brown has notably cast doubt on some of the classic, straightforward approaches to defining hate speech. Here we implicitly assume a rather straightforward approach. Parekh’s definition, as noted by Jeffrey Howard, has the virtue of capturing the concerns of many who have defined hate speech. In his definition, hate speech is: “is directed against a specified or easily identifiable individual or, more commonly, a group of individuals based on an arbitrary or normatively irrelevant feature”; second, it “stigmatizes the target group by implicitly or explicitly ascribing to it qualities widely regarded as undesirable”; and third, it casts the “target group...as an undesirable presence and a legitimate object of hostility. See Parekh, ‘Hate speech’; Jeffrey Howard (2012). “Free speech and hate Speech”, *Annual Review of Political Science* 22 (1), pp.93-109. Alexander Brown (2017). “What is hate speech? Part 1: the myth of hate”, *Law and Philosophy* 36, pp.419-468. Alexander Brown (2017). “What is hate speech? Part 2: family resemblances”, *Law and Philosophy* 36, pp.561-613.

has special reason to refrain from viewpoint discrimination. As Weinstein puts it, an aversion to viewpoint discrimination “is not just some quaint and peculiar detail of American free speech doctrine. It is, rather, a core doctrinal rule implemented to vindicate the basic democratic commitment to formal equality and the equal moral worth of each citizen and to promote political legitimacy”.<sup>167</sup> Indeed, it is a widely held thought, central to many different democratic conceptions, that the government ought to act as a neutral and fair arbiter over the claims of its citizens. And that as such, it must categorically refrain from penalising citizens for simply having the ‘wrong’ kind of views. In this way, this chapter uses Weinstein to consider whether persuasion-control might conflict in a fundamental sense with our considered democratic vision of how the government ought to operate.<sup>168</sup> Indeed, having considered in the previous chapter whether persuasion-control conflicts with our conception of the citizens in a democracy, the chapter before us now reflects upon whether there is a conflict with our conception of the government in a democracy.

This chapter asks: when should we feel sympathy towards those who are burdened by speech regulation? The government may have good, value-neutral reasons to engage in the censorship of various kinds of dangerous speech. But what about when this ends up burdening citizens who, left free to engage with this speech, would not personally endanger others? Does the government wrong its citizens in this case? Can the government legitimately impose burdens on those who would otherwise contribute in relatively minor ways to larger, harmful systems? Or would such interference undermine our democratic vision of the government?

---

<sup>167</sup> James Weinstein (2017). “Viewpoint discrimination, hate speech, and political legitimacy: a reply”, *Constitutional Commentary* 32 (3), pp.715-82, p.750.

<sup>168</sup> Here we characterise the Weinsteinian account canvassed in this chapter, as problematising hate speech regulation and persuasion-control based on a vision of how the government ought to operate. One could, alternatively, characterise Weinstein’s critique as based on a vision of the claims of citizens (most notably the claim to participate ‘on an equal footing’ in public discourse). While none of the substantive arguments considered in this chapter stand or fall on what characterisation we offer, we take this opportunity to nonetheless defend my characterisation. In short, this chapter will argue that among the many different ways of interpreting Weinstein’s claims about hate speech regulation and viewpoint discrimination, the most faithful is as a claim about how the government ought not to let its own biases and political agenda, seep into its decision-making. Meanwhile, the most promising interpretation is as a claim about how the government should be duly anxious about singling out agents for criminal penalty based on their views, and enforcing currently existing political divisions among innocent, responsible citizens. Both of these readings are rooted in a vision of how the government ought to operate (including how it should handle the claims of citizens) as opposed to simply an account of the claims of citizens.

To this end, §I offers a basic sketch of Weinstein’s argument that the government’s decision to regulate hate speech comes at a cost in terms of legitimacy. It is argued that the key to understanding Weinstein’s argument is the claim that hate speech regulation is viewpoint discriminatory. §II seeks to clarify this claim. Having clarified this claim, §III charges that in fact the best kind of persuasion-control (and hate speech regulation) is neither viewpoint discriminatory nor unfair. §IV canvasses and rejects Weinstein’s attempts to defend his account by way of analogy with disenfranchisement. Finally, §V outlines a novel way of understanding the worry that persuasion-control is problematically viewpoint discriminatory. In particular it is argued that some cases of persuasion-control impose burdens on ‘innocent’ citizens in a way that is liable to exacerbate pre-existing problems with political alienation. And that those of us who are worried about viewpoint discrimination and political equality, are right to be anxious about these cases. The chapter ends by considering whether we can respond to these worries, either by changing the way that the government engages in this kind of censorship, or by reconceptualising the role of the government in these cases.

#### §I. A basic sketch<sup>169</sup>

At the heart of Weinstein’s work lies a concern with the ability of citizens to engage in ‘public discourse’.<sup>170</sup> What does this phrase mean precisely? Weinstein casts a broad net; public discourse, we are told, consists of speech “on matters of public concern, or, largely without respect to its subject matter, of expression in settings dedicated or essential to democratic self-governance, such as books, magazines, films, the internet, or in public forums such as the

---

<sup>169</sup> It is no mean feat outlining even a basic sketch of Weinstein’s critique of hate speech regulation. In order to do so, one must draw together his thoughts from at least five different articles; three of which were published in 2011, and two of which were published more recently in 2017. While Weinstein’s thoughts are always subtle and insightful, it is fair to say that they mature throughout these articles. It is with this in mind that we pay most attention to the two articles from 2017, which we take to reflect Weinstein’s most settled views on the topic. See: James Weinstein (2011). “Participatory democracy as the central value of American free speech doctrine”, *Virginia Law Review* 97 (3), pp.491-514; James Weinstein (2011). “Participatory democracy as the basis of American free speech doctrine: a reply”, *Virginia Law Review* 97 (3), pp.633-79; James Weinstein (2011) “Free speech and political legitimacy: a response to Ed Baker”, *Constitutional Commentary* 27 (2), pp.361-83; Weinstein, ‘Hate speech bans, democracy, and political legitimacy’; Weinstein, ‘Viewpoint discrimination, hate speech, and political legitimacy: a reply’.

<sup>170</sup> Weinstein occasionally uses ‘political process’ as a synonym (e.g. in Weinstein, ‘Hate speech bans, democracy, and political legitimacy’). We prefer to refer to ‘public discourse’ because this is the phrase Weinstein most frequently deploys.

speaker's corner of the park".<sup>171</sup> This disjunctive definition will be familiar to those acquainted with First Amendment debates and the two forms of speech that are typically interpreted to fall broadly under its protection. The first disjunct picks out what is usually referred to as speech of 'public concern'. That is, speech protected regardless of the setting in which it is uttered because its content touches upon matters of political, social, or other concern to the community.<sup>172</sup> The second disjunct picks out what is normally referred to as speech in a 'public forum', and affords protection to speech taking place in traditional public fora. Traditionally this has been interpreted to refer to parks and sidewalks, but more recently has been construed to also include the internet.<sup>173</sup>

Broad net notwithstanding, there are limitations on what counts as public discourse. Cases of speech that are directly harmful speech, do not count as public discourse, or at least go beyond being mere public discourse. In this way, Weinstein should not be read as providing a defence of blackmail, threats, incitement to imminent violence, etc. By contrast, what Weinstein does seem to think citizens have a claim to, is being free to express their views (and arguments in favour of those views) concerning political and cultural matters such as the state of society, immigration, the government, climate change, as well as more frivolous matters such as music, film, food, the best diets, beauty, and so on. Citizens should be free to engage in any kind of public discourse so long as it does not lead to imminent danger – meaning that citizens ought also to be free to engage in a good deal of hate speech as well as persuasion-control.<sup>174</sup> Indeed, Weinstein even argues that citizens should be free to express their views on these matters, even in highly abusive language.<sup>175</sup>

Here we must be careful not to overlook an important subtlety in Weinstein's account. Rather than framing his argument straightforwardly in terms of what citizens ought to have the ability to express (i.e. what counts as public discourse), Weinstein rather discusses the status

---

<sup>171</sup> Weinstein, 'Participatory democracy as the central value of American free speech doctrine', p.493.

<sup>172</sup> Chris Hoofnagle (2001). "Matters of public concern and the public university professor", *Journal of College and University Law* 27 (3), pp.669-707.

<sup>173</sup> Stephen Gey (1998). "Reopening the public forum-from sidewalks to cyberspace", *Ohio State Law Journal* 58 (5), pp.1535-1634.

<sup>174</sup> We take it that Weinstein does not support the view that citizens should be free to engage in hate speech when such speech is merely used to intimidate or threaten interlocutors.

<sup>175</sup> Weinstein, 'Viewpoint discrimination, hate speech, and political legitimacy: a reply', p.776.

and class of citizens engaging in public discourse. For instance, Weinstein states that citizens must be free to participate in public discourse as “political equals”.<sup>176</sup> He also tells us that citizens ought to be able to “participate as equals”.<sup>177</sup> Indeed, Weinstein is primarily interested in what it means for citizens to enjoy “equal participation” and “equal opportunity” when it comes to public discourse.<sup>178</sup> Taking a step back then, we see that Weinstein’s account revolves around the concepts of status and class, and in particular what it would take for citizens to engage in public discourse *as equals*.

The next step in Weinstein’s argument is to show that hate speech regulation undermines the ability of citizens to participate as equals in public discourse. Hate speech regulations typically aim to curb the spread of hatred, violence, and discrimination against vulnerable groups, or at least to protect citizens from demeaning abuse – though not all cases fit neatly into this description. Weinstein’s account aims to show that hate speech regulation involves an illegitimate interference with the political equality of citizens. That is, that somehow such well-meaning regulation might in fact undermine our political values. Before we get to the highly vexed issue of why Weinstein thinks this, let us first consider what Weinstein thinks the consequences are of such illegitimate hate speech regulation. As one might expect, Weinstein thinks that this kind of regulation incurs a cost in terms of legitimacy. But his thoughts on this topic are both surprising and subtle. In particular, two points are especially worth highlighting.

The first concerns what particular *kind* of legitimacy is implicated in all of this. Perhaps somewhat surprisingly, Weinstein does not think that hate speech regulation deprives the government itself of legitimacy; he does not, for instance, call into question what he refers to as “systemic legitimacy”.<sup>179</sup> Even more surprisingly, Weinstein nowhere questions the legitimacy of the actual hate speech laws themselves that he finds so troubling. Instead, following Ronald

---

<sup>176</sup> Weinstein, ‘Hate speech bans, democracy, and political legitimacy’ e.g. pp.536, 543, 546, 569, 572.

<sup>177</sup> Weinstein, ‘Hate speech bans, democracy, and political legitimacy’, p.582; Weinstein, ‘Viewpoint discrimination, hate speech, and political legitimacy: a reply’, pp.740, 757.

<sup>178</sup> For the former, see: Weinstein, ‘Hate speech bans, democracy, and political legitimacy’, pp.537, 573; Weinstein, ‘Participatory democracy as the central value of American free speech doctrine’, pp.499, 501. For the latter, see: Weinstein, ‘Hate speech bans, democracy, and political legitimacy’, pp.536-9, 546, 563, 572, 580.

<sup>179</sup> Weinstein, ‘Hate speech bans, democracy, and political legitimacy’, pp.535, 575, 577-8; Weinstein, ‘Viewpoint discrimination, hate speech, and political legitimacy: a reply’, pp.738, 751, 776, 780.

Dworkin, we are introduced to a distinction between so called ‘upstream’ and ‘downstream’ laws.<sup>180</sup> Downstream laws are enacted as a result of the political process and public discourse. These are the laws adopted following discussion and argument about what kind of regulation is appropriate. By contrast, upstream laws are those that affect the political process and public discourse. The key conclusion that Weinstein wants to draw is that citizens who are constrained by upstream hate speech regulation from participating as equals in discussions surrounding discrimination, are not bound by downstream anti-discrimination laws in the same way as other citizens who are free to express themselves.<sup>181</sup> The underlying logic here is that when hate speech regulation excludes citizens from discussing a given law as political equals, that law is thereby deprived of legitimacy with regards to those citizens, and they are no longer under the same kind of obligation to obey that law.

The second point involves clarifying what we mean when we say that a downstream law is deprived of legitimacy. At several points Weinstein declares quite forthrightly, and in-keeping with classic views on First Amendment jurisprudence, that upstream censorship deprives downstream laws of legitimacy in a way that cannot be justified or put right.<sup>182</sup> At other times however, a more conciliatory tone is struck, with Weinstein open to the idea that hate speech regulation ‘might in some circumstance be justified’.<sup>183</sup> Are these views contradictory? No. Here’s why: underlying both these claims is a distinction between ‘political legitimacy’ and what we might call ‘overall legitimacy’. Political legitimacy refers to the legitimacy and binding force a law has in virtue of being issued from a valid political arrangement; or one might say, the legitimacy of a law *qua* law.<sup>184</sup> By contrast, overall legitimacy concerns whether a law is all things considered permissible to enforce, and involves surveying a whole range of factors, including whether it is politically legitimate, but also e.g. whether the law is morally justified. This distinction is crucial for understanding Weinstein’s thoughts on legitimacy. When

---

<sup>180</sup> Dworkin, ‘Foreword’.

<sup>181</sup> Here we focus on Weinstein’s claim that hate speech regulation restricts the ability of citizens to express their views on the matter of discrimination. We focus here on discrimination, because Weinstein focuses on this example. However, one might also imagine that hate speech regulation and persuasion-control similarly restrict the ability of citizens to express their views on immigration, medical and health matters, sexual consent, how to respond to a government that you oppose, and so on.

<sup>182</sup> David O’Brien (2010). *Congress Shall Make No Law: The First Amendment, Unprotected Expression, and the U.S. Supreme Court* (New York: Rowman & Littlefield Publishers).

<sup>183</sup> Weinstein, ‘Viewpoint discrimination, hate speech, and political legitimacy: a reply’, p.732

<sup>184</sup> Weinstein, ‘Hate speech bans, democracy, and political legitimacy’, p.539.

Weinstein argues that upstream censorship deprives downstream anti-discrimination laws of legitimacy, what he is primarily referring to is political legitimacy – leaving open the possibility that these laws are still overall legitimate. After all, even if one has no political obligation to obey a law, one can still be morally obligated to comply with its demands. Or perhaps we should say, one may still be morally obliged to act in a way that happens to correspond with what the law demands.

Tying all of this together, Weinstein’s view is that all citizens have a right to participate as equals in public discourse. Hate speech regulation undermines this right, poisoning public discourse in the sense of undermining the equal status of certain citizens. In doing so, it destroys the political legitimacy of any laws that lie downstream of that now poisoned public discourse. The consequences of all of this vary in terms of overall legitimacy. Sometimes this blow to the political legitimacy of a downstream law will destroy the overall legitimacy of that law.<sup>185</sup> At other times, the laws in question may still nevertheless retain their overall legitimacy. Still, even in this case Weinstein agrees with Dworkin that there is something to regret about the government’s actions, which now operate at a democratic deficit.

Having sketched out most aspects of Weinstein’s account, we now need only consider why he finds hate speech regulation to be so deeply problematic. In other words, what is it about hate speech regulation that leads Weinstein to charge it with being so damaging to political legitimacy? Regarding this matter, Weinstein’s key claim is that hate speech regulation amounts to “*viewpoint discrimination*” – represented in premise four of the standard form argument below.<sup>186</sup> Viewpoint discrimination, roughly, describes the regulation of persons based on the content of their views. Weinstein alleges that hate speech regulation penalises persons based on the content of their views. But in what sense does hate speech regulation (or persuasion-control) do this? The answer is not immediately clear. What’s more, when we take a step back and survey Weinstein’s entire argument, we see that this is not only the most

---

<sup>185</sup> Weinstein suggests that there will be at least some downstream laws where the issue of overall legitimacy rests delicately on a knife’s edge; there are very good reasons both in favour and against the law, and whether the law is politically legitimate may just tip the scales. In cases like these, upstream censorship may well destroy both the political and overall legitimacy of the law. See Weinstein, ‘Hate speech bans, democracy, and political legitimacy’, p.567.

<sup>186</sup> Weinstein, ‘Hate speech bans, democracy, and political legitimacy’, pp.528, 545, 546; Weinstein, Viewpoint discrimination, hate speech, and political legitimacy: a reply’, pp.748, 778.

complicated of his claims, but also the most important. For all its undoubted subtlety and nuance, Weinstein's core critical contention is that certain kinds of speech regulation (e.g. hate speech regulation) are viewpoint discriminatory and hence incompatible with true equality in public discourse.<sup>187</sup> Now we must ask: can this claim be cashed out in a way that is convincing?

- i) *All citizens must be free to participate as equals in public discourse.*
- ii) *If a certain piece of (upstream) regulation undermines the freedom of citizens to participate as equals in public discourse, then the political legitimacy of the relevant downstream regulation will be undermined.*
- iii) *Regulation that is viewpoint discriminatory undermines the freedom of citizens to participate as equals.*

---

<sup>187</sup> Sometimes it is suggested that hate speech regulation does not in fact impose significant burdens. On the contrary, so the thought goes, hate speech regulation merely imposes stylistic restrictions on what can be expressed. As such, so the thought concludes, hate speech regulation is not problematic, and Weinstein is rather making a fuss over nothing. In response to this worry, three points will suffice.

First, when it comes to expression, sometimes style is substance. Take the expression "fuck the draft", the focus of at least one important constitutional dispute. Could the basic propositional content of this phrase be expressed in a less obscene, less vituperative manner? Perhaps. But we should be wary of jumping to a conclusion here. There is more to communication than the basic propositional content and even merely stylistic restriction on obscene language are liable to restrict speakers from being able to express their thoughts and feelings fully. The use of obscene and vitriolic language, for instance, often tells us something about the speaker's feelings and beliefs. It might tell us that he is sick of highfalutin debate, or that he is in a state of immense agitation. It might tell us that he wishes to be provocative. Indeed in some cases, and this seems especially pertinent for hate speech, the use of language that is commonly considered beyond the pale may tell us that the speaker wishes to cease to participate in polite debate and instead wishes to engage in something more sinister. These are all differences in the content of the speech in question.

Second, it may well be that hate speech regulation restricts even some messages that are conveyed in reasonably temperate, measured tones. Think of a speaker who implores his crowd to beware of Muslims. He declares that they are nothing more than a group of criminals, describes how they constitute an immediate threat to their safety, and finally urges that because the state has proven ineffective in securing its citizens against this threat, the time has come for good people to take action into their own hands. This speech may well count as hate speech despite being absent of explicitly vituperative or obscene language. Of course, this will depend on how we define hate speech. But typically hate speech has been defined so as to pick out speech that aims to or has the effect of inciting hatred against a group of persons, or that constitutes a kind of group libel against persons. The speech just mentioned would qualify as hate speech on either definition.

Third, even if we leave these points aside and concede *arguendo* that hate speech regulation restricts only style, this hardly means that said regulation is unobjectionable. After all, stylistic restrictions can be objectionable too. It may be, for instance, that these stylistic restrictions lack proper justification, or are being imposed in some morally arbitrary way. To give a pertinent example, one might ask why hate speech regulation imposes restrictions (stylistic or otherwise) on the communication of certain views, but not others? As we will see in §3, Weinstein's objection to hate speech regulation builds on something close to this worry. See *Cohen v. California*, 403 U.S. 15 (1971). See also Weinstein, 'Viewpoint discrimination, hate speech, and political legitimacy: a reply', p.748.

- iv) Hate speech regulation is viewpoint discriminatory.*
  
- C) Hate speech regulation therefore undermines the freedom of citizens to participate as equals in public discourse, and undermines the political legitimacy of the relevant downstream laws.*

## §II. Analysing Weinstein's key claim

Having clarified *what* the key claim is in Weinstein's account – premise four in the standard form argument canvassed above – this section asks whether that key claim can be cashed-out in a philosophically attractive way. What does it mean to label hate speech regulation, or persuasion-control, as viewpoint discriminatory? And does this really show that such regulation is democratically damaging? In this section we consider these questions by canvassing a number of ways in which we can interpret the claim that hate speech regulation and persuasion-control are viewpoint discriminatory. As will become clear, some of the more straightforward ways of interpreting this claim turn out to fall rather flat upon closer examination.

Perhaps the most straightforward way of reading Weinstein's key claim, is as a claim about how citizens *just have* a right to express their political views, as well as hear the views of others, and that any interference with this right is an unacceptable act of viewpoint discrimination. However, we suspect that few will be willing to accept such a brute claim. Certainly, it is unlikely to be dialectically efficacious when it comes to those readers who do not already hold such a view. Moreover, the brute reading leaves unexplained why the equal political status of citizens depends on *this right* in particular, as opposed to say another right that citizens have, or say the conditions in which citizens live. We do not think, then, that we should be quick to accept the brute reading. Instead, we should rather see if we can cash-out such a right by looking at the interests and claims of citizens.

An alternative reading involves arguing that a law can only be democratically legitimate if citizens have had a chance to freely debate, criticise, and argue for change of that law. The idea here is that the political legitimacy of a law depends on that law being 'the will of the people', or at least having flowed from debates and movements that citizens have had a chance to

participate in. On this reading, the reason why Weinstein judges hate speech regulation to be viewpoint discriminatory and a violation of the equal political rights of citizens, is that it excludes certain citizens from being able to take part in these legitimacy-lending processes.

This is an interesting reading, and it is certainly present in some parts of Weinstein's work. For instance, Weinstein approvingly quotes Judge Learned Hand's suggestion that "public opinion...is the final source of government in a democratic state", and goes on to suggest that any person excluded from public discourse may rightly view the government as no longer a democracy but rather an "illegitimate autocracy" over him.<sup>188</sup> All the same, we want to raise four potential problems with this reading.

First, the suggestion that laws only qualify as politically legitimate so long as citizens have had a chance to participate in public discourse surrounding those laws, is a controversial claim. Why not rather hold, in the mould of Raz for instance, that the legitimacy of laws depends on whether they serve the interests of citizens?<sup>189</sup> Or whether they reflect the right kind of ideal of the citizen?<sup>190</sup> Or whether citizens could hypothetically consent to these laws? Without further substantive argumentation, which Weinstein does not provide, it is unclear why we should opt for such a controversial thesis.

Second, contrary to this reading, at least some laws ought to be established and obeyed, regardless of what citizens think and what kinds of opinions are expressed in public discourse. In particular we have in mind laws that protect the fundamental rights of citizens, such as laws that guarantee the right of all citizens to a fair trial, or laws that protect citizens from religious persecution and racial discrimination. Such laws ought to form the foundation of any functioning democracy that is committed to the equality of citizens, regardless of whether the (e.g. racist) majority vehemently disagrees with them. Now in fairness, it might be suggested that even these laws, while already justified in their own right, would nonetheless gain political legitimacy from being discussed and debated. But if this debate can never have any impact on legal proceeding, then suddenly it looks as though all this talk of the right to equality in public discourse is somewhat overblown, and the framing of our status in public discourse as the key measure of democratic legitimacy rather unjustified.

---

<sup>188</sup> Weinstein, 'Participatory democracy as the central value of American free speech doctrine', p.498.

