

University of Warwick institutional repository: <http://go.warwick.ac.uk/wrap>

**A Thesis Submitted for the Degree of PhD at the University of Warwick**

<http://go.warwick.ac.uk/wrap/2422>

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it. Our policy information is available from the repository home page.

# **Autonomy, Freedom of Speech and Mental Contamination**

by

**Andrés Moles Velázquez**

A thesis submitted in partial fulfilment of the requirements for the degree  
of Doctor of Philosophy in Politics

University of Warwick  
Department of Politics and International Studies  
September 2007

## TABLE OF CONTENTS

ACKNOWLEDGMENTS	v
DECLARATION	vi
ABSTRACT	vii
INTRODUCTION	1
CHAPTER 1	
‘THE PRINCIPLE OF FREE SPEECH AND THE PUBLIC/PRIVATE DISTINCTION’	6
I The Harm Principle	7
II Speech and Harm: Minimalism and Maximalism	11
<i>Minimalism</i>	11
<i>Maximalism</i>	13
III Free Speech and the Public/Private Distinction	18
IV Some Worries about the Public/Private Distinction	24
<i>The ‘Personal is Political’</i>	24
<i>Political Persons and Public Rules</i>	28
<i>Automatic Behaviour and Situated Rationality</i>	29
V Racial Thinking, Cognition and the Costs of Speech	35
<i>Categorising Race</i>	36
<i>Racial Thinking and Hate Speech</i>	40
CHAPTER 2	
‘AUTONOMY MENTAL CONTAMINATION AND FREE SPEECH’	43
I Conditions for Autonomy	46
1) <i>Identification</i>	48
2) <i>Historical Amendments</i>	51
3) <i>Critical Reflection</i>	52
4) <i>Content Neutrality</i>	55
5) <i>Adequacy of Choice</i>	56
6) <i>Classical Rationality</i>	57
<i>Conditions of Autonomy: A Summary</i>	59
II Challenges to Autonomy	61
A) <i>Adaptive Preference Formation</i>	61
B) <i>Abnormal Belief Formation</i>	63
C) <i>Oppressive Socialisation</i>	64
D) <i>Automatic and Unconscious Processes</i>	67
E) <i>Mental Contamination</i>	74
F) <i>Revised Rationality</i>	81

III Relations between Conditions and Challenges	86
<i>Identification</i>	87
<i>Right Causal History</i>	87
<i>Critical Reflection</i>	91
<i>Content Neutrality</i>	97
<i>Adequate Options</i>	99
<i>Optimising Rationality</i>	101
Conclusions	105
<i>Toward Ecological Autonomy</i>	105
<i>Ecological Autonomy and Free Speech</i>	106
CHAPTER 3	
‘THE MILLIAN PRINCIPLE AND AUDIENCE’S INTERESTS’	110
I The Millian Principle	114
<i>The Discontinuous Responsibility Argument</i>	115
<i>Citizen’s Autonomy</i>	119
II Criticisms of the Millian Principle	123
<i>Paternalism</i>	123
<i>Reasons and Means</i>	128
<i>Autonomy and Responsibility</i>	130
<i>Responsibility for What?</i>	136
<i>Conceptions of Autonomy</i>	139
III The Millian Principle, Contractualism and Audience’s Interests	141
<i>Rawls’s Conception of the Person</i>	141
<i>Scanlon’s Contractualism and the Millian Principle</i>	146
<i>Contractualism and the Ecology of Speech</i>	147
Conclusions	151
CHAPTER 4	
‘RONALD DWORKIN AND THE INTERESTS OF SPEAKERS’	153
I Free Speech as a Trump over Social Utility	155
II Authenticity and Equality of Resources	158
III Dignity and Integrity	166
<i>Critical and Volitional Interests</i>	166
<i>Dignity and Political Equality</i>	168
<i>The Model of Challenge, Ethics and Integrity</i>	174
IV Hate Speech, Dignity and Integrity	178
<i>An Internal Conflict</i>	178
<i>Avoiding Cultural Paternalism</i>	183
<i>Justice and Autonomy</i>	185
CHAPTER 5	
‘CONTENT REGULATIONS AND LIBERAL NEUTRALITY’	188
I The Problem of Content-Based Regulation	189
<i>Content-Based and Content-Neutral Regulations</i>	190
<i>Justifications against Content-Based Regulations</i>	193

<i>The Liberal Solution to the Costs of Speech</i>	195
<i>The Dynamics of Belief</i>	197
<i>'More, better speech' and Mental Contamination</i>	202
<i>Autonomy and Content-Based Regulations</i>	205
II Legitimacy, Autonomy and Anti-Perfectionism	208
<i>What Neutrality Is and Is Not</i>	209
<i>Rawls, Liberalism and Legitimacy</i>	212
<i>Perfectionism and Anti-Perfectionism</i>	215
<i>Legitimacy, Publicity and Anti-Perfectionism</i>	220
<i>Autonomy and Anti-Perfectionism</i>	223
CHAPTER 6	
'FREE ASSOCIATION, PUBLIC ECOLOGY AND FREE SPEECH'	228
I Associational Freedom	231
<i>Associational Goods</i>	231
<i>Conditions for Associations</i>	235
<i>Examples of Associations</i>	239
II A Theory of Free Association	243
<i>The 'Moral Uses of Pluralism'</i>	243
<i>The 'Moral Uses of Pluralism': A Critique</i>	250
III Public Reason and Reasonable Association	261
<i>The Condition of Reasonableness</i>	262
<i>Public Reason</i>	265
<i>Reasonable Association</i>	269
CONCLUSIONS	278
BIBLIOGRAPHY	287

## ACKNOWLEDGMENTS

Most of what I know about Political Theory I have learnt it from my supervisors, Matthew Clayton and Susan Hurley. Their encouragement, patience, and inspiration have been invaluable. They provided me with guidance, insights and criticisms without which I would not have finished this thesis. I was devastated by Susan Hurley's death, and this work is a tribute to her memory.

Special thanks are due to the members of the Research Seminar in Political Theory, in which earlier versions of this work were presented and discussed: John Cunliffe, Terry Lovell, Katrina Mitcheson, Bill O'Brian, Will Smith, Victor Tadros, and Andrew Williams. In particular to Ed Page, Fabienne Peter, and Andrew Reeve.

I have also benefited from audiences at seminars and conferences where part of this material has been discussed, especially to CRIPS and fellow postgraduate colleagues at Warwick. The CONACyT kindly funded my studies at Warwick making this thesis possible.

I am also very grateful to Andrew Walton, who generously proofread my work offering suggestions and ways of making the arguments better.

To Susana Cerda, for her love, companionship and support along all these years, I owe much more than words can describe. I dedicate this thesis to my parents, Cristina Velázquez and Luis Moles, for their unconditional support over the years.

## DECLARATION

I declare that this thesis is my own work and an original piece of research and that I have not submitted it for a degree at another University.

Some of the material of Chapter 2 appears is Moles, Andrés 'Autonomy, Free Speech and Automatic Behaviour', *Res Publica*, 13 (2007), 53-75.

## ABSTRACT

The aim of the thesis is to rebut the dominant autonomy-based defence of hate speech within a liberal framework. The thesis argues that liberal egalitarianism is compatible with certain restrictions on free speech. I defend the view that liberal ideals such as equality and autonomy are, contrary to the arguments of many liberals, better achieved by imposing certain restrictions on what citizens are allowed to express.

I examine the problem of freedom of expression in the context of the public/private distinction. In particular, I explore the Rawlsian conception of this distinction, which is based on the idea that principles of justice apply only to the 'basic structure of society'. Citizens are required by justice to treat all others as free and equal citizens, but this seems to hold only when citizens deliberate about 'constitutional essentials and matters of basic justice'. In their private lives and other social contexts citizens are free to treat other people without equal respect and concern, provided that basic rights are not violated. This position is criticised by calling attention to recent developments in Social and Cognitive Psychology. Evidence suggests that much of our behaviour is triggered by features of the environment that bypass individuals' rational control: this includes social stereotypes, non-instrumental behaviour, and goal-oriented activity among others.

I develop these ideas into a discussion of free speech and autonomy. I argue that autonomy defences of free speech need to assess how the environment directly affects rational processes. Moreover, I argue, given the structure of human cognition, there is no guarantee that attitudes and actions cultivated in the private sphere will not 'spillover' into the public sphere. For this reason, I suggest, political morality must also extend to the justice of our private practices. To the extent that autonomy and justice matter, I argue that we have reasons to limit the expression of certain views, in particular those which trigger processes that bypass rational control.

Finally, despite the importance I attribute to the concept of autonomy, I reject the claim that my position endorses a form of liberal perfectionism. I do so by defending a conception of full publicity and demonstrating that the view I articulate is compatible with rejecting perfectionism.

Keywords Automatic Behaviour, Autonomy, Ecological Rationality, Free Speech, Justice, Liberal Legitimacy, Mental Contamination, Racist Speech



## INTRODUCTION

There is a deep divide within liberalism on the subject of freedom of speech. There are important controversies about how the best interpretation of liberal values, such as equality and justice, relate to issues of free expression. In particular, instances of racist and sexist speech are troublesome. Some liberals that endorse equality, such as Thomas Scanlon, Thomas Nagel and Ronald Dworkin, are committed to the protection of speech even when it challenges the most basic assumption of liberal egalitarianism: the idea of moral equality. In this thesis I refute this claim and defend a view which conflicts with it.

Free speech has a peculiar structure. According to a popular liberal view, a minimal principle of liberty holds that it is wrong to interfere with the actions of an individual so long as they do not harm others. However, a principle of free speech extends beyond this, protecting expression even when it is harmful. There are various reasons offered to support this. Popular examples include the claims that free speech is necessary for advancing the truth, that it is a necessary component of a democratic regime, that its restriction risks a slippery slope which threatens the minimal principle of liberty, or that it is necessary for the promotion and protection of personal autonomy. It is the last of these that is the focus of this thesis.

Autonomy figures prominently in the justification of political morality which forms the background of this work. I argue that individuals have a fundamental

interest in developing an effective sense of justice and in living in an autonomous manner. The basic framework within which the argument of this thesis is developed is detailed in Chapter 1. I explain the peculiar structure of freedom of speech and explore the reasons upon which its autonomy-based defence is *prima facie* attractive. I review the defence of free speech in conjunction of considerations of the popular public/private distinction, and suggest that criticisms of this distinction might also be applied to controversies of speech. I identify two problems. First, there is an issue regarding the ‘site’ of justice. Should political principles only apply to coercive institutions, or should they also inform the discrete actions of private citizens? Second, there is an increasing amount of academic evidence which suggest that many forms of behaviour are caused directly by features of the environment. That is, some of our actions are not the result of acts of will, but, rather, caused by external stimuli. I suggest that the behaviour’s automaticity challenges the public/private distinction by highlighting that (at least) some of an individual’s actions are not completely under her control, and, therefore that she might not be able to respond differently in different realms as justice requires.

The second chapter explores the notion of autonomy which underlies defences of free speech. In particular I seek to answer the question: how can we reconcile the idea of personal autonomy with the fact that much of our behaviour occurs automatically? I argue that, in order to accommodate the challenges derived from unconscious processing, we need to make certain revisions to some of the conditions for autonomy.

Discussions of free speech sometimes distinguish between the various interests actors have in expression. In Chapter 3 I ask whether it is always beneficial

for an audience for speech to be protected. To address this question I discuss a famous argument provided by Thomas Scanlon. I conclude that, although it seems free speech is, in general, beneficial to audiences, it can also have bad consequences. Should it not be in an audience's interests to neutralise forms of speech that affect us in non-autonomous ways? I suggest that it is, and defend the view through a contractualist argument.

Nevertheless, it is arbitrary to favour only the interests of an audience; theirs are not the only ones at stake. In Chapter 4 I turn my attention to the other significant interests in free speech: those of speakers. Perhaps, even if audiences are better protected by speech restrictions, speakers' expressive interests trump the former, rendering restrictions illegitimate. There seem to be two approaches by which this conflict can be addressed. On the one hand, it might be suggested that we must balance the two interests and adjudicate between them according to which side's interests are more significant. On the other hand, one might argue that the interests of speakers are conditional upon their being just. I favour the second response, and employ Dworkin's account of ethical integrity to defend this view.

The arguments developed to this point could reasonably suggest to some a form of liberal perfectionism. In Chapter 5 I attempt to dispel such an impression. I defend a qualified version of political liberalism based on a Rawlsian condition of full publicity. This version is anti-perfectionist, but remains concerned with protecting autonomy.

In the last chapter I explore a different liberal institution, whose defence resembles, in several important aspects, that of speech: free association. I itemise some 'associational goods' and explore whether associations which have racist

admissions policies or a 'racist agenda' could be defended by appealing to such goods. Nancy Rosenblum argues they could, asserting that an association is to be assessed exclusively on the impact it has on its members. I criticise this view posing the question: Can attitudes cultivated in private be detrimental to the development of citizens' two moral powers? If so, then perhaps we should re-think the relationship between moral development and free association. I suggest that associational goods can be produced in a manner consistent with justice if the practices of associations are constrained by a condition of reasonableness, and its admissions policies are justifiable by public reasons.

Rawls writes that some unreasonable doctrines must be contained 'like war or disease'<sup>1</sup>. The task of this thesis is to defend an argument according to which restrictions on racist speech are compatible with autonomy and justice. In this respect, restrictions of speech are legitimate means to contain racist doctrines. The position I defend is not complete. Clearly, autonomy and justice are not the only relevant moral concerns in adjudicating on speech regulation, and perhaps they are not decisive. A complete defence must wait for another time.

The thesis is also incomplete in another respect. This is only a first attempt to articulate an ecological account of liberalism. Disciplines other than political theory are developing a deeper understanding of the nature of rationality and the relationship between human cognition and the environment. These advances suggest that the need to revise some of the associations liberals assume between liberal institutions (such as freedom of speech and association) and liberal values (such as

---

<sup>1</sup> John Rawls, *Political Liberalism* (New York: Columbia University Press, 1996), 64, n. 19.

justice, autonomy, equality). How deep these revisions must be remains an open question. What seems clear, though, is that, as Susan Hurley writes ‘if liberal values are to survive and flourish in the face of these discoveries, we need to begin to link questions about how we make up our minds with questions about how our minds are made up’<sup>2</sup>.