<sup>189</sup> Joseph Raz (1986). *The Morality of Freedom* (Oxford: Clarendon Press).

<sup>190</sup> Rawls, 'A Theory of Justice'.

Third, plausibly the best kind of hate speech laws do not limit the ability of citizens to participate in serious and robust debate about the pros and cons of various laws. Granted, this kind of regulation will generally place some limits on the ability of citizens to discuss these laws. For instance, citizens may be forbidden from encouraging others to express their disdain for a law by murdering scores of innocents. Similarly, the law may prevent citizens from attempting to persuade others to violently attack innocent members of minority groups as a way of expressing their disagreement with the law. All the same, these limitations surely leave citizens with ample space to engage in serious critical discussion regarding whether a law should be passed, abolished, obeyed, disobeyed, is flawed, is ineffective, etc.

Fourth and finally, lots of laws prevent citizens from being able to fully engage in public discourse. For instance, incarcerating murderers surely imposes upon those persons significant obstacles to their participating in public discourse (it is a lot harder to transmit your views to the wider public and listen to the views of others, when you spend most of your day in a cell). Why does Weinstein consider this kind of interference with public discourse to be acceptable, but not hate speech regulation? Clearly some interferences with public discourse may be permitted in a functioning democracy, so why not count hate speech regulations and persuasion-control among them? The reading before us does not present a convincing answer.

Moving on to a different reading then: it might be suggested that hate speech regulation is viewpoint discriminatory because of the way it imposes burdens unevenly along the political spectrum. The thought here is that since hate speech regulation (and quite possibly persuasion-control) is more likely to limit the freedom of those on the political right wing to express their views compared to the left, such regulation is therefore viewpoint discriminatory in its consequences. Let us state from the beginning that the empirical aspect of this reading strikes us as generally plausible. It may well be that hate speech regulation (and indeed persuasion-control) tend to more heavily penalise thinkers on the right side of the political spectrum, given, for instance, that these are the persons most likely to wish to present and listen to arguments in favour of e.g. identifying threats posed by immigration, identifying threats posed by vulnerable or minority groups, scepticism towards climate change, scepticism towards medical advice, etc.

Still, the problem with the reading before us is that, generally speaking, we do not object to reasonable and fair regulation just because it imposes burdens unequally along the political

spectrum. It might turn out, for instance, that the criminal regulation of tax evasion disproportionately burdens Republican compared to Democrat voters – or vice versa. However, assuming that tax evaders are prosecuted in a fair and just manner, virtually all will agree that such regulatory measures are uncontroversial. Moreover, an additional problem with this reading is that one could just as well claim that an absence of hate speech regulation tends to favour right wing thinkers. So understood, virtually all regulation, or lack of regulation, will end up favouring one side of the political spectrum more than another. Are we to conclude that virtually all regulation is therefore viewpoint discriminatory, or a threat to political legitimacy? A more palatable option is to simply reject the reading before us.

Having canvassed and quickly rejected a few different readings, let us now consider a more promising way of interpreting Weinstein. What is needed at this point is an explanation of how hate speech regulation is normatively special. That is, we need to hear more about how hate speech regulation involves some additional ingredient that separates it from other forms of regulation that have ostensibly similar effects (e.g. the incarceration of murderers, the regulation of tax evaders). With this in mind, consider the suggestion that what marks out hate speech regulation as viewpoint discriminatory is how the content of the regulation itself is such that it explicitly favours and protects certain views and arguments, while disfavours and excluding others.

This suggestion is certainly present in Weinstein's work. For instance, Weinstein suggests that "unlike a ban on fighting words or profanity or the other restrictions on harmful speech...hate speech bans are inherently viewpoint discriminatory. Britain's hate speech law, for instance, restricts only speech that intends to "stir up racial hatred" but not expression promoting racial tolerance".<sup>191</sup> Similarly, Weinstein asks us to: "[c]ompare, for instance, a law that forbids anyone from protesting on the sidewalk in front of the United States Supreme Court with one allowing protests except for those against the Court's decisions legalizing abortion. Under both laws anti-abortion demonstrators are...being prevented from saying what [they] want to say...But it cannot be seriously gainsaid that the viewpoint-based law infringes core democratic equality concerns far more than the more extensive yet viewpoint-neutral law".<sup>192</sup>

---

<sup>191</sup> Weinstein, 'Hate speech bans, democracy, and political legitimacy', p.545.

<sup>192</sup> Weinstein, 'Viewpoint discrimination, hate speech, and political legitimacy: a reply', p.750.

Here the key idea being expressed is that hate speech regulation, far from incidentally favouring certain political classes over others or happening to exclude certain viewpoints but not others, is in fact explicitly designed in such a way that excludes certain kinds of views and arguments. It is this selective aspect of the censorship – the way it picks and chooses which ideas are too dangerous or too offensive – that Weinstein finds objectionable, and synonymous with his accusation of viewpoint discrimination.<sup>193</sup>

Here at last, then, we get to Weinstein’s substantive objection to hate speech regulation. Above all, Weinstein thinks that the government should not produce regulation that picks out certain arguments and views to exclude them from public discourse. That is, the government should not be in the business of deciding which arguments and ideas are unsuitable for public discourse. Note that this argument tells not only against hate speech regulation, but against persuasion-control too. At its core, persuasion-control involves the government picking out and censoring arguments that it considers too dangerous for public consumption. Persuasion-control must, if it is to be duly selective and piecemeal, make discriminating judgements about what kinds of ideas and arguments pose a threat, and what kinds can be tolerated. It is precisely such selection and discrimination that is being problematised here – presumably in favour of a markedly less selective approach that, say, simply prohibits directly harmful speech and leaves other speech acts unregulated. It is therefore not surprising to find certain passages from Weinstein that seem almost tailor-made to problematise persuasion-control: “[g]overnment may not restrict speech because it fears that the expression of certain viewpoints will generate public opinion that will lead to misguided or even disastrous social policy decisions.”<sup>194</sup>

Finally, note that this objection to governments ‘picking and choosing’ which arguments may feature in public discourse and which are too dangerous, relates to Weinstein’s contempt for ‘ideological’ governmental decision-making.<sup>195</sup> Weinstein, for instance, suggests that the government acts “based on ideology” when it judges that certain views and arguments

---

<sup>193</sup> Weinstein, ‘Viewpoint discrimination, hate speech, and political legitimacy: a reply’, p.529.

<sup>194</sup> Weinstein, ‘Participatory democracy as the central value of American free speech doctrine’, p.498.

<sup>195</sup> The reference to ideology comes from Weinstein, ‘Viewpoint discrimination, hate speech, and political legitimacy: a reply’, p.725, as well as several footnotes in Weinstein, ‘Hate speech bans, democracy, and political legitimacy’.

are too disturbing or offensive.<sup>196</sup> Similarly, he states that the reason hate speech bans are illegitimate is because they are grounded in a “particular ideological position”.<sup>197</sup> The thought here seems to be, and this is a thought one often hears in modern censorship debates, that the government’s decision to exclude some but not other views, to count some views as too dangerous but not others, reflects a deep-seated liberal agenda that is inherently biased against certain viewpoints. Or at any rate, involves drawing arbitrary lines between what is and is not too dangerous to tolerate. Such concerns are often voiced, for instance, in the context of debates about the regulation of hate speech, as well as debates over anti-vaccination arguments and climate change denialism.

### §III. Two objections

Weinstein’s critique of hate speech regulation – and indeed persuasion-control – stems from the view that these regulations are viewpoint discriminatory insofar as they involve picking out and excluding certain ideas and arguments, because they fail to fit in with the government’s liberal ideology. But is this account compelling? Does hate speech regulation (and persuasion-control) really therefore come at a cost to political legitimacy? I’m not so sure.

Below we outline two specific objections to this account, but in general our scepticism can be summed up as follows: so long as we think that certain kinds of speech (e.g. hate speech and some kinds of persuasion) do in fact present a significant danger to innocent persons, and that regulation can play an important role in neutralising this danger, we are unlikely to share Weinstein’s worries with the government selectively picking out these dangerous forms of speech and regulating them. After all, the government regularly picks out certain kinds of actions as being especially dangerous and worthy of regulation. Why should it not engage in the same process when it comes to engaging in and listening to speech? Indeed, some degree of selectivity is required for the regulation in question to avoid the pitfalls of being overly broad and imposing burdens unnecessarily on citizens.

With this worry in mind, Weinstein’s account probably works best when paired with a scepticism regarding the effectiveness of hate speech regulation, along the lines of that

---

<sup>196</sup>Weinstein, ‘Participatory democracy as the central value of American free speech doctrine’, p.448.

<sup>197</sup>Weinstein, ‘Viewpoint discrimination, hate speech, and political legitimacy: a reply’, p.748.

expressed by Eric Heinze.<sup>198 199</sup> Heinze argues that in “longstanding, stable and prosperous democracies” (or LSPDs) like the contemporary U.K. and U.S., hate speech regulations are an ineffective and burdensome means of trying to reduce discrimination and other kinds of attacks on vulnerable groups.<sup>200</sup> He argues, for instance, that even leaving aside hate speech regulation, LSPDs are able to provide vulnerable groups with the direct protection they need from physical attack, and that such advanced democracies have a wide range of alternative tools which can be used in the more long-term fight against hatred. Heinze also uses data (admittedly, in a somewhat fast and loose manner) to suggest that hate speech regulations have not helped to reduce crimes against minority groups, perhaps most notably pointing out that the U.K. and U.S. have roughly similar incidences of hate-based attacks, despite the former’s more censorious policies. Allied with this kind of scepticism, it makes sense for Weinstein to suggest that hate speech regulation (and persuasion-control) not only excludes citizens from equally participating in public discourse, but does so on viewpoint-discriminatory, ideology-based reasons. After all, if regulation is indeed ineffective, what else could the government’s decision to regulate hateful speech stem from, other than an ideological dislike of that speech (or perhaps a misplaced optimism about the effectiveness of regulation)?<sup>201</sup>

However, there are problems with relying on the Heinzean position. Most importantly, more investigation is needed to discover whether Heinze is right to think that these empirical conditions do in fact hold. Weinstein provides us with next to no investigation of this matter. By contrast, as indicated above, Heinze does at least attempt to shed light on this empirical question. But his handling of this matter, and particular his use of empirical data, is far from satisfying. For instance, Heinze focuses on the rather narrow issue of whether hate speech regulation helps prevent discrimination and hate crime (i.e. *assault on a person or damage of their property*, motivated in whole or in part by an offender’s bias against a race, religion, disability, sexual orientation, ethnicity, gender, or gender identity). But as discussed in the second chapter

---

<sup>198</sup> Eric Heinze (2016). *Hate Speech and Democratic Citizenship* (Oxford: Oxford University Press).

<sup>199</sup> We thank Gavin Phillipson for this point.

<sup>200</sup> Here we discuss the issue of whether hate speech inspires hatred and leads to an increase in hate crime and discrimination. Heinze argues that it does not lead to such an increase. However, what we do not discuss, and what Heinze does not contest, is that face-to-face hate speech may be harmful qua amounting to harassment, intimidation, etc.

<sup>201</sup> See, for instance, Weinstein, ‘Viewpoint discrimination, hate speech, and political legitimacy: a reply’, p.579. Here Weinstein responds to Alexander Brown by arguing that: in the absence of “persuasive evidence” for banning racist speech, the suspicion arises that the true motivation for such a law is abhorrence of racist ideology.

of this thesis, many have argued that hate speech is connected with a range of more subtle harms and barriers to autonomous flourishing, that go rather overlooked by Heinze's empirical investigations. Take the suggestion that hate speech can promote contemptuous attitudes and negative stereotypes surrounding certain groups, and that in this way these groups come to be denied a very basic form of respect from their societal peers. If hate speech does in fact work in this way, this strikes us as a significant harm that at the very least establishes a strong prima facie case for regulation. Heinze's analysis is all the worse for overlooking these kinds of issues.

Moreover, scepticism is the only reasonable response to Heinze's attempts to draw conclusions about the effectiveness of hate speech regulation via comparing hate crime incidence in the U.S. and U.K. Heinze's analysis largely overlooks the vast array of cultural, legal, and historical differences between these countries, as well as how hate speech regulation can come in a variety of different forms (e.g. the regulation of online hate speech verses hate speech in physical spaces, the regulation of hate speech by normal citizens verses the regulation of hate speech by political figures), some of which might be more effective than others when it comes to curbing hate crimes. In this sense, more fine-grained empirical research is needed than simply comparing crime rates in two vastly different countries. Moreover, the problem for Heinze and Weinstein, is that several such fine-grained studies have found that hate speech is in fact correlated with an increase in the incidence of hate crime. Some of these studies were discussed in the second chapter of this thesis, but one particular kind of case that was not touched upon, is that of so called 'hyperpolitical hate speech' - hate speech that comes from political figures.

Hyperpolitical hate speech represents a special case. Brown and Sinclair argue compellingly that politicians may have special duties to refrain from engaging in hate speech, tied to the idea that political office comes with duties as well as privileges.<sup>202</sup> Leaving this issue to one side, it is the influence that political figures wield that should especially worry us when it comes to their engaging in hate speech and quasi-hate speech. It is notable that, commenting on the 16 year high of incidences of serious hate crimes in the U.S. (i.e. violent hate crimes and threats), Brian Levin, the director for the Center for the Study of Hate and Extremism, suggested that "[i]n recent years, the spikes in incidents have consistently correlated with

---

<sup>202</sup> Brown and Sinclair, 'The Politics of Hate Speech Laws'.

political attacks against specific marginalized groups. The more we have these derisive stereotypes broadcasted into the ether, the more people are going to inhale that toxin”.<sup>203</sup> Levin’s comments appear to be backed-up by the facts. James Piazza found that “hate speech by politicians is a statistically significant, and substantive, driver of domestic terrorism”. In some countries, the frequent use of hate speech by politicians was reported to result in an almost nine-fold increase in the rate of domestic terrorism. Piazza adds that his “findings also suggest that increased civility on the part of political figures, or projection of tolerance in politicians’ public statements, could yield a substantial security dividend by reducing domestic terrorism”.<sup>204</sup> In recent years, much of the research on hyperpolitical hate speech has centred around Donald Trump. Christian Crandall et al found that Trump’s 2016 presidential campaign resulted in an increase in prejudicial attitudes towards those groups targeted by Trump’s abuse. The authors conclude that Trump’s prejudicial rhetoric “may have played a role in the substantial increase in bias-related incidents” in the U.S.<sup>205</sup> Ayal Feinberg et al consider the impact that Trump’s campaign rallies had in the local area. They find that Trump’s rallies led to a significant increase in the level of hateful actions directed towards vulnerable minorities.<sup>206</sup> Naturally, the problems didn’t end with Trump’s campaign though. Griffin Sims Edwards and Stephen Rushin found that Trump’s election coincided with a statistically significant surge in hate crimes, even when controlling for alternative explanations. They propose that Trump’s rhetoric, as well as the validating effect of Trump’s electoral victory, are significant drivers of hate crimes in the U.S.<sup>207</sup> The authors seem to have a point: a

---

<sup>203</sup> <https://www.theguardian.com/society/2019/nov/12/hate-crimes-2018-latinos-transgender-fbi>

<sup>204</sup> James Piazza (2020). “Politician hate speech and domestic terrorism”, *International Interactions* 46 (3), pp.431-53.

<sup>205</sup> Christian Crandall, Jason Miller, Mark White (2018). “Changing norms following the 2016 U.S. presidential election: the Trump effect on prejudice”, *Social Psychological and Personality Science* 9 (2), pp.186-92. See also: Olga Chyzh, Mark Nieman, and Clayton Webb (2019). “The effects of dog-whistle politics on political violence”, unpublished paper. Iowa State University.

<sup>206</sup> Ayal Feinberg, Regina Branton, and Valerie Martinez-Ebers. “The Trump effect: how 2016 campaign rallies explain spikes in hate”, unpublished paper.

<sup>207</sup> Griffin Sims Edwards and Stephen Rushin (2018). “The effect of President Trump's election on hate crimes”, available at SSRN: <https://ssrn.com/abstract=3102652>

significant number of those perpetrating hate crimes, both during and after Trump's campaign, have explicitly cited Trump as an influence.<sup>208</sup>

In general then, the problem with Weinstein's account is that it relies on a Heinzean scepticism about the harms of hate speech, or at least the effectiveness of hate speech regulation, to really fill in the gaps. Once we cast doubt on the Heinzean account, Weinstein's account is significantly weaker. In particular, two important problems are worth highlighting. First, Weinstein's account charges hate speech regulation with being viewpoint discriminatory, yet this charge is highly contestable given the ostensibly viewpoint-neutral government aim of preventing harm. Second, even if we grant that hate speech regulation is viewpoint discriminatory, it may be that it is nonetheless a just and fair form of viewpoint discrimination.

First then, Weinstein argues that what underwrites hate speech regulation on say, racist and bigoted speech, is the government's own viewpoint discriminatory attitudes towards such speech. Of course, there are a few different things Weinstein might mean by this. He might be suggesting that it is the government's personal feelings of offence and disgust towards such speech, that motivates their censorious interference. This is an ungenerous suggestion. Such speech regulations are in fact justified, not by the government's feelings of offence and disgust, but rather by concerns about the harms that this kind of speech may lead to that the government has a legitimate interest in preventing. For instance, the government might worry that if it permits a certain kind of racist speech, then listeners will be inspired to act violently or discriminate against certain groups. Similarly, when it comes to persuasion-control, it is the worry that certain forms of persuasion (e.g. to do with climate change denialism, anti-vaccination views, anti-medical views, hatred against vulnerable groups, etc.) might lead to substantial harms, that lies behind the government's decision to regulate. Here it is the *harmfulness* of the speech, rather than its content, or the government's dislike of its content, that explains the government's regulatory response.

Indeed, in cases of this kind of regulation, the content of the speech being regulated arguably has no bearing on the government's motivations at all, let alone their attitudes

---

<sup>208</sup> Edwards and Rushin 'The effect of President Trump's election on hate crimes'; Mark Potok (2017). "The Trump effect", *Southern Poverty Law Center*, available at: <https://www.splcenter.org/fighting-hate/intelligence-report/2017/trump-effect>; Dan Barry and John Eligon (2017). 'Trump, Trump, Trump!' How a president's name became a racial jeer", available at: <https://www.nytimes.com/2017/12/16/us/trump-racial-jeers.html>

towards this content. One way of demonstrating this point, is by considering that even the most censorious piece of hate speech regulation that places extensive restrictions on all manner of racist and bigoted speech that are judged to be dangerous, will still in fact permit a good deal of racist and bigoted speech. At the end of the day, no matter how racist and bigoted speech might be, it will still be permitted so long as it does not pose a danger to the community. Contrary to initial appearances then, it seems as though hate speech regulation does not in fact involve the kind of viewpoint discrimination that Weinstein is so worried about. In this way, Weinstein's worries about viewpoint discrimination and liberal ideologies appear to be misplaced.

In his most recent essay on this topic, Weinstein considers something like this objection. In response he writes that “[i]t is the speech that a law prohibits, not that which it leaves unregulated, that is relevant to determining whether a law is content-based or content-neutral”, before moving swiftly on.<sup>209</sup> This response is somewhat terse, yet it hints at an interesting thought that might be used to defend his account. That thought, roughly put, is that there is no easy way of separating the judgement that a piece of hate speech is harmful, from a viewpoint discriminatory judgement about the content of that speech. When the government regulates, say, racist speech that threatens to persuade listeners to develop harmful ideas, it is attempting to prevent harms that are intimately tied up with the content of the racist speech in question. Put simply, it is no coincidence that racist speech (unlike say egalitarian speech) is sometimes liable to promote hatred, sow discontent, or otherwise lead to harm. Indeed, it is precisely because of the content of racist speech – the dangerous ideas and viewpoints inherent in it – that listeners who have understood and accepted this speech might develop harmful attitudes or intentions.

A comparison may help drive this point home. Imagine that each winter a group of racists go for a relaxing log-cabin retreat at the foot of a mountain. They go to share their hateful views, discuss their twisted theories, and to ideally persuade some of the locals about the alleged threat that certain races pose. Imagine also that the local government is reasonably worried that the noise made by these racists is likely to cause a deadly avalanche. As such, the government decides to restrict the racists from voicing their opinions. This form of government

---

<sup>209</sup> Weinstein, 'Viewpoint discrimination, hate speech, and political legitimacy: a reply', p.746.

regulation is not viewpoint discriminatory. Why not? Because the harms the government identifies as caused by the speech it is restricting, are in an important sense independent of the content of that speech. By contrast, hate speech regulation restricts speech out of a concern for harms that are intimately tied up with the content of that speech. It is no accident that hate speech is liable to cause harm, and the government is as aware of this fact as anyone. The same goes for harmful persuasion: these harms are not the product of some random neurological malfunction that makes listeners become violent whenever they hear words with certain sounds next to one another. On the contrary, harmful persuasion involves speakers knowingly persuading listeners to adopt views and intentions (e.g. the intention to murder someone, resentment towards certain groups in society, scepticism towards the advice of doctors, disbelief in climate change), that will predictably lead to harmful outcomes.

Even with all this said, we should still resist the suggestion that hate speech regulation (or indeed persuasion-control) is really viewpoint discriminatory. Granted, harms that justify these regulatory responses are intimately connected with the content of the censored speech. But to put our scepticism plainly, the fact that two things are connected does not mean they are the same. These regulations involve the government responding to the harmfulness of speech. That harmfulness is related to, although separate from, the content of that speech. As such, hate speech regulation and persuasion-control are responses to the harmfulness of certain kinds of speech, rather than to the views or attitudes expressed in that speech per se. For this reason, they are not viewpoint discriminatory in any conventional sense.

Before we move on to the second main objection, we want to quickly consider and reject a few alternative ways of interpreting the suggestion that hate speech regulation and persuasion-control are viewpoint discriminatory. It might be suggested, for instance, that hate speech ordinances are viewpoint discriminatory in the sense of stemming from a governmental bias against and preoccupation with hate speech and hate related crimes. Indeed, so the argument goes, the fact that we have specially designed hate speech ordinances, rather than relying on more general common law regulation (i.e. the regulation of incitement), betrays this viewpoint discriminatory obsession with hate speech. Why else would the government feel the need to design special ordinances, if not for the reason that its ministers find hate speech and hate crimes to be especially offensive?