---

<sup>2</sup> Susan Hurley, 'Bypassing Conscious Control: Unconscious Imitation, Media Violence, and Freedom of Speech', in Pockett, S., W. Banks and S. Gallagher (eds.) *Does Consciousness Cause Behavior* (Cambridge, MA: MIT Press, 2006), 327.

## CHAPTER 1

### THE PRINCIPLE OF FREE SPEECH AND THE PUBLIC/PRIVATE DISTINCTION

In this chapter, I explore some of the basic ideas of political morality upon which the thesis will operate. The discussion is necessarily sketchy and short; most arguments are stated and discussed in more detail in later chapters. In the first section I review a popular liberal idea: the harm principle. I examine what constitutes harm with reference to Joel Feinberg's analysis. In the second section, I explore the relationship between the harm principle and free speech. I rehearse two popular strategies for defending the right to unrestricted expression and conclude that a free speech principle needs to be different and independent of other political principles. Among the different defences of speech I highlight one based on autonomy and I explain the reasons for this focus.

In the third section I connect the notion of autonomy with the public/private distinction in two different ways. First, I argue that liberalism assumes a primacy of private responsibility (which I call a principle of discontinuous responsibility), according to which autonomy is a constraint on governmental action. Second, I consider Rawls's *political liberalism* emphasising the contrast between public and

non-public reasons. I argue that both aspects of the public/private distinction appear to support a principle of free speech.

The fourth section presents a number of objections raised against this distinction. The first regards Rawls's views about the 'site' of justice and is based on the work of G.A. Cohen. The second emerges from important developments of Cognitive and Social Psychology. I highlight that, in virtue of bypassing autonomous control, a significant part of our behaviour threatens the possibility of public reason. Also, contrary to what the principle of discontinuous responsibility suggests, rational agency has a public, rather than private, origin. In section five, I consider the importance of racist speech that motivates this thesis.

## I

### The Harm Principle

Many liberals think that the government is barred from interfering with people's activities unless such actions harm third parties. This idea was formulated by John Stuart Mill, who famously wrote that 'the only purpose for which power might rightfully be exercised over any member of a civilized community, against his will is to prevent harm to others'<sup>1</sup>. This is what has become known as the 'harm principle'. In a contemporary treatment, Joel Feinberg argues that in order to constitute a harm (in the relevant sense) an action must be a setback of certain interests<sup>2</sup>.

---

<sup>1</sup> John Stuart Mill, *On Liberty and Other Essays* (Oxford: Oxford University Press, 1991), 14.

<sup>2</sup> Joel Feinberg, *The Moral Limits of the Criminal Law: Harm to Others* (Oxford: Oxford University Press, 1984), 33-6.

Things people are said to have interests in are those components of their well-being in which they have a stake. Accordingly, the person ‘flourishes or languishes as [their interests] flourish or languish’<sup>3</sup>. A person has interests in many different things, and the importance of an interest varies in relation to the way it promotes or thwarts her well-being. One can distinguish between two kinds of interests. On the one hand, a person has an interest in doing things she wishes to do. Call this category ‘volitional interests’. On this late spring evening I want an ice-cream, and my volitional well-being will increase if I have one, and it will increase *it only for that reason*. I recognise that my interest in eating ice-cream is wholly dependant on my craving for it. But other interests are of a different kind. These interests are independent from wants in the sense that a person has them even if she ignores or rejects them. Call these interests ‘critical interests’<sup>4</sup>. A person’s life goes better and her critical well-being is increased if such interests are satisfied, and the reason an individual (in general) desires these things is because she acknowledges they will do so.

John Rawls’s conceptualisation of the person assumes that, as free and equal moral agents, she has two fundamental moral powers: one is the capacity for an effective sense of justice; the other is the capacity to develop a conception of the good. Accordingly, a person has two critical, ‘highest-order’<sup>5</sup>, interests in exercising those powers. First, the capacity for a sense of justice ‘is the capacity to understand, to apply, and to act from (and not merely in accordance with) the

---

<sup>3</sup> Ibid. 34.

<sup>4</sup> Ronald Dworkin highlights this distinction. See Ronald Dworkin, ‘Foundations of Liberal Equality’, in Darwall, S. (ed) *Equal Freedom: Selected Tanner Lectures on Human Values* (Ann Arbor: University of Michigan Press, 1995), 229-34. I will return to this distinction in Chapter 4.

<sup>5</sup> John Rawls, ‘Kantian Constructivism in Moral Theory’, in his *Collected Papers* (Cambridge, MA: Harvard University Press, 1999), 312.



principles of political justice that specify the fair terms of social cooperation'. Second, the capacity for a conception of the good is 'the capacity to have, to revise, and to rationally pursue a conception of the good. Such a conception is an ordered family of final ends and aims which specifies a person's conception of what is of value in human life'<sup>6</sup>.

This second interest is intimately related to the idea of autonomy. One of the conditions of autonomy is that a person should be able to identify and critically revise her ends and goals<sup>7</sup>. The harm principle, then, attributes special significance to these interests and stipulates that actions which threaten or interfere with their satisfaction (or with the satisfaction of other interests instrumental to them) are harmful in the relevant sense. On this definition, a person also has interests in things the setback of which is not relevant for a conception of harm<sup>8</sup>. In the following chapters, I will argue that the interests racists have in speech are not of the kind protected by this principle, and, as such, that restricting the expression of racist views does not constitute harm.

Feinberg's characterisation of the harm principle includes a second condition, according to which an action is harmful if interests are setback in a wrongful manner. This condition seems to be required in order to exclude situations

---

<sup>6</sup> John Rawls, *Justice as Fairness: A Restatement* (Cambridge, MA: Belknap Press, 2001), 18-9. Rawls argues that these interests belong to a political conception of the person, rather than a moral or a psychological conception. See further development in Chapters 2 and 3. See also Rawls, J. *Political Liberalism*, 18-9, 29-35.

<sup>7</sup> Liberal writers such as Galston and Kukathas do not share this concern for autonomy. In the future I refer to as 'liberals' those who regard it as important. Presumably, authors who do not value autonomy would defend speech by other reasons, such as democracy or truth. Cf. William Galston, *Liberal Pluralism: The Implications of Value Pluralism for Political Theory and Practice* (Cambridge: Cambridge University Press, 2002); Chandran Kukathas, *The Liberal Archipelago: A Theory of Diversity and Freedom* (Oxford: Oxford University Press, 2003).

<sup>8</sup> A complete theory of harm needs to give criteria by which we can distinguish which interests are covered by the harm principle. This task, though, extends beyond the limits of this thesis.

in which someone's well-being is worsened as a result of another's action, but in which the result is not morally objectionable, as Robert Nozick's unlucky suitor exemplifies<sup>9</sup>. According to Feinberg, both conditions are necessary for an action to count as harm, and to make permissible governmental intervention.