In response to this, we will offer two points. First, it may be that the government's decision to construct special hate speech ordinances, simply follows from a value neutral recognition that hate speech is particularly dangerous form of expression and constitutes a special threat in our current climate. Perhaps, for instance, the government has reason to believe that in the current climate hate speech is particularly likely to escalate into violence (and other kinds of harm). Or perhaps, to return to persuasion-control, the government judges that citizens are becoming increasingly liable to develop a harmful scepticism towards medical advice. With this in mind, a specially designed ordinance may reflect the government's desire to provide prosecutors, officers, and indeed citizens, with clear and visible guidance regarding the prohibited nature of this speech.<sup>210</sup>

Second, and relatedly, it may be that certain famous hate speech ordinances are the result of specific political and cultural factors. Consider, for instance, the landmark 2006 Racial and Religious Hatred Act. The Act famously made it an offence in England and Wales to use threatening words or behaviour with the intention of stirring up hatred against people on grounds of their religion. Brown and Sinclair offer a pluralist analysis of the motivation behind this act, whereby the Act was introduced partly out of a public order concern, partly in order to extend to Muslims the kind of legal protection already enjoyed by other religious groups, and partly in response to vocal calls for protection from a now organised and vocal Muslim community who felt (with some justification) that they were being disproportionately exposed to hate speech in the wake of high-profile extremist Islamist terror attacks.<sup>211</sup> Brown and Sinclair's account speaks to both the special need to highlight and tackle the problem of hate speech, as well as the real-life political factors that likely motivated the Labour government at the time. None of these factors need strike us as being viewpoint discriminatory or part of an unacceptable liberal agenda.

Finally, it might be argued that while the government may appeal to ostensibly uncontroversial and neutral harm-prevention rationales in order to justify engaging in hate

---

<sup>210</sup> See for instance:

<https://www.gov.scot/binaries/content/documents/govscot/publications/progressreport/2018/05/independent-review-hate-crime-legislation-scotland-final-report/documents/00535892-pdf/00535892-pdf/govscot%3Adocument/00535892.pdf>

<sup>211</sup> Brown and Sinclair, *'The Politics of Hate Speech Laws'*.

speech regulation, a subtle kind of viewpoint discriminatory impulse may nonetheless lurk behind the scenes. This kind of objection draws heavily on Dworkin's suggestions in "Is there a right to pornography".<sup>212</sup> In this article Dworkin suggests that even ostensibly uncontroversial time and place speech restrictions, may in fact be used by the government as a screen for their more perfectionist impulses. Dworkin was particularly worried that the government might use time and place-based bans on pornography as an excuse to suppress gay smut which the government found distasteful. Along these lines, one might worry that the harm-prevention rationale evoked by the government to justify hate speech regulation and persuasion-control, is simply used as a way of disguising the government's moralistic impulses which are the true motivation behind the interference. Moreover, Dworkin also suggested that harm-prevention justifications can sometimes become entangled in more subtle ways with moralistic impulses. In particular Dworkin worried that otherwise perfectly valid government evaluations about the harmfulness and value of speech, might become subtly interwoven with more troubling perfectionist impulses to do with the tastelessness, vulgarity, or lack of merit of that speech. Along these lines one might worry that even respectable harm-prevention rationales that offer ostensibly good justification for hate speech regulation and persuasion-control, may all the same become sullied by the desire of government members to e.g. tell citizens what to do, impose their moral view of the world on citizens, punish those they see as grossly hateful, get rid of speech that they consider stupid, etc. To be clear, the idea here isn't that government officials cynically hide their true motives behind an ostensibly respectable harm-prevention rationale. Rather, the idea is that the government's desire to pursue this harm-prevention rationale and willingness to respond to these particular kinds of harms, may be subtly informed and framed by less respectable biases and desires.

There is something to this Dworkinian worry; certainly we should be duly vigilant (and perhaps even anxious) when it comes to assessing the true motivations of our political masters. Yet equally, there is something rather cynical, and perhaps even ungenerous, about Dworkin's critique. Note, for instance, that the argument treats Dworkin and Weinstein's opponents with a possibly unfalsifiable distrust and scepticism. I'm not sure this is entirely fair of Dworkin and Weinstein. Indeed, as easy as it is for Dworkin and Weinstein to charge their opponents with being moralistic tyrants, one could just as easily (and unfairly) charge Dworkin and Weinstein

---

<sup>212</sup> Dworkin, 'Is there a right to pornography?'

with being closeted bigots whose complicated democratic arguments merely operate as a screen for the fact that they enjoy voicing and listening to hateful arguments. We do not gain anything, especially when it comes to advancing our philosophical understanding, from engaging in these kinds of psychological critique against our opponents. We should refrain then, at least in the absence of more evidence (e.g. a pattern of bias evident in the government's regulations), from engaging in this kind of psychological objection to the government.

Moving on now to the second important problem with Weinstein's account: imagine we concede to Weinstein that hate speech regulation and persuasion-control amount to a form of viewpoint discrimination, insofar as they involve an attempt to prevent harms that are closely connected with the content of the speech being regulated. And we also concede that such viewpoint discriminatory regulations burden some more so than others. Even so, it does not follow that this kind of viewpoint discriminatory conduct is therefore unjust or unfair, or liable to dent political legitimacy in any way.<sup>213</sup> On the contrary, this kind of viewpoint discrimination may still be perfectly fair and justified. The key, although often unstated, idea that surely underwrites Weinstein's critical treatment of viewpoint discrimination, is that viewpoint discrimination is unfair or unjust in some way. However, when we look at hate speech regulation (and indeed persuasion-control), there is no immediate reason to think that it is any of these things.

Consider the regulation of murder and assault. Such regulation strikes us as just and fair because it simply curtails the freedom of people to harm others. It may be that persuasion-control and hate speech regulation can be justified by a similar kind of rationale. After all, citizens arguably have no claim on endangering others; if expressing or listening to certain arguments exposes others to unacceptable danger, then the government would seem to be perfectly justified in making an intervention. In other words, there is reason to suppose that such regulation involves no normative cost or blow to political legitimacy whatsoever, regardless of whether it turns out to be technically viewpoint discriminatory or not. After all, it simply prevents agents from doing what they have no claim to do in the first place.

---

<sup>213</sup> For more on this point about fairness, see 'Hate speech bans, democracy, and political legitimacy', Weinstein approvingly quotes a passage of Dworkin's that highlights fairness, and compares his account to Dworkin's after characterising Dworkin's view as one in which fair democracy depends on being free to voice one's opinions, no matter what the government thinks of them.

Overall, we have charged Weinstein's account with suffering from two important problems. First, it is unclear whether hate speech regulation or persuasion-control are really viewpoint discriminatory in the way Weinstein suggests. Second, even if these regulatory responses are viewpoint discriminatory in some sense, this does not show that they are unjust or unfair. On the contrary, seeing as they would appear to primarily prevent citizens from doing what they have no right to do in the first place, these regulatory responses should strike us perfectly valid.

#### §IV. Analogies

We want to end our critique of Weinstein by talking about the intuitive appeal of his position, as well as one last way in which Weinstein has sought to defend his account. We have argued that contrary to Weinstein's suggestions, neither hate speech regulation nor persuasion-control need come at a significant moral or political cost. Still, maybe at this point Weinstein can dig in his heels. In particular, he might insist that hate speech regulation and persuasion-control are viewpoint discriminatory and hence unjust in such a way that is neither ameliorated nor scrubbed away by the fact that they also prevent harms. And that as such these regulatory responses always constitute a blow to political legitimacy. To help show the intuitive appeal of this position, Weinstein has drawn an analogy between hate speech regulation and a demonstrably outrageous case where the government disenfranchises those it fears will vote against a proposed bill. Weinstein hopes to show that just as we find the government's actions in the voting case to be unfair or unjust or as involving some kind of illegitimate agenda, we should draw a similar conclusion about hate speech regulation. However, this appeal fails because the cases are crucially disanalogous. What's more, when we modify the voting case so as to make it more analogous, it no longer elicits the kind of anti-regulation intuitions Weinstein is hoping to provoke.

The voting case goes like this: the government has proposed a new bill that will increase property tax. The bill being passed is subject to it being voted for in a city council election. In order to ensure this happens, the government imposes onerous voting restrictions (e.g. age verification measures) that it knows will have the effect of mainly disenfranchising citizens who

oppose the bill and would vote against it.<sup>214</sup> The case is supposed to elicit the intuition that at the very least the government's upstream disenfranchisement undermines the political legitimacy of the downstream tax bill. Weinstein even goes as far as to say that those who have been disenfranchised have no political obligation to obey the new tax law. What's more, we are invited to notice that this blow to political legitimacy is not scrubbed away by the fact that disenfranchisement might be necessary to secure certain benefits or even prevent the disastrous effects that would be brought about by the bill's failing to pass.<sup>215</sup> Finally, Weinstein suggests that just as we find the government's decision to exclude certain citizens from the democratic process to be utterly outrageous in this case, we should draw a similar conclusion regarding hate speech regulation and persuasion-control.

While this is an interesting and provocative case, it is doubtful what lessons we can draw from it seeing as how it is fundamentally disanalogous to hate speech regulation and persuasion-control. Three disanalogies in particular are worth highlighting: first, hate speech regulation and persuasion-control amount to a far more piecemeal restriction compared to disenfranchisement. This is partly due to the difference between voicing an opinion on a law and voting on a law; if John is disenfranchised, he is completely excluded from voting and there is no other way in which he can directly impact upon the passing of the law. By contrast, if John is prevented from engaging in harmful hate speech, he is still able to say a good deal about what he thinks about the law and generally voice his opinion. Granted, hate speech regulation might prevent him from expressing his views with perfect precision. But as discussed earlier, no piece of hate speech regulation would prevent John from voicing even extensive dissent or disagreement with the law.

The second disanalogy concerns the harms in the two cases. Whereas the underlying justification behind hate speech regulation and persuasion-control is the prevention of serious and obvious harms, the underlying justification in Weinstein's voting case is a little more nebulous. In the voting case, the government disenfranchise voters in order to ensure an

---

<sup>214</sup> Weinstein, 'Viewpoint discrimination, hate speech, and political legitimacy: a reply', p.726.

<sup>215</sup> Weinstein, 'Viewpoint discrimination, hate speech, and political legitimacy: a reply', p.728. Weinstein writes that "the existence of all these other legally-conferred benefits, while relevant to the measure of a legal system's legitimacy, does little to ameliorate the damage to a citizen's ability to protect those interests associated with a particular downstream law arising from an upstream speech restriction."

effective property tax bill will be passed. Presumably we are to imagine that, should the bill not pass, people will be harmed as a result. But it is difficult to imagine how exactly this is meant to happen. Granted, we probably *can* imagine such a scenario; it is not beyond our imaginative capacities. But we should be wary about relying on our intuitions concerning under-described cases like this one.<sup>216</sup> Moreover, whereas hate speech involves not only causing harm, but causing harm deliberately or in a way that is foreseeable, nothing in Weinstein's description of the tax example alerts us to this being the case. Third, the disenfranchisement case features deception on the part of the government. That is, the government introduces ID laws that are ostensibly introduced in order to prevent voting fraud, but what the government doesn't reveal is that the real motivation for introducing these laws is so that citizens who oppose the property tax will be disenfranchised. By contrast, the best kind of hate speech regulation involves no such deception.

With these worries in mind, let us sketch a version of the voting case that is more analogous to hate speech regulation and persuasion-control, with which we may test our intuitions: once again, the government imposes voting restrictions. However, this time the government does not disenfranchise any citizens who would normally be able to vote. Instead, it chooses to impose a series of piecemeal regulations and limitations designed to prevent serious and obvious harms from taking place. For instance, the government allows all citizens to vote in the council elections and ensures there are a number of council candidates that represent a wide range of views, however candidates are banned from offering or supporting certain policies that are obviously dangerous or harmful. In particular the government prevents councilors from offering discriminatory policies or policies that would temporarily legalise assault and battery. In effect, then, citizens are prevented from using their vote to support these policies. Finally, we are to imagine that the government is transparent and upfront about these piecemeal regulations and the reasoning behind them.

This case is much more analogous to hate speech regulation and persuasion-control. The problem for Weinstein is that the government's restrictive actions in this case will strike many as perfectly acceptable. Indeed, arguably the U.S. already impose these kinds of limits on citizens and politicians in the form of the constitutional Bill of Rights. Note here that we are

---

<sup>216</sup> See James Wilson (2016). "Internal and external validity in thought experiments", *Proceedings of the Aristotelian Society* 116 (2), pp.127-52.

conceding that, to some extent, these constitutional protections and hate speech regulations impose limits on the ability of citizens to have their say and participate in the democratic process. But the fair aim of preventing serious harms that people have no right to cause in the first place, provides the requisite justification for these regulations – without having a detrimental effect on political legitimacy.

## §V. Zombie Weinstein

In this closing section, we consider one final way in which Weinstein’s worry about viewpoint discrimination and persuasion-control might be revived. In §III, we objected to Weinstein by arguing that the best kinds of speech regulation are: i) not viewpoint discriminatory, but rather the product of a value-neutral harm-prevention rationale, and ii) primarily involve imposing fair burdens that only prevent agents from doing what they have no claim to do in the first place (endanger others). Now, however, we ask if we can proverbially bring Weinstein’s account back to life. In particular, we suggest that there is a certain brand of persuasion-control (and indeed hate speech regulation) that in fact imposes unfair burdens on innocent agents who are not personally endangering others, and which risks exacerbating pre-existing tensions between different political groups in a way that should worry those of us who are sensitive to the issue of viewpoint discrimination and equal respect.

Thus far this chapter has focused on speech regulation that limits the ability of certain citizens to listen to and express generally off-colour views e.g. arguments that question climate science, arguments that encourage disobedience of medical advice, arguments that promote distrust and hatred towards certain groups, incitement of non-imminent violence, etc. In particular we have focused on those citizens who might listen to and express these views in ways that lead to harmful outcomes, archetypical examples of which might be a famous anti-Islamic hate speaker, or an influential religious zealot who promotes anti-gay sentiment in his community, or the listener who is susceptible to these arguments and is moved to eventually engage in violence. Call these ‘irresponsible citizens’, as the way they engage in and with speech ends up endangering those around them. Thus far, it has been argued that persuasion-control and hate speech regulation are neither viewpoint discriminatory nor the result of unfair viewpoint discrimination, when it comes to irresponsible thinkers.

But what about those we might term ‘responsible thinkers’? Here we have in mind those who express off-colour arguments or enjoy engaging with such arguments, but who do so in a way that does not endanger others. Archetypical examples include the bigoted old lady who regularly shares her outdated views with neighbours (who generally pay her no particular mind), or the quiet racist who enjoys hearing anti-Islamic views expressed in public but who does not go as far as to engage in discrimination himself. These citizens do not personally engage in dangerous behaviour. Indeed, neither hate speech regulation nor persuasion-control is designed to curb these more responsible citizens’ freedoms. And yet note, that this is exactly what they do. These citizens are now denied the chance to fully engage in public discourse; either in the sense of being prevented from voicing certain views (and the arguments in favour of them) that they hold, or being denied the chance to hear and engage with these views voiced in public. This is a significant burden to bear, both in terms of the limits it places on the ability of citizens to develop their views upon their being debated in public, and in terms of the sense of validation that goes along with sharing and engaging with political views that are close to one’s heart.

Now it might be replied at this point that, despite what has just been said, these so called ‘responsible’ citizens are in fact endangering others, insofar as they are contributing in some way to the general social milieu that predictably results in harms against vulnerable groups. Consider, for instance, as well as our bigoted old lady and quiet racist, the immigration critic that is sometimes overzealous in his linking of immigration with crime, or the religious homophobe who thinks it is vital to express and share her views about the moral horrors of homosexuality, or the fanatical sceptic who regularly questions medical advice and immunology experts. Granted, these persons may not personally be responsible for any particular harms. But the suggestion here is that we are dealing with structural, aggregative harms; and that while one person holding off-colour views may not be so bad, when these views become popular many people hold them at once, and we are left with a society where certain members of vulnerable groups find themselves generally less respected, less likely to occupy positions of power and status, more likely to be discriminated against in the work place, less sexually desirable, more likely to be ignored, more likely to be viewed as dangerous, and so on. Insofar then as these aggregative harms are nurtured and promoted by certain kinds of speech, we have a powerful reason to engage in censorship of those kinds of speech, and *crucially* should not feel guilty

about depriving persons of the ability to engage with speech in ways that contributes (even in a minor way) to harmful outcomes. Indeed, being able to mobilise the kind of large-scale co-operation needed to address these structural, aggregative harms is one of the unique benefits of having a government in the first place.

The point is well taken, and we do not wish to deny, as Weinstein and Heinze seem to deny at times, that there are subtle, aggregative harms that certain forms of expression and behaviour may well promote. Indeed, these harms may provide us with a powerful reason to engage in censorship. However, we also think that we should feel a degree of sympathy with speakers and listeners who do not contribute in any way to these harms, or who contribute only in very minor and indirect ways. The ‘responsible thinkers’ from the examples above plausibly fall into the latter category. That is to say, while these speakers are not personally responsible for any speech-related harms, it may be that their engaging in and listening to certain kinds of speech contributes in a small way to the general cultural milieu that surrounds them, and which turns out to impose significant burdens on certain groups. Persuasion-control and hate speech regulation are regrettable in the sense that they penalise these citizens who would not normally be liable for regulation because they have not personally harmed or even endangered others, but who must now, because of aggregative considerations and the realities of large-scale governance, take on the burdens of regulation.

At this point a critic might accuse us of blowing matters out of proportion. True, they might say, speech regulation involves burdening ‘responsible’ citizens who have personally committed no wrong. But many forms of perfectly acceptable and uncontroversial government (and non-government) regulations do just this. Consider, for instance, the case of airport security; in the name of preventing just a few irresponsible and dangerous actors, all who wish to fly by plane must endure restrictions on what they can carry on the plane, as well as take the time to pass through a scanning machine. Similarly, one might think that gun ownership regulation burdens many who pose no threat to others, simply in order to prevent a few irresponsible citizens from behaving in ways that might harm others. Another case might be that of subpoena laws, that often order innocent citizens to comply in certain ways with rather onerous demands from the court, so that the court may pursue its goal of achieving justice. Finally, consider the classic Hollywood chase scene where the policeperson has to commandeer the vehicle of an innocent bystander, in order to keep on chasing the baddie. In all these cases,

innocent citizens who have done nothing wrong must take on certain sacrifices, so that other, less responsible citizens can be prevented from harming others. Nonetheless, none of these cases should strike us as objectionable, because the limitations are relatively low, and sometimes small sacrifices can be justifiably imposed in the name of protecting others. Indeed, one could even pursue a rather stronger version of this argument: that allowing ‘innocent’, responsible citizens to go unburdened in these cases, would come at the cost of increasing the exposure of others to serious harm. In this way, the so called innocent and responsible agents in question actually count as ‘innocent threats to others’, even if they themselves personally have done nothing wrong. If we can eliminate these threats via imposing small burdens, so the thought goes, then we ought to do so, and none of this ought to strike us as particularly controversial or objectionable.

The thought before us then, is a rather democracy-orientated way of conceptualising citizens and their government; it is reasonable for the neutral and fair government to impose burdens even on ‘responsible’ and innocent persons, because we are all in ‘this’ (i.e. this democratic, civil society) together. As such, we must occasionally endure sacrifices for others, even when we ourselves have done nothing wrong. This thought is amplified by considering that whereas sometimes we are the ‘responsible’ person who must take on sacrifices because of the stupidity of a few ‘irresponsible people’, at other times the tables will be turned. For instance, John might be perfectly trustworthy when it comes to owning a gun or operating on himself or driving a car while having several units of alcohol in his system. And hence he might have *prima facie* reason to resent having regulatory burdens imposed on him just because other citizens are not so competent or trustworthy in this area. However, John should take solace in the fact that in some areas, he is the irresponsible person spoiling it for the responsible ones. Perhaps, for instance, most people could responsibly be trusted to hear hate speech without it inspiring them to harm others, but John is one of those hot heads who may well do something stupid upon being whipped up into a frenzy by a charismatic speaker. All of this speaks to a view of our democratic project, averred to several times throughout this thesis, where citizens are conceptualised as highly imperfect, highly vulnerable (but also brilliant) creatures who must somehow find a way of co-operating, if we are to ever achieve flourishing on any kind of mass scale. This vision is not always rosy – it highlights the flaws of citizens, and the need for sacrifices. But it is a vision that respects citizens enough to take seriously the needs of all.

Where does all of this leave us? We have concluded that persuasion-control (and hate speech regulation) does impose burdens on ‘responsible’ agents who have personally done nothing wrong. But we have also suggested that this need not strike us as particularly objectionable, seeing as it is reasonable to ask agents (innocent or not) to make relatively small sacrifices in the name of protecting others. All the same, we should be mindful of the way different regulatory responses can impose burdens on innocent agents in more or less worrying ways. Recall, for instance, the way in which standard airport security measures impose the same non-invasive burdens on all who pass through the airport. This kind of blanket imposition of burdens is very different to airport security measures that are applied selectively, as when certain persons are picked out because of their race or religion as posing a higher *ex ante* threat. Note that while both types of airport security measures impose burdens on innocent agents who have done nothing wrong, the latter case is more worrying for a range of reasons. What is most notably worrying is the kind of ‘othering’ involved in this selective security measure, and the way it exacerbates pre-existing tensions between different political groups in society. Even if – and it is quite a large if – there is a sense in which airport security is justified in believing that certain persons are more likely to pose a threat because of their race or religion, those of us who care about social equality and the dignity of different groups have reason to regret this kind of interference.

It is notable then, that persuasion-control (and hate speech regulation) involves a similar kind of ‘othering’ to that which we find the worrying form of selective airport security. Our political views play an important role in our lives and how we view our place in society. It should not be denied that for those who sympathise with these views, it is a substantial benefit to be able to express and hear arguments that e.g. speak to the ‘problems’ of immigration, criticise medical orthodoxy, take aim at climate change science, express moral worries with ‘homosexuality’, etc. The government may well be justified in singling out these kinds of arguments as dangerous, or at least contributing to dangerous social changes. But equally it is likely that this kind of selective interference will exacerbate pre-existing feelings of political alienation, as well as tensions between various political and religious groups. This may well be deeply regrettable given that we are dealing with citizens who have personally committed no wrong. Granted, the analogy with selective airport security is an imperfect one; it may be particularly bad and especially offensive to be picked out for regulation because of one’s

religious views or racial background. But it strikes us that picking out and ‘othering’ innocent citizens based on having the wrong kinds of views – especially political views – might have a broadly similar negative effect on the psyche of the citizen in question, as well as raising some of the same worries about exacerbating social divisions among citizens. Indeed, this is the sense in which persuasion-control may pose a problem in terms of viewpoint-discrimination; that is, in terms of exacerbating pre-existing political tensions and feelings of isolation, that are felt by those who hold certain ‘off-colour’ political views.

In this way, we arrive at the most plausible version of Weinstein’s worry. As Weinstein alerts us to, persuasion-control is selective in nature; it involves picking out certain viewpoints and political arguments as being more dangerous than others. The problem comes in when we consider that such regulation is likely not only to burden ‘responsible’ citizens who have personally committed no wrong, but that the selective nature of this burdening is liable to stigmatise these citizens and exacerbate pre-existing social tensions and feelings of political alienation. This chapter draws to a close by considering two potential responses a reader might have to this way of problematising persuasion-control. After dismissing these responses, we offer some suggestions about how the government might go about engaging in persuasion-control in a way that somewhat defuses this worry about stigmatisation and ‘othering’.