This condition, however, raises an important problem. By making wrongful action necessary for harm, it moralises the principle in a manner that restricts its use. For instance, imagine Colin, a perfectionist who forces Maggie to go to the opera rather than the movies. Most liberals will agree that Colin harms Maggie by not letting her live according to her own values. Nevertheless, Colin might argue that he does not violate the harm principle. Although he does setback Maggie's interests in cinema, he does not do so wrongfully. After all, going to the opera only makes Maggie's life better. In this respect Feinberg's second condition is undetermined; we still need an elaboration of the criteria for wrongfulness. This difficulty is not crucial to the argument I develop in this thesis because understanding harm as a setback of interests in justice and autonomy suffices for the argument to be sustained. Moreover, the argument does not claim that restrictions on hate speech are acceptable because the views expressed are wrong, but only in virtue of the consequences on citizens' two highest-order moral interests.

The harm principle establishes a minimal presumption for liberty: governmental intervention is justified only when an action set back certain interests (namely, those necessary for developing a sense of justice and living an autonomous life). In other words, an individual should be free to do whatever she

---

<sup>9</sup> Robert Nozick, *Anarchy, State, and Utopia* (Oxford: Blackwell, 1974), 237.

wants except those actions which harm (or put at considerable risk of harm) third parties. As Kent Greenawalt states, ‘government should not prohibit people from acting as they wish unless it has a positive reason to do so. The ordinary reason for prohibiting action is that the action is deemed harmful or potentially harmful in some respect’<sup>10</sup>.

## II

### Speech and Harm: Minimalism and Maximalism<sup>11</sup>

The relation between the harm principle and free speech<sup>12</sup> is not a straightforward one. The idea that speech must be free from intervention could be either an instantiation of the more general harm principle (a minimalist strategy), or an independent principle (a maximalist approach).

#### *Minimalism*

In general minimalists argue that expressive activities are covered by the harm principle<sup>13</sup>: their main claim is that speech does not harm third parties and, being thus consistent with the harm principle, it must be protected. There are two main variants within this strategy. The most radical, ‘no-cost minimalism’<sup>14</sup>, claims that

---

<sup>10</sup> Kent Greenawalt, *Speech, Crime, and the Uses of Language* (Oxford: Oxford University Press, 1989), 9.

<sup>11</sup> This taxonomy comes from Joshua Cohen, ‘Freedom of Expression’, *Philosophy and Public Affairs* 22 (1993), 217-22.

<sup>12</sup> In this thesis I use ‘free speech’, ‘freedom of speech’ and ‘freedom of expression’ interchangeably.

<sup>13</sup> I say ‘in general’ because there might be ways of defending an independent principle of free speech within a minimalist strategy, although minimalism does not require that step. For it, speech must be protected as part of the harm principle. Cf. Cohen, ‘Freedom of Expression’, 218.

<sup>14</sup> This term comes from Susan Brison, ‘Speech, Harm, and the Mind-Body Problem in First Amendment Jurisprudence’, *Legal Theory* 4 (1998), 42-3.

‘there can be no legitimate reason to curtail [freedom of expression] since its possession and its exercise do no one any harm’<sup>15</sup>. Famously, Justice Antonin Scalia, stated that ‘[s]ticks and stones will break my bones [...] but words can never hurt me’<sup>16</sup>. According to this view, speech, unlike other forms of action, cannot cause harm because it is *always* mentally mediated<sup>17</sup>. By mental mediation, it is usually meant that communication is always rationally evaluated, and that the ways in which speech affects agents’ preferences and beliefs are always assessed by critical reflection. In this view, mental mediation neutralises the would-be harmful effects of speech.

Another version of minimalism (‘low-cost minimalism’) holds that, although speech has some costs, these are always redressable by ‘more, and better speech’<sup>18</sup>. The thought here is that whatever harm speech does cause, it can always be outweighed by ‘positive’ speech.

Imagine a person is racially insulted, and, as a result, she is shocked. How would these two versions of minimalism respond to this situation? No-cost minimalism denies that the person is harmed in the relevant sense: advocates would probably accept that she is offended, but will argue that offence does not suffice for governmental intervention<sup>19</sup>. They will also argue that she is responsible for her

---

<sup>15</sup> Joseph Raz, ‘Freedom of Expression and Personal Identification’, in his *Ethics in the Public Domain* (Oxford: Clarendon, 1994), 147. Notice, though, that Raz does not endorse this view.

<sup>16</sup> Quoted by Brison, ‘Speech, Harm, and the Mind-Body Problem in First Amendment Jurisprudence’, 39. Brison also quotes Judge Wood who writes ‘[i]t may also be well to remember that often “words die away – leaving no taste, no color, no smell, not a trace’. Ibid. 43.

<sup>17</sup> The problem of mental mediation will be discussed later, first in relation to responsibility (Chapter 3), and later I will discuss empirical evidence showing that the claim is false; indeed a good amount of speech bypasses mental control.

<sup>18</sup> Brison, ‘Speech, Harm, and the Mind-Body Problem in First Amendment Jurisprudence’, 43. For a list of different –unsuccessful- arguments used in defending minimalism see Ibid. 42-47.

<sup>19</sup> Joel Feinberg, *The Moral Limits of the Criminal Law: Offense to Others* (Oxford: Oxford University Press, 1985).

own reaction: just as she can exercise some control over her preferences and beliefs, she could have controlled her response. Low-cost minimalism, on the other hand, does not deny that she may be harmed, but it argues that harm can be rectified through ‘more, better speech’<sup>20</sup>, by discrediting the racist, or by helping her in other ways. She can be counselled and, through mental effort, the damages will disappear. In this sense, speech is not sufficiently harmful as to warrant interference.

The important point in this discussion of minimalism is that it does not give positive reasons for protecting speech. It merely shows that speech warrants protection under the harm principle. As Susan Brison writes: ‘both “no-cost/low-cost” accounts of speech fail to perform the job they are supposed to do, namely to provide a plausible basis for the First Amendment, since they do not say what is *special* about speech such that it merits heightened constitutional protection’<sup>21</sup>. It seems that if no cost/low cost accounts were tenable, then free speech would be no different to the minimal principle of liberty.

### *Maximalism*

Some people are not convinced by minimalism. They want to protect speech beyond the harm principle. That is, to furnish it with an independent justification. This seems to be the natural way to read, for instance, the First Amendment to the US Constitution<sup>22</sup>. This position holds that it is not enough to show that speech

---

<sup>20</sup> I will discuss this strategy in Chapter 5.

<sup>21</sup> Brison, ‘Speech, Harm, and the Mind-Body Problem in First Amendment Jurisprudence’, 47.

<sup>22</sup> The First Amendment the US Constitution is a paradigmatic example of a strong principle of free speech.

causes harm because there are good reasons to protect it beyond the harm principle.

As Schauer argues

The Free Speech principle is an exception or qualification, of no necessary size or strength, to the general rule in force under a particular political theory [say, the harm principle]. When a Free Speech Principle is accepted, there is a principle according to which speech is less subject to regulation [...] than other forms of conduct having the same or equivalent effects<sup>23</sup>.

Maximalism holds that 'expression merits stringent protection because its great value guarantees that the benefits of protection trump the costs'<sup>24</sup>. According to this view, then, there are reasons to protect speech even when speech violates the harm principle. What are these reasons?

A popular argument is that free speech is a necessary condition if we are to discover truths<sup>25</sup>; truth searching requires no opinion or idea to be excluded from discussion. This argument has been combined with different levels of fallibilism<sup>26</sup>, mainly claiming that, because it is possible for at least some of our beliefs to be false, we should not exclude speech from the 'marketplace' of ideas.

Another argument holds that speech is to be protected in order to preserve democracy. Self-government requires that citizens are able to discuss and deliberate about their beliefs and preferences, and censoring speech can thwart the processes by which democratic decisions are made (for instance, by removing ideas from the

---

<sup>23</sup> Frederick Schauer, *Free Speech: A Philosophical Enquiry* (Cambridge: Cambridge University Press, 1982), 7. However, this view has been criticised by Larry Alexander, *Is There a Right of Freedom of Expression?* (Cambridge: Cambridge University Press, 2005). He examines different rationales for a principle of free speech and concludes that none are convincing. His claim is not minimalist, but essentially originates from the belief that liberal governments cannot be neutral in the sense required for such a principle.

<sup>24</sup> Cohen, 'Freedom of Expression', 220.

<sup>25</sup> Cf. Mill, *On Liberty*, Chapter 2.

<sup>26</sup> Cf. Alan Haworth, 'On Mill, Infallibility, and Freedom of Expression', *Res Publica* 13 (2007), 77-100.

agenda, and so on)<sup>27</sup>. In support of this view, Owen Fiss argues that '[t]he purpose of free speech is [...]the preservation of democracy, and the right of a people, as a people, to decide what kind of life it wishes to live [...]. We allow people speak so others can vote'<sup>28</sup>.

Although these arguments are popular, many liberals<sup>29</sup> instead defend free speech on grounds of autonomy<sup>30</sup>. They think that if individuals are to be respected as moral, autonomous agents, then a strong right of free expression must be guaranteed. Normally this defence of free speech has a natural connection with a maximalist strategy, for these authors think that respecting autonomy trumps the harms speech can cause. As Raz explains, some people think that freedom of expression 'lies at the heart of people's humanity, [it] is a requirement of personhood or rationality'<sup>31</sup>. In this respect, George Kateb argues

Freedom of expression is not, then, properly conceptualised as a means or instrument for persons who exercise it [...]. Whatever the goals we attain by expression are [...] conceptually secondary. The primary notion is that a lot of the time, one is one's expression, one lives to express, one lives by expressing. One does not merely use speech; one *is* one's speech; one's life is mostly speech. [...] To be free, one must express oneself freely and freely receive the

---

<sup>27</sup> Cf. Cass Sunstein, *Democracy and the Problem of Free Speech* (New York: Free Press, 1993); Owen Fiss, *The Irony of Free Speech* (Cambridge, MA: Harvard University Press, 1996).

<sup>28</sup> Owen Fiss, *Liberalism Divided* (Boulder: Westview Press, 1996), 13.

<sup>29</sup> Ronald Dworkin, *Freedom's Law* (Cambridge, MA: Harvard University Press, 1996); Ronald Dworkin, *Sovereign Virtue: The Theory and Practice of Equality* (Cambridge, MA: Harvard University Press, 2000); Ronald Dworkin, *Is Democracy Possible Here?* (Princeton: Princeton University Press, 2006); Thomas Nagel, 'Personal Rights and Public Sphere', in *Concealment and Exposure* (New York: Oxford University Press, 2002), 31-52; Thomas Scanlon, 'A Theory of Freedom of Expression', in *The Difficulty of Tolerance* (Cambridge, MA: Harvard University Press, 2003), 6-25; Thomas Scanlon, 'Freedom Of Expression and Categories of Expression' in *The Difficulty of Tolerance*, 84-112; Rawls, *Political Liberalism*, Lecture VIII.

<sup>30</sup> Susan Brison, 'The Autonomy Defense of Free Speech', *Ethics* 108 (1998), 312-39.

<sup>31</sup> Raz, 'Freedom of Expression and Personal Identification', 147.

free expression of one's fellows. Freedom of expression is not the whole of freedom, but its soul'<sup>32</sup>.

Or, from a listener's perspective, the thought is that, if individuals are to develop a sense of justice and a conception of the good, governments must not prevent them from hearing different opinions or communicating their own.

In this thesis I will primarily address the autonomy-based defence of speech. I do so for a number of reasons. First, the relation between autonomy and speech is intrinsically linked to liberal political morality. As noted above, the liberal conception of the person involves an idea of autonomy<sup>33</sup>. Second, we generally think autonomy has significant value. It might not be the only value and it might not always outweigh others, but it is of great importance nevertheless. If there are reasons of autonomy to protect free speech, then those reasons are weighty, and not easily trumped. If, on the other hand, autonomy does not establish a right to free speech, then the case for the latter is weakened. We would then need to explore other arguments, such as truth or democracy, for its protection. Third, other justifications for free speech are (in part or wholly) dependent on the autonomy defence. This seems especially so in the argument from democracy<sup>34</sup>. One of the reasons we believe democracy to be valuable is that it (more or less) reconciles the ideal of people being free and, at the same time, subject to political authority<sup>35</sup>.

---

<sup>32</sup> George Kateb, 'The Value of Association', in A. Gutmann (ed) *Freedom of Association* (Princeton: Princeton University Press, 1998), 53.

<sup>33</sup> This claim is controversial: I am not arguing that liberals believe autonomy to be the most important (or only) value, but rather that some notion of autonomy is necessary. The intuition that agents are able to reflect and decide (within the constraints of their social circumstances) about what is valuable in their lives is one shared by many liberals. Clearly, the notion of autonomy can be thicker or thinner, political or comprehensive, but, ultimately, they are all conceptions of autonomy.

<sup>34</sup> Cf. Fiss, *Liberalism Divided*, 13. He says that the reasons for speech are democratic and not autonomy dependent.

<sup>35</sup> The classical indictment of this idea comes from Jean-Jacques Rousseau, who writes that the fundamental problem is 'to find a form of association which will defend and protect with the whole



Democracy's appeal is, in essence, a function of the manner in which it promotes and/or respects (some sense of) autonomy. Fourth, some other defences of free expression seem not to apply to hate speech. There is no relation, for instance, between truth and racist speech. If the former were the only reason to defend free speech, it is not inconceivable that liberals would support restrictions to the latter. For them, hate speech is to be protected on grounds of autonomy, not knowledge. Moreover, some forms of racist speech are not themselves presented as truths: a Neo-Nazi parade is a manifestation of attitudes, rather than epistemic propositions. That is why liberals justify protection on the basis of autonomy rather than truth.<sup>36</sup>

One important clarification is required before we continue. Sometimes people make a distinction between speech and other forms of conduct<sup>37</sup>. This distinction is, however, unhelpful in understanding how 'speech' is used in the literature. There are many examples of speech not covered by the free speech principle such as falsely shouting 'fire' in a theatre, speaking loudly in libraries, or yelling in avalanche-risk zones, and so on. These do not count as 'speech' in the

---

common force the person and goods of each associate, and in which each, while uniting himself with all, might still obey himself alone and remain as free as before'. Jean-Jacques Rousseau 'The Social Contract' in *The Social Contract and other Discourses*, (London: Everyman, 1993), 191.