The first reply takes issue with our describing as ‘innocent’ and ‘responsible’, those persons who hold off-colour views but who may express and engage with these views in ways that they personally harm no-one. In the course of presenting early drafts of this chapter, one suggestion that has consistently come up in discussion is that these persons are not so ‘innocent’ after all. On the contrary, these persons hold dangerous, hateful, anti-scientific views. As such, they can have no complaint if they are therefore subject to regulatory burdens. However, we are largely unsympathetic towards this reply. Granted, the ‘responsible thinkers’ from our examples – the bigoted old lady and the quiet racist – hold regrettable, racist views. But given that they have personally endangered no-one, the government should nonetheless have a degree of sympathy towards these persons when it imposes regulatory burdens upon them. Indeed, were the government to hold that the lamentable quality of these persons’ views mean that they are especially deserving of regulatory burden, then this would surely count as a rather egregious example of moralistic, viewpoint discrimination.

A second kind of reply pushes the thought that the liberal government has no duty to protect off-colour views. Being committed to liberalism, so the thought goes, involves a commitment to granting all persons a basic kind of respect. Along these lines, a consistent liberal authority, rather than simply showing respect for all views, ought to rather deny special protection to those views that conflict with its own liberal project. This kind of thought is touched upon by the likes of Jonathan Quong, Cecile Laborde, and (somewhat ironically) Ronald Dworkin, and may seem to establish that the government should not offer protection to e.g. hate speech.<sup>217</sup> Quong, for instance, suggests that the consistent liberal government should be committed to certain tenets – such as the view of all persons as free and equal. When we are reviewing, for instance, a proposed system of legal rights that the government might establish, we should ask whether that system lies in conflict with this tenet. If there is such a conflict, the liberal government should reject that system as unreasonable. In this way, the liberal government could coherently refuse to protect the right of citizens to express off-colour views, on the grounds that it has no duty to enable citizens to act unreasonably. However, even if we agree that the liberal government should not provide special protection for hate speech, note that the Weinsteinian argument offered in this section does not demand such special protection. On the contrary, it demands simply that those who hold hateful views are treated just like other citizens; that we have sympathy for them when they are burdened despite personally doing nothing wrong, and we are sensitive to the ways certain regulations may be ‘othering’, alienating or are likely to exacerbate pre-existing political tensions involving these citizens. The liberal government, so the thought goes, do not owe the citizens in these cases special legal protections, but they do owe them their sympathy and respect.

In the end then, there is a sense in which Weinstein was right to worry about the viewpoint discriminatory tendencies of persuasion-control and hate speech regulation. Certainly, we should be duly anxious about the way these regulations are prone to punishing innocent citizens and exacerbating pre-existing feelings of political alienation. Still, I’m not sure this quite delivers the knock-down objection to persuasion-control that this thesis is looking for. Granted, we have pointed to some *prima facie* worrying effects of persuasion-control. But there are a number of ways in which the government might try and either neutralise these

---

<sup>217</sup> Quong, ‘*Liberalism Without Perfection*’; Cécile Laborde (2018). ‘*Liberalism’s Religion*’ (Cambridge, Massachusetts; Harvard University Press); Dworkin, ‘Is there a right to pornography?’.

effects, or indeed embrace them. For instance, the government should make it clear in its justification of persuasion-control and public communications surrounding the regulation, that it recognises that innocent persons may be burdened by this regulation, and that such burdens are the regrettable price we pay for living in a large-scale democracy, rather than any kind of deserved punishment for holding the ‘wrong’ kind of views. Indeed, the government could explicitly express its sympathy towards those citizens who would have benefitted from hearing the now regulated ‘off-colour’ views, without posing even the slightest danger to others. These kinds of public communications may go some way to shielding these citizens from the harms of political alienation and regulatory stigma, that might otherwise befall them.

What about those ‘responsible’ citizens who have not personally harmed others, but who have contributed in some relatively minor way to the general social milieu that disadvantages certain groups? Should the government issue the same kind of apologetic messages to these citizens? That would certainly be one option; the government could also emphasise that the sacrifices they are making (in being excluded to some extent from public discourse) are appreciated and that it is thankful for the way these sacrifices help to prevent other, less responsible persons from harming others.

But there is a rather less sympathetic approach available to the government. The government could argue, for instance, that in a functioning democracy where citizens bear responsibilities to one another, citizens should not so much as even contribute in minor ways to harming others. Indeed, citizens should not even help support the larger systems that harm others. In this way, low-level polluters, quiet racists, bigoted old ladies, those that baselessly question medical advice, may all be held responsible for the harms that result much later down the line from their conduct, and deserve little sympathy for the regulatory burdens that befall them.

This approach to reconceptualising the duties and liabilities of citizens, in turn reconceptualises the role of the government as a neutral arbiter over our competing claims. In a democracy, so the thought goes, one of the important advantages of having a government is that it can compel large scale co-operation among its many citizens, in a way that allows us to tackle structural, aggregative harms. Indeed, this is one of the special jobs that perhaps only a government can do. In this way, it is part of the job of even the anti-perfectionist government committed to the value-neutral aim of preventing harms, to hold its citizens responsible for

contributing to the structural, aggregative harm to others. Such a vision of the government does not strike us as unattractive. Granted, it does permit the government to burden even those who have done nothing personally wrong. But, to touch upon a point made at the end of chapter 3, this is plausibly part of what it means to hold citizens as responsible for one another, and as having reasonable duties of assistance. Good democratic government help enforce these duties.

Both these options have their strengths and weaknesses. The latter may well end up alienating certain citizens even further, although perhaps this shift in the way we approach our responsibilities in a democracy will be necessary if we are to ever to prevent the structural, cumulative harms associated with pollution, anti-scientific thought, and identity hatred. It should also be noted that these two options are not necessarily mutually exclusive. It could well be possible, for instance, to strike a careful balancing act between the less sympathetic approach that demands citizens refrain from contributing even in minor ways to larger harmful systems, while nonetheless expressing understanding that extricating one's self from these larger systems is a burdensome project that may well take many years, significant changes in lifestyle, much debate, and large-scale co-operation to truly achieve. This synthesised approach might also focus on a sense of shared culpability; the idea that there have been collective failures that we have all been involved in, but that starting from now we must all try to do better. Striking this balance is never easy, though it will at least be familiar to all those who have participated in (less than perfectly successful) team sports and groups projects at work. For our purposes what is important is that there are options available for the government that engages in persuasion-control and excludes certain persons from public discourse, but that is nonetheless mindful of the way these regulations exclude (at least to some extent) citizens who have not personally harmed anyone through their engagement in public discourse.

### Conclusion

This fourth chapter has sought to problematise persuasion-control by reflecting on the work of James Weinstein and the issue of viewpoint discrimination. In general, we considered whether persuasion-control might conflict with the democratic vision of the government as neutral arbiter over the claims of its citizens and duty-bound to show equal respect towards all views (and those citizens that hold them). The chapter considered a number of different ways of

interpreting this worry, yet concluded that each of these interpretations turns out to fall short of the mark. In the end a novel version of this worry was explored; that those of us who care about viewpoint discrimination may feel somewhat anxious about the tendency of persuasion-control to burden and politically alienate citizens who have personally done nothing wrong.

That said, the chapter ended by canvassing some of the ways in which the government might neutralise, or even permissibly embrace, this tendency. In the end it is concluded that permitting the government to engage in persuasion-control is in fact, in various ways, compatible with a perfectly attractive conception of the democratic government. Our investigation therefore continues concerning whether persuasion-control conflicts in a deep and principled way with our democratic values. In the final chapter we consider the view that persuasion-control may conflict with a certain democratic vision of how citizens should relate to their government, and the tendency of persuasion-control to grant to governments a rather radical form of moral authority over their citizens.

## Chapter 5

### “This relationship isn’t working”: governments, citizens, and censorship

Cashing-out the principled wrong of persuasion-control has proven to be a complicated and difficult task. Indeed, upon closer examination many of the classic critiques of persuasion-control fall rather flat. For instance, sometimes it is argued that persuasion-control involves a significant interference with the freedom or autonomy of those who are denied the chance to hear the now censored ideas. Yet this suggestion turns out to be surprisingly difficult to cash-out in a satisfying way. Certainly most instances of persuasion-control fall well short of mental coercion, thought control, or even a significant interference with our autonomous, intellectual processes (chapter 2).

Elsewhere in this thesis we have looked at the suggestion that censorship conflicts with a principled commitment to respecting citizens (chapter 3). Yet it was argued that persuasion-control, far from proceeding from an ungenerous or insulting conceptualisation of citizens, may in fact reflect a suitably grounded view of the limitations of citizens and their vulnerability to persuasion-mediated harms – not to mention their duties to one another. Some critics take a different tack, arguing that censorship conflicts with the duties and obligations of the

government. But as long as the government can show that censorship is an effective tool for protecting citizens from harm, a more natural conclusion would be that the government is simply doing its job in promoting and protecting the interests of its citizens – even if this does sometimes involve imposing burdens on ‘innocent’ citizens (see chapter 4).<sup>218</sup>

In this final chapter an alternative approach is presented – *the relationship account* – that seeks to synthesise the approaches from chapters 3 and 4. According to the relationship account, instead of looking in isolation at either the democratic ideal of the citizen, or the democratic ideal role of the government, we should instead focus on the government-citizen relationship. The relationship account holds that permitting persuasion-control changes the government-citizen relationship in at least two ways that are objectionable. First, it grants governments a special kind of intellectual authority over their citizens. In particular it is argued that the government-citizen relationship that is established involves a troubling form of intellectual subordination. Second, given that the government-citizen relationship is already highly asymmetric, freedom from this kind of intellectual subordination turns out to be especially important for preserving the standing of citizens.

The chapter is structured as follows: §I lays the groundwork for our relationship account, explaining how the relationship account fits into, and draw insights from, relational accounts of autonomy. §II says a little bit about how relational accounts generally approach the issue of censorship. §III outlines the relationship account and how it helps problematise persuasion-control. §IV draws the chapter to a close by canvassing and responding to some potential objections to the relationship account, as well as commenting on some of its implications.

### §I. The relational approach to autonomy

This chapter draws inspiration from the insights of ‘relational’ accounts of autonomy.

Relational accounts of autonomy have been notably developed by philosophers like Catriona

---

<sup>218</sup> Indeed, a proponent of the Razian service conception of government legitimacy, might well find cause to praise the government’s censorious regulations, on the grounds that it serves the autonomous interests of citizens. Relatedly, Republican accounts of domination object to agents being subject to arbitrary power and control. Such accounts may be thought to furnish us with an objection to the government having the power to decide what arguments we can and cannot hear. Yet, again, the best kind of censorship (precisely worded, narrowly framed) censorship that we have good reason to think will be effective in promoting the interests of citizens, would seem to be far from arbitrary.

Mackenzie, Natalie Stoljar, Marina Oshana, Jennifer Nedelsky, Andrea Westlund, Paul Benson, and Carolyn McLeod. Such accounts are usually presented in the context of debates surrounding the concept of autonomy. Still, in some ways talk of autonomy is unduly restrictive, and perhaps even downright misleading given the etymology of the word. With this in mind, it would be more accurate to say that relational accounts share a general interest in what it means for autonomous agents like us to flourish and thrive. At their core, relational accounts investigate what it means for us to live well as self-governing agents and to meet our many complicated needs, as well as what might constitute a threat to this condition.

Like many philosophical approaches, the relational approach to autonomy has its roots in a critique of a rival theory. Feminist philosophers like Nedelsky and Martha Fineman have long argued that liberalism is fundamentally flawed insofar as it takes “atomistic individuals as the basic units of political and legal theory”.<sup>219</sup> These philosophers sought to critique the liberal ideal of freedom and autonomy as involving independent agents exercising control over their lives.<sup>220</sup> With this in mind, relational approaches are first and foremost a rejection of this atomistic conception of persons in favour of a more socially embedded vision of persons.<sup>221</sup>

Often this starts with a description of how persons operate in the world. It is argued that our lives are ineliminable shaped by our social surrounding. Mackenzie and Stoljar, for instance, argue that the very concepts with which we identify and understand the world around

---

<sup>219</sup> Jennifer Nedelsky (1989). “Reconceiving autonomy: sources, thoughts and possibilities”, *Yale Journal of Law and Feminism* 1 (1), pp.7-36, p.12. In *Law’s Relations*, Nedelsky writes: “The now familiar critique by feminists and communitarians is that liberalism takes atomistic individuals as the basic units of political and legal theory and thus fails to recognize the inherently social nature of human beings. Part of the critique is directed at the liberal vision of human beings as self-made and self-making men (my choice of noun is, of course, deliberate). The critics rightly insist that, of course, people are not self-made. We come into being in a social context that is literally constitutive of us. Some of our most essential characteristics, such as our capacity for language and the conceptual framework through which we see the world, are not made by us, but given to us (or developed in us) through our interactions with others”: Jennifer Nedelsky (2011) *Law’s Relations: A Relational Theory of Self, Autonomy, and Law* (New York: Oxford University Press), p.120; Martha Fineman (2004) *The Autonomy Myth* (London: New Press). See also: Mackenzie, ‘Relational autonomy, normative authority and perfectionism’.

<sup>220</sup> Edward Dove, Susan Kelly, Federica Lucivero, Mavid, Machirori, Sandi Dheensa, Barbara Prainsack (2017). “Beyond individualism: is there a place for relational autonomy in clinical practice and research?”, *Clinical Ethics* 12 (3), pp.150-65; Natalie Stoljar (2011). “Informed consent and relational conceptions of autonomy”, *Journal of Medicine and Philosophy* 36 (4), pp.375-84; Carolyn McLeod and Emma Ryman (202). “Trust, autonomy and the fiduciary relationship”, in *Fiduciaries and Trust: Ethics, Politics, Economics and Law* (Cambridge: Cambridge University Press), ed. Paul Miller and Matthew Harding, pp.74-86.

<sup>221</sup> Natalie Stoljar (2018) “Feminist perspectives on autonomy”, *The Stanford Encyclopedia of Philosophy*, ed. Edward Zalta, available at: <https://plato.stanford.edu/archives/win2018/entries/feminism-autonomy/>

us, are handed down to us by our social environment.<sup>222</sup> In a somewhat similar vein, Nedelsky points out that the construction of language is a social phenomenon, while Seana Shiffrin suggests that interacting with others is a vital part of the process in which we come to understand and develop our ideas. Indeed, partly because of this, she points out that complete social isolation (as experienced for instance by certain prisoners) is akin to torture.<sup>223</sup> Eva Kittay describes human dependency as an inescapable fact of life. Both Kittay and Mackenzie argue that this kind of universal dependency ('vulnerability' in Mackenzie's words) establishes powerful interpersonal duties.<sup>224</sup> Taking a step back and surveying the overall impact of our social environment on our lives, Carolyn Ells et al write that: "people are integrally connected with a social environment marked by economics, politics, ethnicity, gender, culture, and so on. Their identity is formed and shaped by their social environment, as well as their experience of embodiment, interactions with others, and possibilities for a good life".<sup>225</sup> Or as Marilyn Friedman puts it, "selves 'become who they are – their identities, their capacities, their desires – through the relationships in which they participate'".<sup>226</sup> Annette Baier deploys a striking terminology in order to reflect these insights, suggesting that we are all 'second persons'; the successors and heirs to other persons who formed and cared for us, our personality both developed and revealed to us through our social relations.<sup>227</sup> It is with considerations like these in mind that Edward Dove argues that rather than conceptualising persons as isolated monads, we should instead embrace a conception of "the individual in a socially embedded *network of others*".<sup>228</sup>

---

<sup>222</sup> Mackenzie and Stoljar, 'Introduction: autonomy refigured'.

<sup>223</sup> Nedelsky, 'Reconceiving autonomy: sources, thoughts and possibilities'; Seana Shiffrin (2014). *Speech Matters: On Lying, Morality, and the Law* (Princeton: Princeton University Press).

<sup>224</sup> Eva Kittay (1999). *Love's Labour: Essays on Women, Equality and Dependence* (New York and London: Routledge); Mackenzie, 'The importance of relational autonomy and capabilities for an ethics of vulnerability'.

<sup>225</sup> Carolyn Ells, Matthew Hunt and Jane Chambers-Evans (2011). "Relational autonomy as an essential component of patient-centered care", *International Journal of Feminist Approaches to Bioethics* 4 (2), pp. 79-101.

<sup>226</sup> Marilyn Friedman (2013). "Relational autonomy and individuality" *The University of Toronto Law Journal* 63, (2), pp. 327-41. The quote comes in the context of summing up the insights of Jennifer Nedelsky, 'Law's Relations: A Relational Theory of Self, Autonomy, and Law'.

<sup>227</sup> Annette Baier (1985). *Postures of the Mind* (Minneapolis: University of Minnesota Press).

<sup>228</sup> Dove et al, 'Beyond individualism: Is there a place for relational autonomy in clinical practice and research?'. See also: Jennifer Walter and Lane Friedman Ross (2014). "Relational autonomy: moving beyond the limits of isolated individualism", *Pediatrics* 133 (1), pp.16-23.

Of course, what this involves precisely depends on what we mean when we speak of a “network of others”. A somewhat rough and ready list of relevant aspects of our social relations might include: how others think about us personally, how others think about our identity (e.g. race, religion, gender, sexual orientation), expectations and norms surrounding our identity (e.g. social pressure to ‘appear feminine’ or ‘act manly’), how these aspects of our identity are represented in popular culture, what kind of expectations this leads others to have of us, the explicit relationships with others that we form (e.g. friendships, parent-child relationships, sexual relationships, citizen-citizen relations, etc.), the relationships we form with our surrounding institutions. Each of these social aspects will be discussed more below. For now the point is that many relational theorists ask that when we think about persons, we reflect carefully on these interconnected points of their lives.

But relational theories do not end at the descriptive level. If one had to pick a central tenet of relational approaches to autonomous flourishing, it would be that autonomous flourishing is closely, even inextricably, tied up with our social relations – our network of others. Gerald Dworkin bridges the gap between the description and prescription, when he suggests that our accounts of autonomy must accommodate the fact that people are rarely, if ever, fully independent; that very few seek total isolation, and even fewer persevere with it.<sup>229</sup> The previous Ells quote at least gestures towards this idea through its reference to the good life. Martha Nussbaum makes a similar point, when she suggests that focusing on the normative importance of our social relations, reflects the actual lived experience of social persons – what Oshana terms their “flesh and blood functioning in the real world”.<sup>230</sup> So our conception of autonomy should reflect the ways in which the autonomous flourishing of persons depends on their network of others. On this score, two (compatible) suggestions are made; that our network of others has an instrumental role in our autonomous flourishing, and that it has a constitutive role in our autonomous flourishing. Let us clarify both these suggestions.

---

<sup>229</sup> Dworkin, ‘*The Theory and Practice of Autonomy*’.

<sup>230</sup> Martha Nussbaum (2000). “The future of feminist liberalism”, *Proceedings and Addresses of the American Philosophical Association*, 74 (2), pp. 47-79; Marina Oshana (2014). “A commitment to autonomy is a commitment to feminism”, in *Autonomy, Oppression, and Gender*, ed. Mark Piper and Andrea Veltman, pp.141-60.

First, the instrumental view: one often finds in philosophical discussion of autonomy, a focus on rather individualistic questions: am I successfully pursuing my goals?<sup>231</sup> Are these the right kind of goals? Do I have an adequate range of options?<sup>232</sup> Am I able to conceive, revise, and pursue my conception of the good?<sup>233</sup> Am I happy?<sup>234</sup> Am I living a suitably dignified life?<sup>235</sup> Am I living in accordance with my own reflective values?<sup>236</sup> Were these values freely chosen in the absence of manipulation and other such liberty-impairing influences?<sup>237</sup> Are they the product of an appropriately critical reflective process?<sup>238</sup> Do I have an adequate sense of self-worth?<sup>239</sup> Do I have a suitable range of functioning capabilities?<sup>240</sup> Do I possess, for instance, bodily health, bodily integrity, an adequate range of emotional capabilities, a sense of self-respect, the ability to pursue play and leisure activities?<sup>241</sup> These are all questions that relate, on the surface, to how I am doing, quite apart from my surrounding network of others.

Relational approaches to autonomy hold that our network of others plays an important instrumental role in these aspects of our autonomous flourishing. That is to say, our surrounding network of others can both promote and hinder the aforementioned individualistic aspects of our flourishing. Consider the successful pursuit of goals and the living of a life in accordance with our own deep values. Our network of others can help us revise and reflect upon our goals, as when there are opportunities to take part in meaningful debate about what goals to have and our ideas are listened to and taken seriously. Similarly, forming

---

<sup>231</sup> Andrea Westlund (2009). "Rethinking relational autonomy", *Hypatia* 24 (4), pp.26-49.

<sup>232</sup> Raz, 'The Morality of Freedom'.

<sup>233</sup> Rawls, 'A Theory of Justice'.

<sup>234</sup> Jeremy Bentham (1770) [1789]. *An Introduction to the Principles of Morals and Legislation* (London: Athlone Press), ed. James Henderson, H.L.A. Hart.

<sup>235</sup> Mill, 'On Liberty'.

<sup>236</sup> John Christman (2013). "Autonomy", in *The Oxford Handbook of the History of Ethics* (Oxford: Oxford University Press), ed. Roger Crisp, pp.281-93.

<sup>237</sup> John Christman, 'Autonomy and personal history'.

<sup>238</sup> Westlund, 'Rethinking relational autonomy'.

<sup>239</sup> Paul Benson (1994). "Autonomy and self-worth", *Journal of Philosophy* 91 (12), pp.650-68; Trudy Govier (1993). "Self-trust, autonomy and self-esteem", *Hypatia* 8 (1), pp.99-119.

<sup>240</sup> Amartya Sen (1985). *Commodities and Capabilities* (Amsterdam and New York: North-Holland); Amartya Sen (1989). "Development as capability expansion", *Journal of Development Planning* 19, pp.41-58; Amartya Sen (2009). *The Idea of Justice* (London: Allen Lane).

<sup>241</sup> Nussbaum 'Women and Human Development: The Capabilities Approach'; Nussbaum, 'Capabilities as fundamental entitlements: Sen and social justice'.

friendships and healthy working relationships can, as Shiffrin has argued, help the development and maintenance of our minds and personalities, as well as helping us to realise and meet our duties to others.<sup>242</sup> At other times our network of others can be a liability. Research suggests that exposure to racial hatred, discrimination, and stigmatised cultural images, for instance, is correlated not only with lower self-esteem, but also expectations, anxieties, and reactions that can adversely affect one's psychological wellbeing as well as result in decreased motivation for socioeconomic attainment.<sup>243</sup> Indeed, this is just part of the damaging consequences of what has been termed 'internalised-racism'. We should also be mindful that sometimes it can be extremely difficult to know what decision will best reflect our deep values, as when one is under significant emotional stress or there is a lack of certainty about the consequences of one's decision. In these cases, social support may be particularly important. For instance, Tineke Abma and Kaja Heidenreich (and their co-authors) have convincingly argued that patients need emotional support from their clinicians in order to help them make decisions that are true to their own deeply held values.<sup>244</sup>

Or consider the development of our emotional capacities and maintenance of emotional well-being. Sarah McKenzie finds that social connections can act as a buffer against the impact of stressful life events.<sup>245</sup> These social relations also increase the likelihood of those with mental health problems, seeking out professional help.<sup>246</sup> Similarly, partaking in caring,

---

<sup>242</sup> Shiffrin, 'Speech Matters: On Lying, Morality, and the Law'.