<sup>36</sup> Racist science is indubitably false; there is sufficient scientific evidence showing that the racial differences defended by racists simply do not exist. The same holds for Holocaust Deniers. Their arguments have been thoroughly discredited in academia. Cf. Deborah Lipstadt, *Denying the Holocaust: The Growing Assault on Truth and Memory* (London: Penguin, 1994); Michael Shermer, and Alex Grobman *Denying History: Who Says the Holocaust Never Happened and Why They Say It?* (Berkeley: University of California Press, 2000); Richard Evans, *Lying About Hitler: History, the Holocaust and the David Irving Trial* (New York: Basic Books, 2001); Catriona McKinnon, 'Should We Tolerate Holocaust Denial'. *Res Publica* 13 (2007), 9-28.

<sup>37</sup> Both minimalists and maximalists can accept this distinction, although they will take opposite sides in how it is assessed: the latter will say that speech is too important to be regulated (speech's transcendental value outweighs other forms of 'physical' behaviour), the former will say that speech is innocuous and cannot hurt (as in the 'sticks and stones' view).

relevant sense: in terms of the 'free speech doctrine' they are only noise<sup>38</sup>. On the other hand, other forms of conduct usually not regarded as 'speech' receive protection under a free speech principle: pictures, films, burning flags, silent protestations, graffiti, etc. In this sense 'speech' is a term of art used to describe items that fall under the reasons for protecting speech<sup>39</sup>. The distinction can also be challenged at a more general level by noting that speech is in fact a form of action and that we can 'do things' with it<sup>40</sup>. Through speech we perform actions and bring about new states of affairs. By saying 'I do', one can get married and create new rights and obligations. Under certain conditions uttering some words might scare, intimidate or subordinate others. Conversely, other forms of conduct can have a communicative dimension, and might mean things to us. Preparing a sandwich 'communicates' that I am hungry, rushing out of the library 'means' that I am late, and so on. Trying to understand 'free speech' while employing this distinction is not likely to succeed<sup>41</sup>.

### III

#### Free Speech and the Public/Private Distinction

There is a relation between the public/private distinction and the idea of autonomy important for understanding the liberal defence of free speech. Liberals hold that autonomy places a constraint on the coercive use of government's action. There is a

---

<sup>38</sup> Cf. *R.A.V. v. City of St. Paul, Minnesota*, 505 U.S. (1992) 377.

<sup>39</sup> Hurley, 'Bypassing Conscious Control', 317.

<sup>40</sup> Two classic works on the subject are John Austin, *How to Do Things with Words* (Oxford: Clarendon Press, 1962); and John Searle, *Speech Acts: An Essay in the Philosophy of Language* (Cambridge: Cambridge University Press, 1969).

<sup>41</sup> Scanlon, 'A Theory of Freedom of Expression', 9.

distinction between two realms: the private sphere, in which government has no right to interfere, and the public sphere in which governmental intervention is not illegitimate. Although there are different ways of differentiating these two spheres, for the sake of the argument defended two are especially important.

One sense of the distinction gives private responsibility a special priority over governmental intervention<sup>42</sup>. According to this principle we can distinguish between the direct and indirect effects of an agent's actions. The priority of private responsibility states that a person is to be held responsible only for the direct effects of her actions. It is to be understood, that is, as 'discontinuous'<sup>43</sup>. For instance, John sells a gun to Eva; the following week Eva shoots Jim dead. According to the principle of discontinuous responsibility it is only Eva (not John) who can be held responsible<sup>44</sup>. The reasoning for this is that, although John's action *indirectly* harmed Jim, it was Eva who *directly* killed Jim<sup>45</sup>. An example of the priority of private responsibility argument is Ronald Dworkin's 'ethical individualism'. One of its dimensions holds that 'each person has a special responsibility for realizing the success of his own life, a responsibility that includes exercising his judgement

---

<sup>42</sup> The relation between responsibility and autonomy is that the former is necessary but not sufficient for the latter. Only responsible agents can be autonomous, but not every responsible agent is autonomous.

<sup>43</sup> Susan Hurley calls this principle 'the principle of proximal agency' 'Public Ecology of Responsibility', (2006), Unpublished paper.

<sup>44</sup> This view is different from what Nozick calls the 'bucket theory of responsibility'. This model holds that there is a fixed amount of responsibility for each act. If we hold Eva fully responsible for shooting Jim, there would be no responsibility left to attribute to John. According to this model if we wanted to hold John (partly responsible) for Jim's death, then we would need to subtract that amount of responsibility from Eva's, such as if the amount of responsibility for the act is ' $Tr$ ', then John's responsibility ' $Jr$ ' plus Eva's responsibility ' $Er$ ' must equal ' $Tr$ ' ( $Tr = Jr + Er$ , or  $Jr = Tr - Er$ ). The model is clearly mistaken 'because responsibility is not a bucket in which less remains when some is apportioned out'. Nozick, *Anarchy, State, and Utopia*, 130.

<sup>45</sup> This principle leads naturally to a defence of free speech: A expresses a view. After being convinced by it, B acts in a way that harms C. According to this principle, only B is responsible, not A. Cf. my assessment of T. Scanlon's 'Millian Principle' in Chapter 3.

about what kind of life would be successful for him'<sup>46</sup>. Responsibility and autonomy place limits on governmental action: an individual's responsibility for her actions renders paternalistic policies illegitimate<sup>47</sup>.

Another (related) popular sense in which the public/private distinction is employed within liberalism is connected to Rawls's idea of 'public reason'. He argues that special constraints apply to the reasons citizens can legitimately offer each other when making certain political decisions. Public reasons are those acceptable to reasonable citizens in an 'overlapping consensus'. An overlapping consensus, according to Rawls, means that a 'political' conception of justice 'is supported by the reasonable though opposing religious, philosophical, and moral doctrines that gain a significant body of adherents and endure over time from one generation to the next'<sup>48</sup>. Because Rawls thinks that in a well-ordered society the free use of practical reason will lead to a 'reasonable pluralism' of comprehensive and controversial doctrines, the only way in which a government's use of coercive power might be acceptable to citizens (who disagree about their conceptions of the good) is if it is being justified independently of any of those comprehensive doctrines<sup>49</sup>. A political conception of justice is worked out as to be acceptable to reasonable citizens. To qualify as a 'political' conception, a conception of justice

---

<sup>46</sup> Dworkin, *Is Democracy Possible Here?*, 10. I will have more to say about Dworkin's defence of free speech in Chapter 3.

<sup>47</sup> Or at least make them problematic: in order to be acceptable, paternalism needs special justification. I will discuss further issues related to paternalism in Chapters 3 and 4.

<sup>48</sup> Rawls, *Justice as Fairness*, 32.

<sup>49</sup> This very much simplifies Rawls's position, though I will return to these ideas in greater in due course. See his *Political Liberalism*, Introduction, Lectures II and IV; John Rawls, 'The Idea of Public Reason Revisited' in *The Law of Peoples and the "Idea of Public Reason Revisited"* (Cambridge, MA: Harvard University Press, 1999), 131-80; and John Rawls, 'The Domain of the Political and Overlapping Consensus' in his *Collected Papers*, 473-96.

must have three features. First, its specific subject is the ‘basic structure of a constitutional democratic regime’<sup>50</sup>. Thus, Rawls argues that

the primary subject of justice is the basic structure of society, or more exactly, the way in which the major social institutions distribute fundamental rights and duties and determine the division of advantages from social cooperation. By major institutions I understand the political constitution and the principal economic and social arrangements. Thus the legal protection of freedom of thought and liberty of conscience, competitive markets, private property in the means of production, and the monogamous family are examples of major social institutions<sup>51</sup>.