<sup>243</sup> David Williams (2018). 'Stress and the mental health of populations of color: advancing our understanding of race-related Stressors', *Journal of Health and Social Behaviour* 59 (4), pp.466-85; Naa Oyo Kwate and Ilan Meyer. (2011). "On sticks and stones and broken bones: stereotypes and African American health", *Du Bois Review: Social Science Research on Race* 8 (1), pp.191-8; Naomi Priest, Yin Paradies, Brigid Trenerry, Mandy Truong, Saffron Karlsen, and Yvonne Kelly (2013). "A systematic review of studies examining the relationship between reported racism and health and wellbeing for children and young people", *Social Science and Medicine* 95, pp.115-27.

<sup>244</sup> Tineke Abma, Barth Oeseburg, Guy Widdershoven, and Marian Verkerk (2009). "The quality of caring relationships", *Psychology research and behavior management*, 2, pp.39-45; Kaja Heidenreich, Anders Bremer, Lars Johan Materstvedt, Ulf Tidefelt, Mia Svantesson (2018). "Relational autonomy in the care of the vulnerable: health care professionals' reasoning in moral case deliberation", *Medicine, Health Care and Philosophy* 21, pp. 467-77.

<sup>245</sup> Sarah McKenzie, Sunny Collings, Gabrielle Jenkin, and Jo River (2018). "Masculinity, social connectedness, and mental health: men's diverse patterns of practice", *American Journal of Men's Health*, pp.1247-126; Cecilia Aslund, Peter Larm, Bengt Starrin, Kent Nilson (2014). The buffering effect of tangible social support on financial stress: influence on psychological well-being and psychosomatic symptoms in a large sample of the adult general population", *International Journal for Equity in Health* 13 (85), pp.1-9.

<sup>246</sup> McKenzie et al, 'Masculinity, social connectedness, and mental health: men's diverse patterns of practice'; Sarah Andrea, Sarah, Siegel, Alan Teo (2016). "Social support and health service use in depressed adults: findings from the national health and nutrition examination survey", *General Hospital Psychiatry* 39, pp.73-9.

supportive relationships has been found to foster the development of a robust sense of self-worth, which can play a key role in helping to cope with stressful events.<sup>247</sup> By contrast, a lack of social support, including a lack of understanding and exposure to belittlement, has been linked to poor mental health.<sup>248</sup> Related to this issue, social norms around gender are another key component of our network of others, that can have a significant impact upon our emotional well-being. Daniel Coleman et al find that traditional masculinity is a key driver of suicide in men, emphasising aggressiveness and emotional restriction as potentially key contributors in this respect.<sup>249</sup> Perhaps part of what explains this is that, as David Vogel et al have found, traditional masculine norms attach a stigma to seeking psychological help.<sup>250</sup> McKenzie suggests that the ‘unwritten boundaries’ and norms of hegemonic masculinity can prevent men from forming emotionally supportive relationships with other men – although she emphasises that by no means do all men fit into this description. McKenzie suggests that when this isolation lasts for years, it leads to feelings of shame and failure. Katelyn Sileo and Trace Kershaw find that men who conform to anti-feminine norms often struggle with anxiety control, which can have long term negative implications for psychological well-being.<sup>251</sup>

Several philosophers focus on the harmful impact that oppression and identity-based hatred can have on our psychological and emotional well-being. Mari Matsuda, for instance, drawing on the research of the likes of Harry Haruo Kitano, finds that racial prejudice is

---

<sup>247</sup> Vasudha Gupta and Charles Korte (1994). “The effects of a confidant and a peer group on the well-being of single elders”, *The International Journal of Aging and Human Development* 39 (4), pp.293-302; Jennifer Thomas and Kimberly Daubman (2001). “The relationship between friendship quality and self-esteem in adolescent girls and boys”, *Sex Roles* 45 (1), pp.53-65; Michelle Harris and Ulrich Orth (2020). “The link between self-esteem and social relationships: a meta-analysis of longitudinal studies”, *Journal of Personality and Social Psychology* 119 (6), pp.1459-77; Andrea Gruenenfelder-Steiger, Michelle Harris, Helmut Fend (2016). “Subjective and objective peer approval evaluations and self-esteem development: a test of reciprocal, prospective, and long-term effects”, *Developmental Psychology* 52 (10), pp.1563-77.

<sup>248</sup> McKenzie et al, ‘Masculinity, social connectedness, and mental health: men’s diverse patterns of practice’.

<sup>249</sup> Daniel Coleman, William Feigelman, and Zohn Rosen (2020). “Association of high traditional masculinity and risk of suicide death: secondary analysis of the add health study”, *JAMA Psychiatry* 77(4), pp.435-37; Daniel Coleman (2015). “Traditional masculinity as a risk factor for suicidal ideation: cross-sectional and prospective evidence from a study of young adults”, *Archives of Suicide Research* 19 (3), pp.366-84.

<sup>250</sup> David Vogel, Sarah Heimerdinger-Edwards, Joseph Hammer, and Asale Hubbard (2011). “Boys don't cry”: examination of the links between endorsement of masculine norms, self-stigma, and help-seeking attitudes for men from diverse backgrounds”, *Journal of Counselling Psychology* 58 (3), pp.368-82.

<sup>251</sup> Katelyn Sileo and Trace Kershaw (2020). “Dimensions of masculine norms, depression, and mental health service utilization: results from a prospective cohort study among Emerging adult men in the United States”, *American Joournal of Men’s Health* 14 (1), pp.1-16.

correlated with displaced aggression, avoidance, retreat, withdrawal, alcoholism, and suicide.<sup>252</sup> Indeed, exposure to discrimination and identity-based hate has also been found by a number of studies to correlate with stress and anxiety.<sup>253</sup> While there is also evidence that discrimination and exposure to racial hatred are linked to depression, even when individual personality differences are controlled for.<sup>254</sup> As we have seen elsewhere in this thesis, exposure to identity-based hatred is correlated with ‘internalised racism’ and self-hatred.<sup>255</sup> In a similar vein, José Medina argues that social oppression can lead to a lack of self-confidence and epistemic self-trust.<sup>256</sup> Many relational theorists take epistemic self-trust to be a necessary condition of autonomy. Mackenzie argues that autonomous agents possess a suitably confident self-conception. Autonomy, she suggests, involves identifying as a legitimate source of authority, able to speak for one’s own self. It is only through relations with others of a certain kind, Mackenzie supposes, that we can sustain this robust sense of ‘normative authority’.<sup>257</sup>

These points tease out the perfectionist quality latent in many relational accounts. This point has been touched upon already, but it is worth making explicit: flourishing autonomous agents, so the relational theorist claims, don’t just live in accordance with their values and pursue their goals; they have the ‘right’ kind of values and pursue the ‘right’ kind of goals.<sup>258</sup> This might involve holding that autonomous agents have a certain degree of self-respect, or are enjoying a certain kind of interpersonal dignity. It might involve viewing autonomous agents as refusing to pursue humiliating goals and demeaning values, or refusing to systematically defer

---

<sup>252</sup> Matsuda, ‘Public response to racist speech: considering the victim’s story’. See also: Matsuda, ‘*Words That Wound: Critical Race Theory, Assaultive Speech, and the First Amendment*’.

<sup>253</sup> Williams, ‘Stress and the mental health of populations of color: advancing our understanding of race-related stressors’; Tené Lewis, Courtney, Cogburn, and David Williams (2015). “Self-reported experiences of discrimination and health: scientific advances, ongoing controversies, and emerging issue”, *Annual Review of Clinical Psychology* 11, pp.407-40; Brown, ‘*Hate Speech Law: A Philosophical Examination*’.

<sup>254</sup> Frederick Gibbons, Meg Gerrard, Michael Cleveland, Thomas Wills, and Gene Brody (2004). “Perceived discrimination and substance use in African American parents and their children: a panel study”, *Journal of Personal Social Psychology* 86 (4), pp.517-29.

<sup>255</sup> Williams, ‘Stress and the Mental Health of Populations of Color: Advancing Our Understanding of Race-related Stressors’; Delgado, ‘Words that wound: a tort action for racial insults, epithets, and name-calling’.

<sup>256</sup> José Medina (2013). “Color blindness, meta-ignorance, and the racial imagination”, *Critical Philosophy of Race* 1 (1), pp.38-67; Lynne Steinberg (2010). “When friends make you blue: the role of friendship contingent self-esteem in predicting self-esteem and depressive symptoms”, *Personality and Social Psychology Bulletin* 36 (3), pp.384-97.

<sup>257</sup> Mackenzie, ‘Relational autonomy, normative authority and perfectionism’.

<sup>258</sup> Marina Oshana (1998). “Personal autonomy and society”, *Journal of Social Philosophy* 29 (1), pp.81-102.

to others for no good reason. With this in mind, the crucial question then becomes whether our network of others – our cultural representations, our gender expectations, the relationships we find ourselves in, etc. – promote this perfectionist ideal.

It is along these lines, that relational theorists often discuss relationships that involve stark inequality, self-abnegation, and deference. Consider, for instance, the classic case of the overly deferential housewife who ‘voluntarily’ defers to her husband on virtually all matters; she dresses how he wants her to dress, cooks what he wants to eat, modulates her diet in accordance with his wishes, puts her goals on hold so that she might better serve his.<sup>259</sup> Part of what has gone so wrong in this case, is the way in which this highly asymmetric relationship likely damages the housewife’s sense of self-trust and self-respect. A related worry is that the housewife is not attending to her own long-term goals and needs, and that the relationship in which she finds herself in fact makes it harder for her to do this.<sup>260</sup> Indeed, there may also be a deep sense in which, despite the absence of physical threat or harm, and the fact that the housewife ostensibly volunteers to be in this relationship, the striking deference and psychological subordination involved qualifies the relationship as genuinely coercive.<sup>261</sup> Put simply, in various ways this picture of self-abnegation and striking deference, is incompatible with the perfectionist vision of autonomy outlined by many relational theorists.

One might worry that this objection to overly deferential relationships, actually points to a contradiction in the relational approach. After all, many relational theorists emphasise the importance of entering into caring relations with others, part of which involves making sacrifices and taking on burdens for one another. Indeed, a striking instance might be that of a mother making great sacrifices to take care of her seriously ill child. Is there a way of defending the mother’s sacrifices in this case, while nonetheless holding that something has gone deeply wrong in the deferential housewife case? Two disanalogies between these cases are striking. The first is that while the mother would seem to have a powerful and rational reason to make sacrifices – her child really needs these sacrifices to be made – such reasons are absent in the

---

<sup>259</sup> A modification of Thomas Hill’s classic case: Thomas Hill (1991). *Autonomy and Self-Respect* (Cambridge: Cambridge University Press).

<sup>260</sup> Oshana, ‘Personal autonomy and society’.

<sup>261</sup> Laura Richards has spent years drawing research together and developing the concept of coercive control. See for instance: <https://www.laurarichards.co.uk/coercive-control/> and [www.laurarichards.co.uk/domestic-abuse-and-coercive-control/](https://www.laurarichards.co.uk/domestic-abuse-and-coercive-control/). See also: <https://www.cps.gov.uk/legal-guidance/controlling-or-coercive-behaviour-intimate-or-family-relationship>

housewife case. The housewife makes inordinate sacrifices for her husband, simply because he wants her, or expects her, to make these sacrifices. So understood, we see that the housewife is devaluing herself, insofar as she takes the trivial wants and needs of her husband to have outsized importance compared to her own wants and needs. Indeed, it is precisely because the housewife responds to her husband with the same devotion and self-sacrificing attitude as that of a mother attending to her seriously ill child, that we see her situation as so troubling. The other important disanalogy, is that while the mother may well sacrifice her time and energy and goals in the name of caring for her child, she nonetheless retains the kind of normative, epistemic authority that philosophers like Mackenzie and Govier think is so important. Whatever else we might say about the mother, unlike the deferential housewife, she retains a sense of trust in her own competency to decide what is in her and her son's best interests.

Moving on now to the second way in which an account of autonomous flourishing must be sensitive to the importance of our "network of others"; the suggestion that our network of others is a constitutive part of our flourishing as autonomous agents. This view does not deny that the more individualist aspects, touched upon above, are important to our autonomous flourishing, or that our network of others can play a key instrumental role in promoting these aspects. But the constitutive view goes one step further in assuming that what it means to be a fully autonomous agent is to have a network of others of a certain kind. Or, as Oshana puts it, relations of a certain kind are mandatory if agents are to count as genuinely self-determining.<sup>262</sup> In what way might we consider our network of others to be a constitutive part of our autonomous flourishing? Several (compatible) ideas present themselves.

One somewhat complicated idea is that even when our network of others promotes various aspects of our autonomous flourishing, such that it could be classified as a useful tool for promoting our flourishing, this talk of tools and instrumental value in fact misses a key insight. The idea runs that persons are so thoroughly relational, that their surrounding network of others – their friendships, families, social standing, etc – is *the* primary tool used by agents to flourish. This primacy speaks to the idea that our network of others plays a constitutive role in our autonomous flourishing. We might explore this idea by reflecting on

---

<sup>262</sup> Marina Oshana, 'A commitment to autonomy is a commitment to feminism'.

cases where our network of others plays a key positive role in promoting some individualistic aspect of our flourishing. But what about the opposite case? What about when some individualistic aspect of an agent's flourishing goes well, even though a supportive network of others is absent? In such a case, it is notable that agents keenly feel this absence, and must take great lengths to make up for losing the primary way in which persons promote their autonomous flourishing. Consider, for instance, just how difficult it would be to develop one's emotional capacities in total isolation, and the tremendous amount of work that it would take to overcome this obstacle. Modulating one's emotional responses, without the input of others, would seem to require an almost inhuman effort. Similarly, it may be possible to develop a robust sense of self-worth while being systematically exposed to identity-based hate or placed in humiliating conditions. But clearly special efforts are needed in these cases – not to mention likely an increased dependence on other aspects of one's network of others.

These cases tease-out the idea that even when our network of others functions as a kind of tool for promoting flourishing, it is a tool like no other. Indeed, so much so, that it would be more accurate to say that our network of others is constitutively valuable. An analogy might be drawn here with the idea that persons are 'perceptual beings'. The meaning of this is not simply that persons use their perceptual faculties to come to understand the world – although that is undoubtedly part of the picture. Rather, the key point is that our perceptual faculties are the most fundamental and important tools we use to understand the world. In other words, we are perceptual beings, through and through.<sup>263</sup> The analogous point runs that we are also relational beings, through and through. And that if our network of others is indeed an instrument we use to (e.g.) cultivate our emotional capacities or develop rationally chosen goals, it is an instrument we could hardly do without.

A more straightforward way in which our network of others counts as a constitutive aspect of our autonomous flourishing, runs as follows: we have a special need for social relations of a certain kind, and thus our autonomous flourishing is set-back when our social needs do not have this required character. This is especially clear when we reflect on what we might call, our need for social relations that include human warmth: that is to say, social relations that include friendship and affection. Monika Betzler describes how often the best

---

<sup>263</sup> We thank Douglas Lavin for suggesting this analogy.

relationships involve ‘intimacy, mutual concern and care, and trust’.<sup>264</sup> But Betzler’s point here isn’t that such relationships are a useful tool for meeting related needs. On the contrary, it would seem that many (if not, all) people have a direct need for such interpersonal warmth. Perhaps this is why, as Kathleen Lynch et al found, for many adults, forming and maintaining close relationships are the most important part of their overall life plan.<sup>265</sup> Their findings suggest that caring relationships aren’t merely useful as a tool for meeting some more important need or achieving some more important goal, but rather that we have a direct need for such relationships themselves. The idea here is that my friendship with my next-door neighbour, on top of being a useful way of finding out what’s going on in the neighbourhood and helping to keep my household secure from thieves, also meets a need I have to form warm social relations. Anca Gheaus similarly suggests that most, if not all, people have a need to take part in loving relationships in which they are valued.<sup>266</sup> Rather than this kind of ‘loving care’ being merely a tool for promoting our emotional well-being, Gheaus takes it to be itself a constitutive part of our well-being. Abma et al agree. Writing in the context of bioethics, they stress the importance of doctors providing patients with a supportive space in which to share their emotions and feelings. Allowing patients to express themselves in this way may well be instrumentally valuable insofar as it e.g. helps them to make rational decisions about what is in their best interests. But it is also constitutively valuable, the authors argue, insofar as it helps patients meet their relational need to share painful experiences with others and feel as though they are being truly listened to.<sup>267</sup> Dove et al, writing in the same context, emphasise the value of relationships that involve mutual support, love, dependency and trust.<sup>268</sup> Overall, the key point being made here isn’t that having caring social relations is merely constitutively valuable.

---

<sup>264</sup>Monika Betzler (2019). “The relational value of empathy”, *International Journal of Philosophical Studies* 27 (2), pp.136-61; Anca Gheaus (2009). “How much of what matters can we redistribute? Love, justice, and luck”, *Hypatia* 24 (4), pp.63-83.

<sup>265</sup> Kathleen Lynch (2009). “Affective equality: who cares?”, *Development* 52 (3), pp.410-15.

<sup>266</sup> Gheaus, ‘How much of what matters can we redistribute? Love, justice, and luck’.

<sup>267</sup> “Giving space to suffering implies that one deliberately refrains from acting. The focus shifts from instrumental values to the intrinsic values of attentiveness, being present and being related to the patient. This is hard to practice for those health care professionals who define their identity in terms of being an expert. Creating a space for emotions and feelings of patients requires that one redefines one’s self-image as a health care professional. Being a good health care professional is not only a matter of technical expertise, but includes the moral virtues of being there and listening, responding to and taking part in the life story of the patient as it unfolds”: Tineke Abma et al, ‘The quality of caring relationships’. Also see: Tineke Abma (2005) “Struggling with the fragility of life: a relational-narrative approach to ethics in palliative nursing”, *Nursing Ethics* 12 (4), pp.337-48.

<sup>268</sup> Dove et al, ‘Beyond individualism: is there a place for relational autonomy in clinical practice and research?’.

On the contrary, clearly such relations are instrumentally valuable insofar as they help us e.g. develop our capacities, maintain self-esteem, meet our goals, etc. But in addition to this, it seems that these relations also form a constitutive part of our autonomous flourishing, insofar as we humans simply desire this kind of closeness.

Overall then, relational theorists suppose that our social relations (our ‘network of others’) plays both an important instrumental and constitutive role in our autonomous flourishing. In some ways, the key underlying thought here is Aristotelian: we learn about what it means for a person to flourish as an autonomous agent, by reflecting on the nature of that autonomous agent. Philosophers who endorse a relational approach to autonomy hold that autonomous agents are fundamentally social creatures. It is not surprise then, that what follows from this insight, is that our flourishing is bound up with our social relations. What perhaps is somewhat surprising, is just how tightly bound up our flourishing is with our social relations; for not only do our social relations turn out to be instrumentally important for meeting our various needs and developing our capacities, they are also constitutively important insofar as we have a direct need for social relations of a certain character – e.g. relations that are caring. Why is this important? How does this relate to a thesis on censorship and free speech? Shortly we will propose a new account of the way having social relations of a certain kind, is constitutively valuable – one that has important implications for freedom of speech and censorship. Our proposal, which we discuss in relation to Republican accounts of freedom, revolves around the idea that persons have a need for social relations that are absent of degrading, hierarchical features.

## §II. Persuasion-control and relational accounts

Taking a step back and considering the relational needs of citizens, it is an open question whether relational philosophers should support or oppose persuasion-control. Broadly speaking there has been a tendency for relational theorists to emphasise the potential upside of censorious regulations like persuasion-control. There are at least three reasons for this. First, relational theorists tend to be optimistic about the idea that censorious regulation might promote autonomy. In other words, they recognise that censorious interferences can help promote a network of others that better meets the relational needs of citizens. In practice this involves reflecting upon the autonomy-undermining aspects of our network of others –

prominent disrespectful cultural representations, damaging social stereotyping, unequal levels of respect towards certain groups – and framing censorious regulations as having the power to change these aspects and render them more supportive of our autonomous flourishing. Or, at the very least, regulation might render these social aspects of our life less corrosive of autonomy.<sup>269</sup> Second, when relational theorists consider someone like the deferential housewife, they are far more likely than say proponents of procedural theories of autonomy, to judge that this person’s decisions lack autonomy.<sup>270</sup> This is important because those opposing regulation often seek to paint censorship as interfering with fully autonomous agents, and therefore the regulation itself as a threat or insult to their autonomy. At the very least, the more nuanced understanding of autonomy presented by relational theorists tends towards a less alarmist view of regulation.<sup>271</sup> Third, relational philosophers tend to reject the liberal ideal of the self-sufficient individual free from outside interference, and the almost neurotic libertarian rejection of governmental regulation that goes along with it.<sup>272</sup> In this sense, relational theorists adopt a relatively open attitude towards government regulation, so long as that regulation meets our relational needs.

The point is well taken. Still, at the same time, the relational account of autonomy may provide us with the resources to reanimate a certain kind of worry about the threat that persuasion-control poses.<sup>273</sup> In the next section we present what we call the *relationship account*.

---

<sup>269</sup> Brison ‘The autonomy defence of free speech’; Susan Brison (2001). “Contentious freedom: sex work and social construction”, *Hypatia* 21 (4), pp.192-200.

<sup>270</sup> Oshana, ‘Personal autonomy and society’.

<sup>271</sup> Mackenzie and Stoljar, ‘Introduction: autonomy refigured’.

<sup>272</sup> It may be countered that we are drawing too sharp a distinction between relational approaches to autonomy, with their focus on the social nature of persons, and allegedly more individualist-minded liberal theories. After many liberal theories in political philosophy in fact embrace the idea, at least to some extent, that we ought to care about people’s social relations. The point is well taken. All the same, we refer the reader to our discussion in §1, of the ‘instrumental view’. There we give a list of popular approaches to well-being and autonomy that focus too narrowly on what we identify as individualist aspects of our autonomous flourishing, and not enough on the more social aspects. Nonetheless, we agree that in many cases, especially when dealing with liberal theories in political philosophy, that these theories under-emphasise, rather than ignore entirely, the importance of the social aspects of flourishing.

<sup>273</sup> Some, like John Christman, might worry that this kind of relational account risks silencing or excluding or even re-victimising those persons it judges to be lacking in autonomy. What if a person claims to enjoy their demeaning friendship, or their subordinate marital relationship, or their problematic relationship with the government, or the humiliating gender norms imposed upon them? Does my account claim that we must dismiss this person’s voice? See: John Christman (2004). “Relational autonomy, liberal individualism, and the social constitution of selves”, *Philosophical Studies* 117 (1-2), pp.143-64.

The account focuses on the government-citizen relationship, and what kind of position citizens occupy within this relation, relative to those in political power. It is argued that permitting the government to engage in persuasion-control changes and indeed damages the government-citizen relationship – placing citizens in an even more subordinate position before government members, than they were already. This subordination is a result, not of any underlying message of persuasion-control, nor of the material restrictions it imposes on our autonomous capacities, but rather taking part in an important and inescapable relationship that demands intellectual deference. It is from this standpoint that we can best understand the deep-seated Scanlonian anxiety many have with persuasion-control. Moreover, our discussion of relational autonomy provides us with the tools to better understand why agents find such subordinating relationships to be objectionable: the answer being, in short, that as social beings we are sensitive to the social hierarchies we find ourselves in.

### §III. The relationship account

The key question at hand then, is: how might permitting government censorship change the government-citizen relationship in a way that is objectionable? In particular we might ask how the position of citizens relative to members of the government changes, such that citizens find themselves in a starkly subordinate position? There are *two parts* to our answer.

*First*, let us start with the observation, touched upon in chapter 2, that allowing the government to engage in censorship grants the government a special kind of authority over its citizens. Persuasion-control involves the government attempting to tell its citizens what to think; altering what kind of persuasive materials citizens have access in their environment, in an attempt to steer them away from developing dangerous beliefs and intentions. In this sense,

---

In response, we will reiterate what has been said elsewhere in this thesis. While Christman is right to suppose that we should not ignore or dismiss the voice of citizens, willy nilly. And that we should be duly anxious about labelling citizens as non-autonomous. Nonetheless, we should not satisfy ourselves with thinking that the question of whether someone is fully autonomous or living her best kind of life is settled simply by her own self-assessment. Whether we like it or not, we are not always the best judge of what is in our own interests. We all make mistakes when it comes to objectively appraising our own autonomous condition. Indeed, Christman's own broadly procedural approach to autonomy says just as much. After all, Christman thinks that agents lack autonomy when their desires are the product of non-consensual hypnosis, whether those agents agree with this assessment or not. Sometimes, then, we will have to look beyond a person's self-appraisal in order to gain a full picture of whether she is autonomous. We do not think this should strike us as overly dismissive or insulting. On the contrary, it speaks to the complexity of the issue at hand and indeed of people.

the government installs itself as an arbiter over not simply how citizens should act, but indeed how they should think and what kind of materials they may reflect upon.

Why might we find this objectionable? The answer has to do with the position or standing of citizens, relative to those in political power. We will now present our answer as a kind of relational theory of flourishing, before then discussing how one might alternatively construe our account as a modified republican theory of freedom.

The core idea of our relationship account is that, as social beings, we care about our position within our surrounding social hierarchy. We are acutely sensitive to the social relations we have with others, and in particular the dynamics of power within those relations. In other words, when we are evaluating something like persuasion-control, we must reflect not only on how persuasion-control impacts upon the more individualistic aspects of our flourishing (e.g. whether it coerces us or substantially undermines our ability to use our rational faculties), but also how it impacts upon our relationships and the social hierarchies in which we find ourselves. This has very important implications for evaluating persuasion-control. As argued elsewhere in this thesis, persuasion-control does not necessarily amount to coercion, thought control, or even a particularly significant interference with our abilities to reason. Likewise, persuasion-control needn't involve an insulting message about the incompetence of citizens, nor involve the government engaging in viewpoint discrimination. Rather, what is worrying is the kind of relationship between government and citizen that persuasion-control establishes. As noted above, to grant the government the power to engage in persuasion-control, is to establish them as an arbiter over what materials we have access to, and what kind of beliefs and intentions we may hold. The flipside of the government possessing this kind of authority, is that citizens are placed in the position of intellectual subordination and deference. It is this stark intellectual subordination and deference before members of the political authority, that renders persuasion-control objectionable from the standpoint of relational autonomy. As we will argue below, standing in such a subordinated position, leaves one unable to 'look members of the political authority in the eye'. It is *this* aspect of persuasion-control, that renders it objectionable. Or, to put it another way, it is hierarchical nature of the relationship that persuasion-control involves, rather than say its implicit message or its alleged unfairness, that should worry us.

To help capture the intuition behind this first suggestion, it will be productive to reflect upon three cases. The purpose of these cases is to consider the normative implications of intellectual subordination, and how such subordination damages our social relations. In the first case, we imagine that I have been living for the past ten years in a house with my wife. One day I am in the attic, when I come across a strange cardboard box containing leaflets and pamphlets from local religious groups. I ask my wife where this box has come from, and after some cajoling she admits that these are documents that have, over the years, been delivered through our letterbox, but that she has been intercepting these documents so as to keep them out of my sight. When I ask her why, she says that while I am generally speaking likely to find these documents unpersuasive and not my taste, she has found that I am quite liable when I am very stressed, to act in quite impetuous and unwise ways. In particular she suspects that, when I am suitably stressed, I would likely be swayed by these kinds of propagandistic materials. Thus, out of concern that I might some day come across one of these documents and be duped by it, she has been efficiently intercepting and hiding them from me. We take it that, as a matter of intuition, readers will find the wife's behaviour to be highly objectionable. Admittedly, there may be a number of reasons for this. What we would like to focus on, is how the wife's behaviour changes, and ultimately significantly damages the relationship she has with her husband. In the end what is so damaging, is that the wife installs herself as the intellectual authority in the relationship, and concomitantly I (the husband) now occupy a position of intellectual subordination. Note that, in terms of my individual well-being, my wife's behaviour has had little, if any kind of, negative effect. We can imagine, for instance, that had I in fact come across these documents, then I would have, contrary to my wife's assumptions, rejected them. We could even alter the example so that my wife's action in fact saved me a good deal of pain that would have resulted from my impetuously joining a religion. Nonetheless, we suppose that the wife's behaviour dramatically changes, and in fact possibly even breaks, the relationship. Why might this be? Because, as people, we are acutely sensitive to the hierarchy of the relationships we find ourselves in. Indeed, so much so that our autonomous flourishing is partly constituted by those hierarchies.

The second case involves the government suppressing crime statistics for a certain area, because it is worried that should the local citizens find out about how high the crime rate is, they will be sent into a state of acute anxiety and panic. The idea here is that the noble lie that

'local crime isn't that bad' is preferable to the stress and fear that accompanies finding out the truth. While this case is very close to that of persuasion-control, there are some disanalogies worth drawing out. For instance, rather than the government suppressing persuasive appeals, this case rather involves the suppression of a straightforward disclosure of the facts. Moreover, in this case the government attempts to protect citizens from harming themselves through stress, whereas our concept of persuasion-control has tended to focus more so on cases where the government prevents persons from developing beliefs and intentions that harm others. Finally, the crime statistics case may be special insofar as failing to disclose the facts might endanger citizens. After all, it is possible that armed with the facts about how dangerous their local area is, citizens would be more likely and more able to take defensive precautions – although note that this is a contingent, rather than necessary fact about the case.

These differences notwithstanding, it is interesting to note just how intuitively objectionable the government's actions are in the crime statistics case. We assume that virtually all readers will find the government's actions in this case to be quite outrageous.<sup>274</sup> One reason for this, is that many of us do not wish for the government to act as arbiter over what facts we can and cannot handle. Granted, the government exercises many forms of authority over us. Perhaps we should simply say that its claiming of intellectual authority over us in this case is at a piece with other governmental interferences in our lives, such as setting speed limits and demanding tax contributions. However, this would be a mistake. There is something distinctive when the government assumes the mantle of arbiter over which aspects of reality are too shocking for us to handle. In such a case, it claims authority over us on a matter so integral to what it means to be an autonomous agent, as what facts about the world are 'too much' for us to handle. We would not be distributed by having to stand before another person – in this case members of the political authority – as someone who must defer to them on so central a matter?

For the third case, we offer a modern case of persuasion-control. Through research, the government comes to learn that anti-vaccination speech, especially on online platforms, significantly increases vaccination hesitancy. Indeed, the predictable increase in vaccination hesitancy is high enough, that it will plausibly lead to innocent persons being harmed, as lower

---

<sup>274</sup> While falling well short of an empirical investigation of the case, we note that this at least has been the response I've encountered when discussing this case with others.

rates of vaccination are related to higher rates of virus transmission, (thus) hospitalization, and even virus mutation. Looking at all the research, the government with some justification, decides to ban the publication of anti-vaccination speech on online platforms, unless the speech in question has been peer-reviewed and published in a scientific journal. Why might we find such an act of effective censorship to be objectionable? Hopefully by now it is clear that in such a case, the government installs itself not simply an authority that can physically coerce us, but as an intellectual authority that can tell us what we may and may not believe. This places citizens in a striking position of subordination before those who occupy political power. The key insight we have drawn from the literature on relational autonomy is that we have a deep interest in how we stand, relative to other people. To stand before government officials as subjects who may be told what to think, is no small matter. We suggest that it is one thing to, within limits of course, be required to defer to the government's judgement about how we may act. But there is an even deeper level of subordination involved in standing before the government as a person who may be told what thoughts they may entertain.

Thus far, then, we have explored the relational account as an off-shoot of the relational approach to autonomous flourishing. For it is by reflecting upon the literature and research surrounding the social nature of persons, that we learn that persons are acutely sensitive to the social hierarchies in which they find themselves. However, the relationship account can alternatively be finessed as a kind of modified Republican account of freedom. According to Republican accounts, civic freedom involves being free from the domination of others. Crucially, domination is defined as arbitrary interference.<sup>275</sup> What constitutes 'arbitrary' interference is a matter of much debate, but generally speaking non-arbitrary interference in someone's life, is interference that they have some say of the matter in. For instance, it is often taken that government interference in our lives is not 'arbitrary', so long as it proceeds from a suitably democratic process – the key idea here being that democratic processes give power back to those people who are being interfered with. As Danielle Wenner puts it: “[d]emocratically-imposed legal restrictions are, on this view, controlled by those upon whom they are imposed”,

---

<sup>275</sup> Phillip Pettit (1997). *Republicanism: A Theory of Freedom and Government* (Oxford: Clarendon Press); Arthur Ripstein (2009). *Force and Freedom: Kant's Legal and Political Philosophy* (Cambridge, Massachusetts: Harvard University Press).

and thus do not qualify as domination.<sup>276</sup> Taking a step back for a moment, the core Republican thought is that when we move from a state of nature into a political community, it is necessary that we create non-dominating relations with others. This, of course, is no mean feat, considering not only how liable agents are to exercise arbitrary power and control over others, but also that in a political community certain persons are granted explicit political power over others.

Underlying all of this, but which we have chosen not to yet mention, is the idea that non-domination allows agents to, as Phillip Pettit, puts it, be “able to walk tall amongst others and look any in the eye without reason for fear or deference”.<sup>277</sup> In other words, Republicanism is, at least in some sense, concerned with creating relations of a certain kind between citizens in a political community. More to the point, it associates domination and loss of freedom with the existence of a subordinating social hierarchy. Indeed, in recent times, Pettit has been rather explicit on this topic. He writes, for instance, of “the eyeball test, which figures prominently in the Republican tradition”, that demands that whether citizens, “however they differ otherwise should be able to look one another in the eye without reason for fear or deference”.<sup>278</sup> In this sense, our relationship account could be construed as proceeding from the Republican tradition. After all, at its core, our account is concerned with social hierarchy, and that ordinary citizens are able to stand before members of the political elite, eye to eye, without having to bend or scrape before those in political authority.

At this point we wish to highlight two general problems with the Republican account of freedom. We suggest that both these problems can be remedied by introducing insights from the school of relational theories of autonomy. Thus, in the end, it is argued that our relationship account may be read as a kind of modified Republican account, that draws fruitfully on the work of relational theorists. Those two problems then: first, while the Republican approach professes a concern with whether citizens can look one another in the eye, it underestimates just what is required for this to take place. Granted, John’s exercising

---

<sup>276</sup> Danielle Wenner (2020). “Nondomination and the limits of relational autonomy”, *International Journal of Feminist Approaches to Bioethics* 13 (2), pp.28-48.

<sup>277</sup> Philip Pettit (2012) *On the People's Terms: A Republican Theory and Model of Democracy* (Cambridge: Cambridge University Press), pp.7-8. Also, see p.82.

<sup>278</sup> Mark Haugaard and Philip Pettit (2017). “A conversation on power and republicanism: an exchange between Mark Haugaard and Philip Pettit”, *Journal of Political Power* 10 (1), pp.25-39.

arbitrary power of Jane, may be sufficient to mean that he dominates Jane, and that Jane can no longer truly look John in the eye. But there are other ways in which fate can conspire against us, and undermine our ability to look one another in the eye. As we have argued, following the relational approach to autonomy, persons have a deep interest in social hierarchy. All of us are loathed to being placed in a subordinate position before another person – being made to defer to their judgements, being made to grovel, being made to flatter and pander to them. Crucially, being subordinated like this undermines our ability to look others in the eye, even when such subordination is ‘non-arbitrary’ in the sense of being productive enough that it might be in our overall best interests and even consented to within a democracy. For instance, one can imagine a case where being made to grovel before government officials, turns out to have highly beneficial societal consequences, and thus becomes an ordinary requirement in a democratic political community. Even in this case, however, as reflecting upon the relational autonomy literature teaches us, the autonomous flourishing of the citizens involved would be damaged, and their ability to look their government officials in the eye plausibly destroyed. It is our nature as social creatures, who care about social hierarchies we find ourselves in, that we are sensitive in this way to being placed in a (productive) subordinate position before others. Republican accounts of domination would do well to incorporate this relational insight.

Second, it is natural when reflecting upon Republican theories of domination, to ask: why should we even care, much less be distressed, by the thought of others dominating and exerting arbitrary control over us? We suggest that Republican theories lack the resources to fully answer this question, instead taking it as a given that we should care about these things. How might one try to fill this argumentative gap? Again, we suggest that relational theories of autonomy can help Republicans out. As we have seen, relational theories of autonomy reflect carefully on the nature and needs of persons. They find that we are thoroughly social animals, through and through. We depend on others in a myriad of ways – not least in the sense that we have a desire to stand in social hierarchies of a certain kind. These reflections, among others, help us to explicate why we might justifiably be deeply concerned with how others exercise power over us.

We have outlined, then, the first part of our relationship account, and indeed, its main insight: that persuasion-control is objectionable in light of the way it places citizens in a starkly subordinate position before those in political authority. Granted, we are often in a somewhat

subordinate position before those in political authority, yet standing in a relation of intellectual deference distinguishes persuasion-control and the species of subordination it involves, from typical government interference in our lives. Moreover, we have explored two ways of understanding this insight. The first reads the relationship account as simply a new development in relational theories of autonomous flourishing, while the second reads the account as rooted in the Republican tradition, albeit incorporating at least two insights from the work of relational theorists. We believe both options are promising, and in this project, do not commit to either option in particular.

Let us now move on to the *second* way in which censorship is troubling on our relationship account. Leaving aside the especially worrying intellectual deference involved in persuasion-control, proponents of relational theories of autonomy have reason to generally worry about the asymmetric nature of the government-citizen relationship. The government-citizen relationship, by its very nature, involves an extreme asymmetry of power and various forms of deference; governments can physically coerce their citizens, put them in prison, seize their property, control what kind of agreements they can make, where they can go, how they may act, tax their earnings, and so on and so on. Does the government therefore automatically rule me in a demeaning way that undermines my relational need for respectful relationships? The answer is 'no'. In particular, what is important is that the government-citizen relationship involves what we term '*reframing factors*'.

Consider, by way of analogy, the asymmetric nature of the relationship I have with my thesis supervisor. My supervisor wields a great deal of asymmetric power over me, his supervisee; he reports back to the department on my progress, he will provide a reference which will ultimately affect my future employment chances, he can help arrange opportunities for me to exchange my work with others, he can provide invaluable feedback on my work, etc. There is then, a rather large imbalance of power in the relationship. Are we to conclude that my supervisor-supervisee relationship is in fact morally problematic? It would be nice to resist this conclusion; I've always enjoyed the supervisor-supervisee relationship and found it to be both respectful and productive. Similarly, many other supervisees, not to mention soldiers and patients, would no doubt speak up in favour of the relationships in which they are the 'subordinated' party. Is there any way of making sense of this? Or are we all mistaken, our happiness with these relationships perhaps the product of adaptive preferences?

The key point to consider is that while these relationships may be asymmetric and deferential in various ways, they also involve what we might call '*reframing factors*'. In a relationship that involves dominant and subordinate parties, a reframing factor is some fact about that relationship that helps recontextualise the relationship and re-empower the subordinate party. Reframing factors are thus facts about that relationship that change the context of the asymmetry, such that the position of the subordinate party is no longer as demeaning as it might otherwise be. My supervisor may indeed wield a great deal of power over me. But at the same time: I chose to be in this relationship, I have also since chosen to stay in this relationship, I can leave this relationship at any time, and if I feel like the relationship is becoming problematic then I can go over my supervisor's head and report him. These are all examples of reframing factors. These factors help explain how I am empowered in this relationship, despite the asymmetric power structure.

The government that cares about forming a relationship that allows citizens to 'look government officials in the eye', must carefully construct and commit to reframing factors. The problem with this is that the government-citizen relationship is (almost always) absent of all the '*reframing factors*' listed just above; very few choose to reside in a country out of a desire to be in a relationship with that country's particular government, none of us have any real individual power to elect our government, very few of us can easily leave the government-citizen relationship, nor easily go over the head of our government. Indeed, it may well be that the very nature of the scale of power that governments must necessarily wield over their citizenry, and the impersonal, technocratic nature of ruling over huge numbers of people, means that the standard gamut of reframing factors tends to be unavailable. With this in mind, an agreement to abstain from persuasion-control can operate as an important reframing factor in the government-citizen relationship. It is surely not the only reframing factor in this relationship. Perhaps basic legal rights and democratic voting laws, for example, also act as bulwarks that stem the tide of governmental authority over their citizens. All the same, an agreement to refrain from persuasion-control and treat the beliefs of citizens as private, can operate as a commitment to the dignified standing of citizens; a commitment by the government, despite all their power, to addressing citizens as thinking beings that can decide for themselves what beliefs and intentions to develop. Crucially, this commitment need not imply that citizens are perfectly rational thinking beings, or that citizens would not fare better were the government to

remove materials it believes might have a corrupting effect. Nor does it ensure that citizens and the government possess the same kind of overall power in their relationship. Instead, it simply reflects a desire to preserve a relationship where citizens enjoy at least authority over their own minds when standing before the government.

We have arrived then at what we take to be the most interesting and plausible expression of the scepticism towards persuasion-control that this thesis set out to investigate: persuasion-control is problematic because it changes the government-citizen relationship. In particular, it establishes a relationship where citizens stand before their government as intellectual subordinates – the government granted authority to tell citizens what materials and thoughts are beyond the pale. Moreover, permitting the government to engage in persuasion control removes one of the key ‘reframing factors’ that helps ensure the standing of citizens within this relationship.

One loose thread is still in need of tying up: how does all of this pertain to the issue of our democratic values? There are several ways in which we might draw a connection between, on the one hand, our ideals of democracy and political legitimacy, and on the other hand the relationship account. The most compelling, we think, is to trace the Republican thought, sketched earlier in this chapter, that the movement out of the state of nature and into a political community, creates a special kind of moral challenge. On the one hand, the establishment of a coercive political authority that places limitations on the freedom of citizens, is necessary if we are to establish a community where citizens can have substantial freedoms. However, this means that certain citizens – those who form part of the political authority – will wield special control over other citizens. How can we grant citizens this kind of authority, in a way that nonetheless ensures that all citizens can look one another in the eye? To put it another way, how can we ensure that our political community treats citizens as moral equals – none of whom are required to senselessly debase themselves or grovel or pander before others? Our democratic ideals are designed to answer these questions, and different theories of democracy and political legitimacy provide different answers. As for the relationship account outlined in this thesis, we suggest that ensuring that citizens may look one another in the eye, requires not only that government officials refrain from exerting arbitrary power over citizens, but that they are also sensitive to the social hierarchy their non-arbitrary use of power creates.

#### §IV. Objections and responses

To end this chapter, we canvass and respond to some potential objections to our relationship account, as well as comment on some of the practical implications that follow from all of this. To help keep things organised, we split the potential objections into three categories: objections to the way we have constructed the government-citizen relationship, objections that stem from an alternative approach to the issue of censorship, and objections to the effect that our account exaggerates the wrong of censorship.

##### *a. objections to how we have constructed the government-citizens relationship*

First, one might object that our relational account rests on a chimerical, illusory view about the nature of people. Our account problematises the government exercising striking asymmetric power over us, and in particular demanding intellectual deference from us. But, so the objection goes, the truth is that we need to be treated this way. That our lives will go much better if we have a meddling nanny-state telling us what to do, what to think, what to believe, etc. Why then, shouldn't the government operate in this way and ask that we defer to its judgement? Our account complains that this treatment is demeaning. But the objector will insist that this treatment is in fact apt, and that intellectual deference is an epistemically valid option.<sup>279</sup>

This is a deep worry: certainly, if the relationship account is to be convincing, it must rest on a reasonable demand for respect and recognition from the government, rather than mere vanity. With this in mind, let us offer the following response: just because we citizens are imperfect (something which most of us realise by the way!), does not mean we want other parties to take over our lives so that they can better direct them. Perhaps it is not demeaning for the government to point out that its citizens are often inadequate when it comes to forming beliefs and intentions. But there is something *prima facie* demeaning in the government forcing us to therefore defer to its judgement on these matters. Indeed, even when the government has good reason to force us to defer in this way, it does not necessarily take away from the demeaning nature of it all. Consider, by way of analogy, the case where the government forces

---

<sup>279</sup> Alvin Goldman (2001). "Experts: which ones should you trust?", *Philosophy and Phenomenological Research* 63 (1), pp.85-110.

citizens to grovel before the state, or announce their loyalty on a daily basis, because it turns out that this will predictably produce an increase in citizen compliance with the state's directives and a drop in crime-rate. All the same, those of us that care about how citizens stand before their state and their relational need for respect in their institutional relationships, have reason to find this deeply worrying. In a similar vein, insofar as we view the government-citizen relationship as a source of meaning and value in our lives, we have reason to object to persuasion-control.

*b. alternatives*

Here we consider an alternative approaches to evaluating persuasion-control, that one might think fares better than our relationship account. In particular, one can imagine a kind of Razian approach to analysing censorship that judges persuasion-control to be not in fact particularly objectionable. Censorship, so the thought goes, is just another tool that the government uses to help agents comply with their reasons. That is to say, it is another tool that the government can use to ensure that it serves its citizens in a way a government should. In this way, persuasion-control is not as problematic as our account makes out.

The point is well taken, but it is hardly a knockdown objection to our account. This thesis has never contested the idea that censorship might be used to make society a safer, less discriminatory, more pleasant place to live. The point is rather that sometimes we can lodge a principled, deontological complaint even with government policies that turn out to be useful. A ban on democratic voting, for instance, might well turn out to benefit society greatly in terms of ensuring the best, most productive government is elected. Still, even in this case, such a ban would strike many sensible commentators as outrageous, because of the way it flies in the face of some principle or deontological value e.g. fairness, the ideal that the government is a representation of the people, etc. The relationship account concedes, *arguendo*, that censorship may well be highly useful, but contends nonetheless that it flies in the face of an important value many of us are right to take seriously; the value of citizens being able to look their government officials in the eye.