The specific subject of a conception of justice is distinguished from general moral theories, which ‘appl[y] to a wide range of subjects, and in the limit to all subjects universally’<sup>52</sup>. While a political conception is public (in the sense that its scope is restricted to political actions), a general conception applies to both the public and the private. The latter provides guidelines about how to behave in relation to one’s friends, family, private associations, career, etc.<sup>53</sup>

The second feature of a political conception of justice is that its acceptance ‘does not presuppose accepting any particular comprehensive doctrine’<sup>54</sup>. Due to the ‘fact of reasonable pluralism’<sup>55</sup>, a conception of justice must not be derived from any comprehensive doctrine; otherwise it would not be acceptable to all reasonable citizens. Although it is ‘freestanding’<sup>56</sup>, this does not mean that it cannot

---

<sup>50</sup> Rawls, ‘The Domain of the Political and Overlapping Consensus’, 480.

<sup>51</sup> John Rawls, *A Theory of Justice* (Cambridge, MA: Harvard University Press, 1999), 6.

<sup>52</sup> Rawls, *Political Liberalism*, 13.

<sup>53</sup> I will briefly discuss the problems related with the ‘site’ of justice in section IV.

<sup>54</sup> Rawls, ‘The Domain of the Political and Overlapping Consensus’, 480.

<sup>55</sup> Rawls, *Political Liberalism*, 36.

<sup>56</sup> Rawls, *Political Liberalism*, 13.

be derived from, or supported by, comprehensive doctrines. It does demand, though, that its acceptance does not commit persons to accept any particular doctrine. A political conception can form the basis for an overlapping consensus only if this condition is satisfied.

The third feature of a political conception of justice is that it ‘is formulated as far as possible solely in terms of certain fundamental intuitive ideas viewed as implicit in the public political culture of a democratic society’<sup>57</sup>. Rawls sees his own conception, justice as fairness, as embodying the values of freedom and equality as have been developed in recent centuries<sup>58</sup>. What kind of justifications can citizens offer each other when fundamental political issues arise? Rawls’s answer is that

[c]itizens must be able [...] to present one another publicly acceptable reasons for their political views in cases raising fundamental political questions. This means that our reasons must fall under the political values expressed by a political conception of justice. If free and equal persons are to cooperate politically on a basis of mutual respect, we must justify our use of corporate and coercive power, where those essentials matters are at stake, in the light of public reason<sup>59</sup>.

Among the different aspects of public reason, two are relevant here: the first concerns the relevant sort of questions and subject; the other specifies the persons to whom it applies. First, Rawls argues that the subject of public reason is ‘the public good concerning questions of fundamental political justice [which are two

---

<sup>57</sup> Rawls, ‘The Domain of the Political and Overlapping Consensus’, 480.

<sup>58</sup> In his recent book, Dworkin argues that ‘in spite of the popular opinion, [Americans] actually can find shared principles of sufficient substance to make a national political debate possible and profitable’ Dworkin, *Is Democracy Possible Here?*, 6. Notice, however, that Dworkin’s theory is not a political conception in the Rawlsian sense

<sup>59</sup> Rawls, *Justice as Fairness: A Restatement*, 91.

kinds], constitutional essentials and matters of basic justice'<sup>60</sup>. Second, public reason applies only to the 'public political forum', which comprises 'the discourse of judges in their decisions', 'the discourse of government officials' and 'the discourse of candidates for public office'<sup>61</sup>. It does not apply to the 'background culture' comprised of civic society, private associations, churches, universities, or the media<sup>62</sup>. It is worth noting here that legislators are not bound to honour the ideal of public reason in *all* decisions. If public reason is restricted to questions relating to the basic structure (constitutional essentials and matters of basic justice), then it seems that legislators can use non-public reasons when deciding whether to fund a public park, build a sport centre, a parking lot or an opera venue. In representative democracies, the *ideal* of public reason applies to citizens when voting for their representatives and, if there are any, referenda and other democratic procedures. Ideally, citizens must think of themselves as legislators 'and ask themselves what statutes, supported by what reasons satisfying the criterion of reciprocity, they would think it most reasonable to enact'<sup>63</sup>.

How does the public/private distinction relate to the protection of hate speech? On the one hand, an overlapping consensus demands racist views be excluded from political discussions. Given that citizens reasonably reject the content of those opinions, they cannot be used as justification for policies. In this sense, hate speech will not be tolerated in the public sphere. On the other hand, however, and perhaps more importantly, it seems that within the background culture there could be no restrictions on speech. If there are sound reasons to defend

---

<sup>60</sup> Rawls, 'The Idea of Public Reason Revisited', 133.

<sup>61</sup> Ibid. 133.

<sup>62</sup> Ibid. 134.

<sup>63</sup> Ibid. 135.

a free speech principle then it can accommodate hate speech in the non-public realm. Persons, qua individuals who endorse a racist conception of the good, are not to be restricted in expressing their views, it would seem, provided that these are not enacted by legislatures in constitutional essentials and matters of basic justice.

#### IV

#### Some Worries about the Public/Private Distinction

In this section I will address two concerns surrounding the public/private distinction. The first worry rehearses familiar critiques made by some feminists and a powerful variant professed by G.A. Cohen<sup>64</sup>. The second is rather different; it is motivated by recent developments in Social and Cognitive Psychology.

##### *The 'Personal is Political'*

Feminists have argued that the private sphere, and especially the domestic realm, is politicised<sup>65</sup>. To claim that domestic relations are unrelated to justice amounts to condoning the inequalities present in gender relations. For instance, they have argued that the unequal division of domestic labour between men and women affects the principle of fair equality of opportunity by placing the latter in a disadvantaged position. For some, the institution of marriage, in virtue of being a

---

<sup>64</sup> G. A. Cohen, 'Where the Action Is: On the Site of Distributive Justice'. *Philosophy and Public Affairs*, 26 (1997), 3-30. See also his 'Incentives, Inequality, and Community', in Darwall, S. (ed.) *Equal Freedom* (Ann Arbor: Michigan University Press, 1995), 331-97; 'The Pareto Argument for Inequality', *Social Philosophy and Policy* 12 (1995), 160-85; *If You're an Egalitarian How Come You're so Rich?* (Cambridge, MA: Harvard University Press, 2000).

<sup>65</sup> Susan Okin, *Justice, Gender, and the Family* (New York: Basic Books, 1989); Susan Okin, 'Gender, the Public and the Private', in Held, D. (ed.) *Political Theory Today*. (Cambridge: Polity Press, 1991); Iris Young, 'Taking the Basic Structure Seriously', *Perspective in Politics* 4 (2006), 91-97.



voluntary association between competent adults, is an area in which unequal practices are not a concern for justice (insofar as rights are not violated –obviously marital rape and other forms of violence are condemned). Feminist writers argue, however, that due to gender inequalities, private practices are translated into inequalities in the public sphere. Think, for instance, of how some parents choose to invest more resources in a son's education, leaving daughters with fewer opportunities for career development. Or of how many women work part time (in worse paid jobs), because they carry the burdens of childrearing.