*c. exaggeration*

Let us end the chapter by looking at some of the ways in which it might be claimed that our account exaggerates the problem of censorship. First, one might contend that contrary to our analysis, good relationships frequently involve censorship, and that this kind of interference does not in fact make the relationship problematically unbalanced. For instance, perhaps my supervisor hears about a conference that he thinks I might be interested in attending, but declines to mention it to me because he thinks I should rather focus on finishing my thesis. Here my supervisor forms a judgement about what kind of work I should be focusing on, and ‘filters’ what information I have access to, for fear that I might form an opposing and incorrect judgment. Many readers would surely judge that supervisor’s actions to be perfectly acceptable, not to mention quite compatible with a functioning supervisor-supervisee relationship. Does our account implausibly deny this?

Three points are worth noting in response. First, in the example above, my supervisor does not demand the same kind of moral, intellectual deference as that which is involved in persuasion-control. Sure, he filters information as a means to getting me to act how he thinks I *ought* to act. But this is not a moral ought. It is merely how he thinks I ought to act given my stated aim of finishing my PhD in good shape. This means that my supervisor is wielding a less radical form of authority over me. Were we to alter the case so that my supervisor was acting so as to demand moral deference from me concerning my own beliefs, then we suspect that most readers would suddenly find his behaviour to be rather more problematic. At the very least, it should surely be admitted that this kind of moral management would change our supervisor-supervisee relationship.

Second, as noted earlier, my supervisor-supervisee relationship involves a number of what we have termed ‘reframing factors’ that may well function to draw the string of my supervisor’s moral management. These reframing factors help empower me within the relationship. They help me hold my head up high and look my supervisor in the eye, despite the imbalance of power in our relationship. Third, and this will complicate matters, we have avoided suggesting that just because a relationship is unbalanced or degrading, this necessitates that the relationship is therefore all-things-considered impermissible or should be immediately dissolved. We have instead suggested that such relationships are pro tanto objectionable, and that the government-citizen relationship may be especially objectionable given that it is especially unbalanced and (often) lacking in reframing factors. This leaves the door open to, at

least sometimes, embracing a more demeaning relationship. Perhaps, for instance, an objectionable government-citizen relationship is worth persevering with because it offers important kinds of benefits.

For now, let us consider a separate but related objection. This objection amounts to a version of Brison's inversion argument, and goes like this: permitting censorship can lead to a worryingly subordinating relationship between the government and its citizens. But our account exaggerates matters insofar as it draws the conclusion that we should reject censorship. Granted, persuasion-control may strike a blow against one aspect of our autonomy. But declining to engage in persuasion-control simply reinforces pre-existing patterns of domination, allowing certain groups to subordinate others and creating a network of social relations that undermines their autonomy.<sup>280</sup> In this way, we ought to weigh up the extent to which censorship protects and promotes autonomy, against the way it undermines autonomy, and draw the appropriate conclusion.

Again, this point is well taken; there may be serious, negative consequences to declining to engage in censorship. There is surely no doubt that our current network systematically undermines the autonomous flourishing of many people. This would appear to leave us with a dilemma: either we resort to censorship and undermine to some degree the relational autonomy of all citizens, or else we refrain from censorship and proceed with the pre-existing 'network of others' that no doubt imposes a wide range of significant burdens on the autonomy of many citizens.

This dilemma admits of no easy resolution. After all, comparing two distinct values is notoriously difficult. How are we to compare the importance of the standing of citizens before their government, versus other aspects of the relational autonomy of citizens such as their social standing in society or their ability to form meaningful interpersonal relationships? One thing to say is that despite their superficial similarity, these are starkly different aspects of our relational autonomy. The latter involves how our peers perceive us and whether they grant us the kind of social respect and recognition that we need. Whereas the former involves the unique issues of intellectual deference and social hierarchy. Some would suggest that we consider which of these aspects of our relational autonomy would be more valuable, in

---

<sup>280</sup> Brison, 'The autonomy defence of free speech'; Cécile Laborde (2008). *Critical Republicanism: The Hijab Controversy and Political Philosophy* (Oxford: Oxford University Press).

isolation, across our whole life. Others might ask us to consider which one of these values would be most helpful in allowing to achieve some more important, ‘master’ aspect of autonomy, such as our ability to conceive, revise, and pursue our conception of the good, or our overall happiness. Some might argue that the government has special political obligations to prioritise securing certain aspects of our autonomous flourishing over others. For instance, certain theories may, from the standpoint of political legitimacy, prioritise ensuring that citizens can look one another in the eye, over ensuring that citizens have other needs met. Investigating whether any of these approaches work, is a task for another time. For now we note that the relationship account presented sharpens our focus on these suggestions and the importance of comparing these different aspects of our autonomous flourishing, even if it provides no easy answers.

With all this said, we finish by emphasising two important practical implications of the relationship account, for how we approach the issue of persuasion-control. First, the relationship account reveals a new way in which even effective persuasion-control is objectionable. This should raise the stakes for those governments that are considering engaging in persuasion-control. For instance, we might plausibly think that persuasion-control requires higher evidentiary thresholds to be met, before it is acceptable, compared to other forms of censorship that do not undermine the relational flourishing of agents. That is to say, compared to other forms of censorship, persuasion-control requires more certainty from the government that it will be effective. Similarly, the threshold for what counts as effective or productive governmental interference, should be higher in the case of persuasion-control, compared to other forms of censorship.

Second, the censoring government should be sure, when possible, to put in place significant ‘reframing factors’, that may serve to repair the government-citizen relationship and improve the standing of citizens within it. Here the idea is that if the government is to engage in persuasion-control, it can at least be mindful to do so in a way that reduces the demeaning aspects of it and the extent to which it demands intellectual (moral) deference. For instance, one ‘reframing factor’ might involve ensuring that (where possible) the government conducts public debates about whether to engage in persuasion-control, that themselves (the debates, that is) do not involve persuasion-control. In essence, the government would be at least allowing their citizens to discuss their continued intellectual subordination, as intellectually

unsubordinated agents. Relatedly, transparency may be key. The government can be explicit and vocal about its reasons for censorship, compatible of course with the censorship still being effective. It must resist the temptation of plausible deniability. In his ground-breaking essay on the different types of freedom, Isiah Berlin warned of the special wrong of interfering with someone's freedom, while trying to obscure the fact of this interference or even convince the person being interfered with that the interference is actually in accordance with their own wishes.<sup>281</sup> Such disguised interferences actively subvert our having a respectful and somewhat equal standing with the interferer, and so would be particularly destructive in the already highly asymmetric government-citizen relationship, and especially so in the case of persuasion-control.

Likewise, the government should be open about cases of censorship that proved unsuccessful; cases where they overstepped the boundaries of what was necessary, or proved overzealous in their interference. A reader might respond that these kinds of measures are a given in any reasonable account of the government's obligations. Still, the point the relationship account makes clear, is that these measures are especially important from the standpoint of ensuring that the government protects the relative standing of the citizens that it exercises power over.

## Conclusion

This thesis has sought to analyse and evaluate some of the most important philosophical critiques of what we have termed 'persuasion-control'. More precisely, the thesis has examined several arguments as to why, allegedly, persuasion-control lies in conflict with our democratic values. While the thesis canvassed a number of deeply interesting, stimulating arguments, offered by some of the greatest political philosophers of the last 50 years, it ultimately concluded that none of these arguments are successful. The reasons for rejecting these arguments vary, but a general theme that emerges is that these arguments do not fully appreciate what we might term the relational aspect of persons and the democratic community. For instance, some of these arguments fail because they fail to appreciate the interpersonal duties that political citizens owe one another, that making sacrifices for one's fellow citizens is

---

<sup>281</sup> Isiah Berlin (2002) [1969]. 'Two concepts of liberty', in *Four Essays on Liberty* (London: Oxford University Press), ed. Isiah Berlin.

neither degrading nor unfair, that co-operation between citizens is vitally important if flourishing is to be achieved on any kind of large scale, and that social antagonisms deeply threaten the autonomous flourishing of persons.

Still, there is an ironic twist to all of this. For the thesis supposes that while reflecting on the relational nature of persons helps to defuse some of the classic worries with persuasion-control, further reflection reveals persuasion-control itself to constitute a unique threat to our relational flourishing. The most important contribution of the thesis is the outline of what is termed 'the relationship account'. The relationship account draws heavily on insights from the feminist literature on relational autonomy, as well as Republican approaches to freedom. The account argues that as social beings, we are highly sensitive to the social hierarchy in which we find ourselves – including our standing before those citizens that possess political authority. Building on this insight, it is further argued that persuasion-control damages the relationship between ordinary citizens and those citizens possessing political authority, by requiring that the former stand as intellectual subordinates before the latter. This fundamentally changes this relationship, and compromises the ability of former to look the latter 'in the eye'.

In the end, then, the thesis not only raises the stakes for governments tempted to engage in persuasion-control. It also applies pressure on such governments from a somewhat new direction – that is, it applies pressure from those concerned with the relational, social needs of agents, who, as this thesis has argued, are often generally sympathetic to government interference aimed at e.g. combating social hatred.

## Bibliography

Tineke Abma (2005). "Struggling with the fragility of life: a relational-narrative approach to ethics in palliative nursing", *Nursing Ethics* 12 (4), pp.337-48.

Tineke Abma, Barth Oeseburg, Guy Widdershoven, and Marian Verkerk (2009). "The quality of caring relationships", *Psychology research and behavior management* 2, pp.39-45.

Larry Alexander (1996). "Banning hate speech and the sticks and stones defense", *Constitutional Commentary* 600 (13), pp.71-100.

Robert Amdur (1980). "Scanlon on freedom of expression", *Philosophy & Public Affairs* 9 (3), pp.287-300.

Elizabeth Anderson (2011). "Democracy, public policy, and lay assessments of scientific testimony", *Episteme* 8 (2), pp.144-64.

Sarah Andrea, Sarah, Siegel, Alan Teo (2016). "Social support and health service use in depressed adults: findings from the national health and nutrition examination survey", *General Hospital Psychiatry* 39, pp.73-9.

Liisa Antilla (2010). "Self-censorship and science: a geographical review of media coverage of climate tipping points", *Public Understanding of Science* 19 (2), pp.240-56.

Richard Arneson (2015). "Nudge and shove", *Social Theory and Practice* 41 (4), pp.668-91.

Cecilia Aslund, Peter Larm, Bengt Starrin, Kent Nilson (2014). The buffering effect of tangible social support on financial stress: influence on psychological well-being and psychosomatic symptoms in a large sample of the adult general population", *International Journal for Equity in Health* 13 (85), pp.1-9.

Annette Baier (1985). *Postures of the Mind* (Minneapolis: University of Minnesota Press).

Edwin Baker (1989). *Human Liberty and Freedom of Speech* (Oxford: Oxford University Press).

Edwin Baker (2011). "Autonomy and free speech", *Constitutional Commentary* 27, pp.251-82.

Edwin Baker (2012). "Hate speech", in *The Content and Context of Hate Speech: Rethinking Regulation and Responses* (Cambridge: Cambridge University Press), ed. Michael Herz and Peter Molnar, pp.57-80.

Dan Barry and John Eligon (2017). 'Trump, Trump, Trump!' How a president's name became a racial jeer", available at: <https://www.nytimes.com/2017/12/16/us/trump-racial-jeers.html>

Philip Baugut and Katharina Neumann (2020). "Online propaganda use during Islamist radicalization", *Information, Communication and Society* 23 (11), pp.1570-92.

Paul Benson (1994). "Autonomy and self-worth", *Journal of Philosophy* 91 (12), pp.650-68.

Isiah Berlin (2002) [1969]. 'Two concepts of liberty', in *Four Essays on Liberty* (London: Oxford University Press), ed. Isiah Berlin.

Johan Christiaan Bester (2017). "Measles vaccination is best for children: the argument for relying on herd immunity fails", *Bioethical Inquiry* 14, pp.375-84.

Monika Betzler (2019). "The relational value of empathy", *International Journal of Philosophical Studies* 27 (2), pp.136-61.

Jennifer Blumenthal-Barby (2013). "Choice architecture: a mechanism for improving decisions while preserving liberty", in *Paternalism: Theory and Practice* (Cambridge: Cambridge University Press), ed. Christian Cowons and Michael Weber, pp.178-97.

Jason Brennan (2018). "A libertarian case for mandatory vaccination", *Journal of Medical Ethics* 44 (1), pp.37-43.

Susan Brison (1998). "The autonomy defense of free speech", *Ethics* 108 (2), pp.312-39.

Susan Brison (2001). "Contentious freedom: sex work and social construction", *Hypatia* 21 (4), pp.192-200.

Alexander Brown (2015). *Hate Speech Law: A Philosophical Examination* (New York: Routledge).

Alexander Brown (2017). "Hate speech laws, legitimacy, and precaution: a reply to James Weinstein", *Constitutional Commentary* 32 (3), pp.599-617.

Alexander Brown (2017). "What is hate speech? Part 1: the myth of hate", *Law and Philosophy* 36, pp.419-468.

Alexander Brown (2017). "What is hate speech? Part 2: family resemblances", *Law and Philosophy* 36, pp.561-613.

Alexander Brown and Adriana Sinclair (2019). *The Politics of Hate Speech Laws* (New York: Routledge).

Allen Buchanan (2002). "Political legitimacy and democracy", *Ethics* 112 (4), pp.689-719.

Andrea Caputo (2013). "A literature review of cognitive biases in negotiation processes", *International Journal of Conflict Management* 24 (4), pp.374-98.

Eric Cave (2007). "What's wrong with motive manipulation?" *Ethic Theory Moral Practice* 10 (2), pp.129-44.

Amanda Cawston (2019). "The feminist case against pornography: a review and re-evaluation", *Inquiry* 62 (6), pp.624-58.

Leah Ceccarelli (2011). "Manufactured scientific controversy: science, rhetoric, and public debate", *Rhetoric and Public Affairs* 14 (2), pp.195-228.

Michael Cholbi (2017). "Paternalism and our rational powers", *Mind* 126 (501), pp.123-53.

John Christman (1991). "Autonomy and personal history", *Canadian Journal of Philosophy* 21 (1), pp.1-24.

John Christman (2004). "Relational autonomy, liberal individualism, and the social constitution of selves", *Philosophical Studies* 117 (1-2), pp.143-64.

John Christman (2013). "Autonomy", in *The Oxford Handbook of the History of Ethics* (Oxford: Oxford University Press), ed. Roger Crisp, pp.281-93.

Olga Chyzh, Mark Nieman, and Clayton Webb (2019). "The effects of dog-whistle politics on political violence", unpublished paper. Iowa State University.

Daniel Coleman (2015). "Traditional masculinity as a risk factor for suicidal ideation: cross-sectional and prospective evidence from a study of young adults", *Archives of Suicide Research* 19 (3), pp.366-84.

Daniel Coleman, William Feigelman, and Zohn Rosen (2020). "Association of high traditional masculinity and risk of suicide death: secondary analysis of the add health study", *JAMA Psychiatry* 77(4), pp.435-37.

Sandra Coliver (1992). "Hate speech laws: do they work?", in *Striking a Balance: Hate Speech, Freedom of Expression, and Non-Discrimination* (Article 19, International Centre Against Censorship, Human Rights Centre), ed. Sandra Coliver, pp.363-74.

Sarah Conly (2013). "Coercive paternalism in health care: against freedom of choice", *Public Health Ethics* 6 (3), pp.241-45.

Irene Couzigou (2019). "The criminalization of online terrorism preparatory acts under international law", *Conflict and Terrorism*, pp.1-20.

Christian Crandall, Jason Miller, Mark White (2018). "Changing norms following the 2016 U.S. presidential election: the Trump effect on prejudice", *Social Psychological and Personality Science* 9 (2), pp.186-92.

- Sue Curry Jansen and Brian Martin (2004). "Exposing and opposing censorship: backfire dynamics in freedom-of speech struggles", *Pacific Journalism Review* 10 (1), pp.29-45.
- Sue Curry Jansen and Brian Martin (2019). "The Streisand effect and censorship backfire", *International Journal of Communication* 9, pp.656-71.
- Stephen Darwall (1977). "Two kinds of respect", *Ethics* 88 (1), pp.36-49.
- Richard Delgado (1982). "Words that wound: a tort action for racial insults, epithets, and name-calling", *Harvard Civil Rights-Civil Liberties Law Review* 17, pp.133-83.
- Richard Delgado and David Yun (1994). "The neoconservative case against hate-speech regulation – Lively, D'Souza, Gates, Car, D'Souza, Gates, Carter, and the toughlove crowd", *Vanderbilt Law Review* 47 (6), pp.1807-25.
- Richard Delgado and Jean Stefancic (2001). *Critical Race Theory: An Introduction* (New York: New York University Press).
- Peter de Marneffe (2006). "Avoiding paternalism", *Philosophy & Public Affairs* 34 (1), pp.68-94.
- Gail Dines (2012). "A feminist response to Weitzer". *Violence Against Women* 18 (4), pp.512-20.
- Edward Dove, Susan Kelly, Federica Lucivero, Mavid, Machirori, Sandi Dheensa, Barbara Prainsack (2017). "Beyond individualism: is there a place for relational autonomy in clinical practice and research?", *Clinical Ethics* 12 (3), pp.150-65.
- Gerald Dworkin (1988). *The Theory and Practice of Autonomy* (Cambridge: Cambridge University Press).
- Ronald Dworkin (1981). "Is there a right to pornography?", *Oxford Journal of Legal Studies* 1 (2), pp.177-212.
- Ronald Dworkin (1996). *Freedom's Law: The Moral Reading of the American Constitution* (Cambridge, Massachusetts; Harvard University Press).
- Ronald Dworkin (2009). "Foreword", in *Extreme Speech and Democracy* (Oxford: Oxford University Press), ed. Ivan Hare and James Weinstein, pp.v-ix.
- Ronald Dworkin (2012). "Reply to Jeremy Waldron", in *The Content and Context of Hate Speech: Rethinking Regulation and Responses* (Cambridge: Cambridge University Press), ed. Michael Herz and Peter Molnar, pp.341-4.

Susan Easton (1995). "Autonomy and the free speech principle", *Journal of Applied Philosophy* 12 (1), pp. 27-39.

A.W. Eaton (2007). "A sensible antiporn feminism", *Ethics* 117 (4), pp.674-715.

Carolyn Ells, Matthew Hunt and Jane Chambers-Evans (2011). "Relational autonomy as an essential component of patient-centered care", *International Journal of Feminist Approaches to Bioethics* 4 (2), pp.79-101.

David Enoch (2016). "What's wrong with paternalism: autonomy, belief, and action", *Proceedings of the Aristotelian Society* 116 (1), pp.21-48.

Ayal Feinberg, Regina Branton, and Valerie Martinez-Ebers. "The Trump effect: how 2016 campaign rallies explain spikes in hate", unpublished paper.

Martha Fineman (2004). *The Autonomy Myth* (London: New Press).

Agneta Fischer, Eran Halperin, Daphna Canetti, and Alba Jasini (2018). "Why we hate", *Emotion Review* 10 (4), pp.309-320.

Stanley Fish (1994). *There's No Such Thing as Free Speech; And It's a Good Thing Too* (New York: Oxford University Press).

Jessica Flanigan (2014). "A defense of compulsory vaccination", *HEC Forum* 26, pp.5-25.

Miranda Fricker (1991). "Reason and emotion", *Radical Philosophy Archive*, issue 057.

Marilyn Friedman (2013). "Relational autonomy and individuality" *The University of Toronto Law Journal* 63, (2), pp.327-41.

Katharine Gelber and Luke McNamara (2016). "Evidencing the harms of hate speech", *Social Identities* 22 (3), pp.324-41.

Anca Gheaus (2009). "How much of what matters can we redistribute? Love, justice, and luck", *Hypatia* 24 (4), pp.63-83.

Frederick Gibbons, Meg Gerrard, Michael Cleveland, Thomas Wills, and Gene Brody (2004). "Perceived discrimination and substance use in African American parents and their children: a panel study", *Journal of Personal Social Psychology* 86 (4), pp.517-29.

Jonathan Glover (1999). *Humanity: a Moral History of the Twentieth Century* (London: J. Cape).

Alvin Goldman (2001). "Experts: which ones should you trust?", *Philosophy and Phenomenological Research* 63 (1), pp.85-110.

Paul Gomberg (1994). "Autonomy and free expression", *Journal of Social Philosophy* 25 (2), pp.97-104.

Trudy Govier (1993). "Self-trust, autonomy and self-esteem", *Hypatia* 8 (1), pp.99-119.

Amanda Greene (2016). "Consent and political legitimacy", in *Oxford Studies in Political Philosophy* (Oxford: Oxford University Press), ed. David Sobel, Peter Vallentyne, and Steven Wall, pp.71-97.

Stephen Grey (1998). "Reopening the public forum-from sidewalks to cyberspace", *Ohio State Law Journal* 58 (5), pp.1535-1634.

Kalle Grill (2007). "The normative core of paternalism", *Res Publica* 13 (4), pp.441-581.

Andrea Gruenfelder-Steiger, Michelle Harris, Helmut Fend (2016). "Subjective and objective peer approval evaluations and self-esteem development: a test of reciprocal, prospective, and long-term effects", *Developmental Psychology* 52 (10), pp.1563-77.

Vasudha Gupta and Charles Korte (1994). "The effects of a confidant and a peer group on the well-being of single elders", *The International Journal of Aging and Human Development* 39 (4), pp.293-302.

Michelle Harris and Ulrich Orth (2020). "The link between self-esteem and social relationships: a meta-analysis of longitudinal studies", *Journal of Personality and Social Psychology* 119 (6), pp.1459-77.

H.L.A. Hart (1959). "Prolegomenon to the principles of punishment", *Proceedings of the Aristotelian Society* 60, pp.1-26.

Mark Haugaard and Philip Pettit (2017). "A conversation on power and republicanism: an exchange between Mark Haugaard and Philip Pettit", *Journal of Political Power* 10 (1), pp.25-39.

Kaja Heidenreich, Anders Bremer, Lars Johan Materstvedt, Ulf Tidefelt, Mia Svantesson (2018). "Relational autonomy in the care of the vulnerable: health care professionals' reasoning in moral case deliberation", *Medicine, Health Care and Philosophy* 21, pp.467-77.

Eric Heinze (2013). "Hate speech and the normative foundations of regulation", *International Journal of Law in Context* 9 (4), pp.590-617.

Eric Heinze (2016). *Hate Speech and Democratic Citizenship* (Oxford: Oxford University Press).

Steven Heyman (2009). "Hate speech, public discourse, and the first amendment", in *Extreme Speech and Democracy* (Oxford: Oxford University Press), ed. Ivan Hare and James Weinstein, pp.151-81.

Judith Hill (1987). "Pornography and degradation", *Hypatia* 2 (2), pp.3-54.

Thomas Hill (1991). *Autonomy and Self-Respect* (Cambridge: Cambridge University Press).

Chris Hoofnagle (2001). "Matters of public concern and the public university professor", *Journal of College and University Law* 27 (3), pp.669-707.

Jennifer Hornsby and Rae Langton (1998). "Free speech and illocution", *Legal Theory* 4, pp.21-37.

Jeffrey Howard (2012). "Free speech and hate speech", *Annual Review of Political Science* 22 (1), pp.93-109.

Jeffrey Howard (2019). "Dangerous speech", *Philosophy & Public Affairs* 47 (2), pp.208-54.

Joseph Jaconelli (2018). "Incitement: a study in language crime", *Criminal Law and Philosophy* 12 (2), pp.245-65.

Peter Jacques, Riley Dunlap, and Mark Freeman (2008). "The organisation of denial: conservative think tanks and environmental scepticism", *Environmental Politics* 17 (3), pp.349-85.

Nadine Jung, Christian Wranke, Kai Hamburger, Markus Knauff (2014). "How emotions affect logical reasoning: evidence from experiments with mood-manipulated participants, spider phobics, and people with exam anxiety", *Frontiers in Psychology* 5, pp.570-81.

Shelly Kagan (1991). "Replies to my critics", *Philosophy and Phenomenological Research* 51 (4), pp.919-28.

Frances Kamm (1996). *Morality, Mortality: Volume II: Rights, Duties, and Status* (Oxford: Oxford University Press).

Immanuel Kant (1996) [1785]. *The Groundwork for the Metaphysics of Morals*, in *Practical Philosophy* (Cambridge: Cambridge University Press), ed. Mary Gregor.

Jordan Kiper, Richard Ashby Wilson, Christine Lillie, and Lasana Harris (2019). "Propaganda, empathy, and support for intergroup violence: the moral psychology of international speech crimes": <https://ssrn.com/abstract=2580521>

Eva Kittay (1999). *Love's Labour: Essays on Women, Equality and Dependence* (New York and London: Routledge).

Howard Koh (2015). "Communicating the health effects of climate change", *JAMA* 315 (3), pp.239–240.

Naa Oyo Kwate and Ilan Meyer (2011). "On sticks and stones and broken bones: stereotypes and African American health", *Du Bois Review: Social Science Research on Race* 8 (1), pp.191-8.

Will Kymlicka and Wayne Norman (1994). "The return of the citizen: a survey of recent work on citizenship theory", *Ethics* 104 (2), pp.352-81.

Cécile Laborde (2008). *Critical Republicanism: The Hijab Controversy and Political Philosophy* (Oxford: Oxford University Press).

Cécile Laborde (2018). *'Liberalism's Religion'* (Cambridge, Massachusetts; Harvard University Press).

Rae Langton (1993). "Speech acts and unspeakable acts", *Philosophy & Public Affairs* 22 (4), pp.293-330.

Rae Langton and Caroline West (1999). "Scorekeeping in a pornographic language game", *Australasian Journal of Philosophy* 77 (3), pp.303-19.

Charles Lawrence (1990). "If he hollers let him go: regulating racist speech on campus", *Duke Law Journal* 39 (3), pp.431-83.

Elissa Lee and Laura Leets is fascinating (2002). "Persuasive storytelling by hate groups online examining its effects on adolescents", *American Behavioral Scientist* 45, pp.927-57.

Brian Leiter (2016). "The case against free speech", *Sydney Law Review* 38, pp.407-39.

Maxime Lepoutre (2017). "Hate speech in public discourse: a pessimistic defense of counterspeech", *Social Theory and Practice* 43(4), pp. 851-83; Geoffrey Stone (1994). "Hate Speech and the U.S. constitution" *East European Constitutional Review* 78, pp.78-82.

Maxime Lepoutre (2019). "Can "more speech" counter ignorant speech?", *Journal of Ethics and Social Philosophy* 16 (3), pp.155-91.

Primo Levi (1987). *If This Is A Man: The Truce* (London: Abacus).

Tené Lewis, Courtney, Cogburn, and David Williams (2015). "Self-reported experiences of discrimination and health: scientific advances, ongoing controversies, and emerging issue", *Annual Review of Clinical Psychology* 11, pp.407-40.

Ariel Victoria Lieberman (2017). "Terrorism, the internet, and propaganda: a deadly combination", *Journal of National Security Law and Policy* 9 (1), pp.95-124.

Alison Liebling (2011). "Moral performance, inhuman and degrading treatment and prison pain", *Punishment and Society* 13 (5), pp.530-50.

Emma Lozon (2018). "Communicating the health impacts of climate change: from conviction to persuasion", *Open Access Master's Thesis, Michigan Technological University*: <https://doi.org/10.37099/mtu.dc.etr/603>

Kathleen Lynch (2009). "Affective equality: who cares?", *Development* 52 (3), pp.410-15.

Cormac McCarthy (1992). *Blood Meridian, or, The Evening Redness in the West* (New York: Vintage International).

Mary Kate McGowan (2019). *Just Words: On Speech and Hidden Harm* (Oxford: Oxford University Press).

Sarah McKenzie, Sunny Collings, Gabrielle Jenkin, and Jo River (2018). "Masculinity, social connectedness, and mental health: men's diverse patterns of practice", *American Journal of Men's Health*, pp.1247-61.

Catriona Mackenzie and Natalie Stoljar (2000). "Introduction: autonomy refigured", in *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self* (Oxford: Oxford University Press), ed. Catriona Mackenzie and Natalie Stoljar.

Catriona Mackenzie (2008). "Relational autonomy, normative authority and perfectionism", *Journal of Social Philosophy* 39 (4), pp.512-33.

Catriona Mackenzie (2013). "The importance of relational autonomy and capabilities for an ethics of vulnerability", in *Vulnerability: New Essays in Ethics and Feminist Philosophy* (Oxford: Oxford University Press), ed. Catriona Mackenzie, Wendy Rogers and Susan Dodds, pp.33-59.

Catharine MacKinnon (1992). "Pornography, civil rights and speech", in *Pornography: Women, Violence and Civil Liberties* (Oxford: Oxford University Press), ed. Catherine Itzin, pp.456-512.

Catharine MacKinnon (1995). *Only Words* (London: Harper Collins).

Catriona McKinnon (2016). "Should we tolerate climate change denial?", *Ethics and Global Climate Change* 40 (1), pp.205-16.

Carolyn McLeod and Emma Ryman (2020). "Trust, autonomy and the fiduciary relationship", in *Fiduciaries and Trust: Ethics, Politics, Economics and Law* (Cambridge: Cambridge University Press), ed. Paul Miller and Matthew Harding, pp.74-86.

Ishani Maitra and Mary Kate McGowan (2012). "Introduction and Overview" in *Speech and Harm: Controversies over Free Speech* (Oxford: Oxford University Press), ed. Ishani Maitra and Mary Kate McGowan, pp.1-23.

Ishani Maitra (2012). "Subordinating Speech", in *Speech and Harm: Controversies Over Free Speech* (Oxford: Oxford University), ed. Mary Kate McGowan and Ishani Maitra, pp.94-120.

Avishai Margalit (1996). *The Decent Society*. (Cambridge, Massachusetts: Harvard University Press).

Mari Matsuda (1989). "Public response to racist speech: considering the victim's story", *Michigan Law Review* 87 (8), pp.2320-38.

Mari Matsuda (1993). *Words That Wound: Critical Race Theory, Assaultive Speech, and the First Amendment* (Boulder, Colorado: Westview Press).

Jonathan Maynard and Susan Benesch (2016). "Dangerous speech and dangerous ideology: an intergrated model for monitoring and prevention", *Genocide Studies and Prevention* 9 (3), pp.70-95.

José Medina (2013). "Color blindness, meta-ignorance, and the racial imagination", *Critical Philosophy of Race* 1 (1), pp.38-67.

Alexander Meiklejohn (1948). *Free Speech and Its Relation to Self-Government* (New York: Harper and Brothers).

Alexander Meiklejohn (1960). *Political Freedom* (New York: Harper and Brothers).

Hugo Mercier (2017). “How gullible are we? A review of the evidence from psychology and social science”, *Review of General Psychology* 21 (2), pp.103-22.

John Stuart Mill (2003) [1859]. *On Liberty*, in *Utilitarianism and On liberty: including Mill's Essay on Bentham' and selections from the writings of Jeremy Bentham and John Austin* (Malden, Massachusetts: Blackwell Publishers), ed. Mary Warnock, pp.88-180.

Karsten Müller and Carlo Schwarz (2020). “Fanning the flames of hate: social media and hate crime”: <https://ssrn.com/abstract=3082972>

Atsuo Murata, Tomoko Nakamura, and Waldemar Karwowski (2015) “Influence of cognitive biases in distorting decision making and leading to critical unfavorable incidents”, *Safety* 1 (1), pp.44-58.

Thomas Nagel (1995). “Personal rights and public space”, *Philosophy & Public Affairs* 24 (2), pp.83-107.

Mark Navin (2015). *Values and Vaccine Refusal: Hard Questions in Ethics, Epistemology, and Health Care* (New York: Routledge).

Jennifer Nedelsky (1989). “Reconceiving autonomy: sources, thoughts and possibilities”, *Yale Journal of Law and Feminism* 1 (1), pp.7-36.

Jennifer Nedelsky (2011). *Law's Relations: A Relational Theory of Self, Autonomy, and Law* (New York: Oxford University Press).

Laura Beth Nielsen (2002). “Subtle, pervasive, harmful: racist and sexist remarks in public as hate speech”, *Journal of Social Issues* 58 (2), pp.265-90.

Martha Nussbaum (2000). *Women and Human Development: The Capabilities Approach* (Cambridge: Cambridge University Press).

Martha Nussbaum (2000). “The future of feminist liberalism”, *Proceedings and Addresses of the American Philosophical Association*, 74 (2), pp. 47-79.

Martha Nussbaum (2003). "Capabilities as fundamental entitlements: Sen and social justice", *Feminist Economics* 9 (2-3), pp.33-59.

David O'Brien (2010). *Congress Shall Make No Law: The First Amendment, Unprotected Expression, and the U.S. Supreme Court* (New York: Rowman & Littlefield Publishers).

Marina Oshana (1998). "Personal autonomy and society", *Journal of Social Philosophy* 29 (1), pp.81-102.

Marina Oshana (2006). *Autonomy in Society* (Aldershot: Ashgate Publishing).

Marina Oshana (2014). "A commitment to autonomy is a commitment to feminism", in *Autonomy, Oppression, and Gender*, ed. Mark Piper and Andrea Veltman, pp.141-60.

Michael Otsuka (1994). "Killing the innocent in self-defense", *Philosophy & Public Affairs* 23 (1), pp.74-94.

Bhikhu Parekh (2006). "Hate speech", *Public Policy Research* 12, pp.213-23.

Richard Perloff (2020). *The Dynamics of Persuasion: Communication and Attitudes in the Twenty-First Century*, (New York: Routledge).

Michael Peters (2020). "Limiting the capacity for hate: hate speech, hate groups and the philosophy of hate", *Educational Philosophy and Theory* (forthcoming).

Phillip Pettit (1997). *Republicanism: A Theory of Freedom and Government* (Oxford: Clarendon Press).

Philip Pettit (2012). *On the People's Terms : A Republican Theory and Model of Democracy* (Cambridge: Cambridge University Press).

James Piazza (2020). "Politician hate speech and domestic terrorism", *International Interactions* 46 (3), pp.431-53.

Isabella Poggi (2005). "The goals of persuasion", *Pragmatics and Cognition* 13 (2), pp.297-336.

Robert Post (2017). "Legitimacy and hate speech", *Faculty Scholarship Series* 524, pp.651-9.

Mark Potok (2017). "The Trump effect", *Southern Poverty Law Center*, available at: <https://www.splcenter.org/fighting-hate/intelligence-report/2017/trump-effect>

Naomi Priest, Yin Paradies, Brigid Trenerry, Mandy Truong, Saffron Karlsen, and Yvonne Kelly (2013). "A systematic review of studies examining the relationship between reported racism and health and wellbeing for children and young people", *Social Science and Medicine* 95, pp.115-27.

Warren Quinn (1989). "Actions, intentions, and consequences: the doctrine of doing and allowing", *The Philosophical Review* 98 (3), pp.287-312.

Warren Quinn (1994). *Morality and Action* (Cambridge: Cambridge University Press).

Jonathan Quong (2010). *Liberalism Without Perfectionism* (Oxford: Oxford University Press).

Matthew Rabin (2003). "Projection bias in predicting utility", *The Quarterly Journal of Economics* 118 (4), pp.1209-48.

Peter Railton (1984). "Alienation, consequentialism, and the demands of morality", *Philosophy & Public Affairs* 13 (2), pp.134-71.

John Rawls (1972). *A Theory of Justice* (Oxford: Oxford University Press).

John Rawls (1980). "Kantian constructivism in moral theory", *The Journal of Philosophy* 77 (9) pp.515-72.

John Rawls (1964). "Legal obligation and the duty of fair play", in *Law and Philosophy: A Symposium* (New York: New York University Press), ed. Sidney Hook, pp.3-18.

Joseph Raz (1986). *The Morality of Freedom* (Oxford: Clarendon Press).

Martin Redish (1982). "The value of free speech", *University of Pennsylvania Law Review* 130, pp.591-645.

Andrew Reid (2020). "Does regulating hate speech undermine democratic legitimacy? A cautious 'no'" *Res Publica* 26, pp.181-99.

Judith Resnik and Julie Chi-hye Suk (2003). "Adding insult to injury: questioning the role of dignity in conceptions of sovereignty", *Stanford Law Review* 55 (5), pp.1921-62.

David Richards (1999). "Constitutional legitimacy, the principle of free speech, and the politics of identity", *Kent Law Review* 74 (2), pp.779-822.

Laura Richards: <https://www.laurarichards.co.uk/coercive-control/> and [www.laurarichards.co.uk/domestic-abuse-and-coercive-control/](https://www.laurarichards.co.uk/domestic-abuse-and-coercive-control/).

Laura Richards: <https://www.cps.gov.uk/legal-guidance/controlling-or-coercive-behaviour-intimate-or-family-relationship>

Elihu Richter, Dror Kris Markus, and Casey Tait (2018). “Incitement, genocide, genocidal terror, and the upstream role of indoctrination: can epidemiologic models predict and prevent?”, *Public Health Review* 39 (30), pp.1-22.

Arthur Ripstein (2009). *Force and Freedom: Kant’s Legal and Political Philosophy* (Cambridge, Massachusetts: Harvard University Press).

Sara Rubinelli. (2013). “Argumentation as rational persuasion in doctor-patient communication”, *Philosophy & Rhetoric* 46 (4), pp.550-69.

T.M. Scanlon (1972). “A Theory of Freedom of Expression”, *Philosophy & Public Affairs* 1 (2), pp. 204-26.

T.M. Scanlon (1979). “Freedom of expression and categories of expression”, *Pittsburgh Law Review* 40, pp.519-50.

Frederick Schauer (1985). “Slippery slopes”, *Harvard Law Review* 99 (2), pp.361-83.

Frederick Schauer (1993). “The phenomenology of speech and harm”, *Ethics* 103 (4), pp.635-53.

Andreas Schmidt and Bart Engelen (2000). “The ethics of nudging: an overview”, *Philosophy Compass* 15 (4), pp.1-13.

Jonathan Seglow (2016). “Hate speech, dignity and self-respect”, *Ethical Theory and Moral Practice* 19 (5), pp1103-16.

Amartya Sen (1985). *Commodities and Capabilities* (Amsterdam and New York: North-Holland).

Amartya Sen (1989). “Development as capability expansion”, *Journal of Development Planning* 19, pp.41-58.

Amartya Sen (2009). *The Idea of Justice* (London: Allen Lane).

Shlomo Sher (2011). "A framework for assessing immorally manipulative marketing tactics", *Journal of Business Ethics* 102 (1), pp.97-118.

Seana Shiffrin (2000). "Paternalism, unconscionability doctrine, and accommodation" *Philosophy & Public Affairs* 29 (3), pp.205-50.

Seana Shiffrin (2014). *Speech Matters: On Lying, Morality, and the Law* (Princeton: Princeton University Press).

Katelyn Sileo and Trace Kershaw (2020). "Dimensions of masculine norms, depression, and mental health service utilization: results from a prospective cohort study among Emerging adult men in the United States", *American Joournal of Men's Health* 14 (1), pp.1-16.

Evan Simpson (2006). "Responsibilities for hateful speech", *Legal Theory* 12 (2), pp.157-77.

Robert Mark Simpson (2013). "Dignity, harm, and hate speech", *Law and Philosophy* 32 (6), pp.701-28.

Robert Mark Simpson (2013). "Intellectual agency and responsibility for belief in free-speech theory", *Legal Theory* 19 (3), pp.307-30.

Griffin Sims Edwards and Stephen Rushin (2018). "The effect of President Trump's election on hate crimes", available at SSRN: <https://ssrn.com/abstract=3102652>

Peter Singer (1972). "Famine, affluence, and morality", *Philosophy & Public Affairs* 1 (3), pp.229-43.

Tara Smith (2017). "Vaccine rejection and hesitancy: a review and call to action", *Open Forum Infectious Diseases* 4 (3), pp.1-7.

Wiktor Soral, Michał Bilewicz, and Mikołaj Winiewski (2018). "Exposure to hate speech increases prejudice through desensitization", *Aggressive Behaviour* 44 (2), pp.136-46.

Sarah Sorial (2011). *Sedition and the Advocacy of Violence: Free Speech and Counter-Terrorism* (New York: Routledge).

Sarah Sorial (2015). "Hate speech and distorted communication: rethinking the limits of incitement", *Law and Philosophy* 34 (3), pp.299-324.

Daniel Statman (2000). "Humiliation, dignity and self-respect", *Philosophical Psychology* 13 (4), pp.523-40.

Lynne Steinberg (2010). "When friends make you blue: the role of friendship contingent self-esteem in predicting self-esteem and depressive symptoms", *Personality and Social Psychology Bulletin* 36 (3), pp.384-97.

Natalie Stoljar (2011). "Informed consent and relational conceptions of autonomy", *Journal of Medicine and Philosophy* 36 (4), pp.375-84.

Natalie Stoljar (2018). "Feminist perspectives on autonomy", *The Stanford Encyclopedia of Philosophy*, ed. Edward Zalta, available at:  
<https://plato.stanford.edu/archives/win2018/entries/feminism-autonomy/>

Scott Straus (2007). "What is the relationship between hate radio and violence? Rethinking Rwanda's "radio machete"", *Politics and Society* 35 (4), pp.609-63.

David Strauss (1991). "Persuasion, autonomy, and freedom of expression", *Columbia Law Review* 91 (2), pp.334-71.

Nadine Strossen (2012). "Interview with Nadine Strossen", in *The Content and Context of Hate Speech: Rethinking Regulation and Responses* (Cambridge: Cambridge University Press), ed. Michael Herz and Peter Molnar, pp.378-98.

Nadine Strossen (2018). *Hate: Why We Should Resist It With Free Speech, Not Censorship* (Oxford: Oxford University Press).

Thomas Szanto (2018). "In hate we trust: the collectivization and habitualization of hatred", *Phenomenology and the Cognitive Sciences* 19, pp.453-80.

Christine Tappolet (2006). "Autonomy and the emotions", *European Journal of Analytic Philosophy* 2 (2), pp.45-59.

Christine Tappolet (2006). *Emotions, Value, and Agency* (Oxford: Oxford University Press).

Richard Thaler and Cass Sunstein (2008). *Nudge: Improving Decisions About Health, Wealth, And Happiness* (London: Yale University Press).

Jennifer Thomas and Kimberly Daubman (2001). "The relationship between friendship quality and self-esteem in adolescent girls and boys", *Sex Roles* 45 (1), pp.53-65.

Obdulia Torres González (2018). “Claves para comprender la resistencia de los colectivos antivacunas una controversia científico-tecnológica pública”, *Revista de Humanidades de Valparaíso* 11, pp.7-37.

George Tsai (2014). “Rational persuasion as paternalism”, *Philosophy and Public Affairs* 42 (1), pp.78-112.

Alexander Tsesis (2002) *Destructive Messages: How Hate Speech Paves the Way for Harmful Social Movements* (New York: New York University Press).

Alexander Tsesis (2013). “Inflammatory hate speech: offense versus incitement”, *Minnesota Law Review* 97 (4), pp.354-405.

William Tucker (2012). “Deceitful tongues: is climate change denial a crime?”, *Ecology Law Quarterly* 39 (3), pp.831-94.

UN Office on Drugs and Crime (2012). “The use of the internet for terrorist purposes”, pp.3-11.

[https://www.unodc.org/documents/frontpage/Use\\_of\\_Internet\\_for\\_Terrorist\\_Purposes.pdf](https://www.unodc.org/documents/frontpage/Use_of_Internet_for_Terrorist_Purposes.pdf)

David Vogel, Sarah Heimerdinger-Edwards, Joseph Hammer, and Asale Hubbard (2011). “Boys don't cry”: examination of the links between endorsement of masculine norms, self-stigma, and help-seeking attitudes for men from diverse backgrounds”, *Journal of Counselling Psychology* 58 (3), pp.368-82.

Jeremy Waldron (2012). *The Harm in Hate Speech* (Cambridge, Massachusetts: Harvard University Press).

Jennifer Walter and Lane Friedman Ross (2014). “Relational autonomy: moving beyond the limits of isolated individualism”, *Pediatrics* 133 (1), pp.16-23.

James Weinstein (2011). “Participatory democracy as the central value of American free speech doctrine”, *Virginia Law Review* 97 (3), pp.491-514.

James Weinstein (2011). “Participatory democracy as the basis of American free speech doctrine: a reply”, *Virginia Law Review* 97 (3), pp.633-79.

James Weinstein (2011) “Free speech and political legitimacy: a response to Ed Baker”, *Constitutional Commentary* 27 (2), pp.361-83.

James Weinstein (2017). "Hate speech bans, democracy, and political legitimacy", *Constitutional Commentary* 32 (3), pp.527-83.

James Weinstein (2017). "Viewpoint discrimination, hate speech, and political legitimacy: a reply", *Constitutional Commentary* 32, pp. 715-78.

Andrea Westlund (2009). "Rethinking relational autonomy", *Hypatia* 24 (4), pp.26-49.

Danielle Wenner (2020). "Nondomination and the limits of relational autonomy", *International Journal of Feminist Approaches to Bioethics* 13 (2), pp.28-48.

David Williams (2018). 'Stress and the mental health of populations of color: advancing our understanding of race-related Stressors', *Journal of Health and Social Behaviour* 59 (4), pp.466-85.

Matthew Williams, Pete Burnap, Amir Javed, Han Liu, Sefa Ozalp (2020). "Hate in the machine: anti-black and anti-Muslim social media posts as predictors of offline racially and religiously aggravated crime", *The British Journal of Criminology* 60 (1), pp.93-117.

James Wilson (2011). "Why it's time to stop worrying about paternalism in health policy", *Public Health Ethics* 4 (3), pp.269-79.

James Wilson (2016). "Internal and external validity in thought experiments", *Proceedings of the Aristotelian Society* 116 (2), pp.127-52.

Robert Wolff (1970). *In Defense of Anarchism* (New York: Harper & Row).