Another criticism, launched by Okin, claims that a gendered family is unjust to children because they are born into it involuntarily. Basic socialisation, the argument holds, takes place (usually) within families, and children (both boys and girls) educated in a sexist environment will fail to develop a sense of justice required to be autonomous citizens<sup>66</sup>. Families nurture future citizens and that provides a strong reason to have just families<sup>67</sup>.

G. A. Cohen employs the form of the feminist critique against the public/private distinction in his critique of Rawls's difference principle<sup>68</sup>. His main claim is that 'choices not regulated by the law fall within the purview of justice'<sup>69</sup>. Cohen believes that there is an inconsistency in Rawls's work. The difference principle holds that 'inequalities are just if and only if they are necessary to make

---

<sup>66</sup> The upshot of the argument is not limited to justice between the sexes; it covers other instances such as religion. Cf. Matthew Clayton, *Legitimacy and Justice in Upbringing* (Oxford: Oxford University Press, 2006).

<sup>67</sup> Cf. Okin, *Justice, Gender, and the Family*, 17, 131. I will examine the relationship between moral development, justice and free association in Chapter 5.

<sup>68</sup> Cohen's work has motivated a number of responses, among which see Andrew Williams, 'Incentives, Inequality and Publicity', *Philosophy and Public Affairs* 27 (1998), 225-47; Joshua Cohen, 'Taking People as They Are', *Philosophy and Public Affairs* 30 (2001), 363-86; Samuel Scheffler, 'Is the Basic Structure Basic?' in Sypnowich, Ch. (ed.) *The Egalitarian Conscience* (Oxford: Oxford University Press, 2006), 102-29.

<sup>69</sup> Cohen, 'Where the Action Is: On the Site of Distributive Justice', 4.

the worst off people in society better off than they would otherwise be'<sup>70</sup>. Cohen argues that there are two ways of reading the idea of 'necessary inequalities'<sup>71</sup>: a 'lax' interpretation and a 'strict' one. The latter implies that talented people *need* the extra money in order to produce more (and to benefit the worst-off). Without the extra they simply *could not* be more productive. The lax interpretation, on the other hand, suggests that talented people are not willing to work harder unless they are paid better. They *could* produce the extra benefits for the worst-off, but they do not want to. Under this interpretation, it seems that a society in which incentives for inequalities are necessary is not a just society. For how can the same person affirm and uphold correct principles of justice and simultaneously deem inequalities necessary to elicit his most productive performance<sup>72</sup>?

Cohen believes that the natural answer open to Rawls is the 'basic structure objection': the latter might say that because principles of justice apply only to the basic structure, they leave individual choices outside the purview of justice. The idea is that qua citizen, a person must honour the principles of justice but in his daily life he is free to act in a different fashion.

Cohen criticises this response by noting that there is an important ambiguity within the definition of the basic structure. Rawls seems to suggest sometimes that 'coercive (in the legal sense) institutions exhaust it, or, better, that institutions

---

<sup>70</sup> Ibid. 5.

<sup>71</sup> See also his *If You're an Egalitarian How Come You're so Rich?*, 127.

<sup>72</sup> For it is her denial to produce more without the incentives which generates the 'just' inequalities. Cohen writes: 'for [the talented] could not claim that, in self justification, at the bar of the difference principle, that their high rewards are necessary to enhance the position of the worst-off, since [...] it is they themselves who make those rewards necessary, through their unwillingness to work for ordinary rewards as productively as they do for exceptionally high ones, an unwillingness which ensures that the untalented get less than otherwise would'. *If You're an Egalitarian How Come You're so Rich?*, 127

belong to it only in so far as they are (legally) coercive'<sup>73</sup>. This interpretation has some force, for it is clear that coercion requires special justification<sup>74</sup>. However, it is not convincing. Despite coercion being an important moral concern, taking it as the defining element seems arbitrary. For there are non-coercive institutions whose effects in people's lives are equally as important as the coercive ones. We are thus led to the second interpretation of the basic structure. Rawls says '[t]he basic structure is the primary subject of justice because its effects are so profound and present from the start'<sup>75</sup>. Partly, this profound impact affects '[society's] members and determines in large part the kind of persons they are'. It fosters and limits certain ambitions, desires and talents while discouraging others. 'More generally, the basic structure shapes the way the social system produces and reproduces over time a certain form of culture shared by persons with certain conceptions of their good'<sup>76</sup>. Even though they are not coercive, many of these private institutions apply to us involuntarily, and this fact seems to make the need for special justification compelling. Legal coerciveness cannot be a necessary condition (although it might be sufficient) for inclusion as part of the basic structure. So, Cohen concludes, Rawls must 'admit application of the principles of justice to (legally optional)

---

<sup>73</sup> Cohen, 'Where the Action Is', 18. For instance Rawls writes: 'The basic structure of society is the way in which the main political and social institutions of society fit together into one system of social cooperation, and the way they assign basic rights and duties and regulate the division of advantages that arises from social cooperation over time [...]. The political constitution with an independent judiciary, the legally recognized forms of property, and the structure of the economy (for example, as a system of competitive markets with private property in the means of production), as well as the family in some form, all belong to the basic structure. The basic structure is the background social framework within which the activities of associations and individuals take place. A just basic structure secures what we may call background justice' Rawls, *Justice as Fairness: A Restatement*, 10.

<sup>74</sup> For a sympathetic reading of this interpretation, see Scheffler, 'Is the Basic Structure Basic?'

<sup>75</sup> Rawls, *A Theory of Justice*, 7.

<sup>76</sup> Rawls, *Political Liberalism*, 269.

social practices, and indeed, to patterns of personal choice that are not legally prescribed'<sup>77</sup>.

### *Political Persons and Public Rules*

There is, however a problem with Cohen's argument. Rawls argues that 'the underlying unity [of the basic structure] is provided by the idea that free and equal moral persons are to construct *reasonable* and *helpful guidelines* for moral reflection in view of their need for such organizing principles and the role in social life that these principles and their corresponding subjects are presumed to have'<sup>78</sup>. Principles of justice, then, must be public in the sense that they can guide people with confidence<sup>79</sup>, and their fellow citizens can see others being guided by these principles. It is imperative that the principles of justice provide public rules to citizens acting in good faith<sup>80</sup>. The requirement of publicity helps to understand one aspect of the relationship between citizens and the basic structure: citizens might be better able to comply with the reasons of justice which apply to them than following their own judgement<sup>81</sup>. So the basic structure objection is not an avenue for citizens to opt for lesser obligations of justice; rather it is more effective means of ensuring compliance.

There are, then, choices and patterns of behaviour that cannot be included as part of the basic structure because they cannot be public in the sense of providing reasonable and helpful guidance to citizens. Nevertheless, they produce unjust

---

<sup>77</sup> Cohen 'Where the Action Is', 22.

<sup>78</sup> Rawls, *Political Liberalism*, 262, emphasis added.

<sup>79</sup> I defend the 'full publicity condition' in Chapter 4.

<sup>80</sup> Williams, A. (1998) 'Incentives, Inequality, and Publicity'.

<sup>81</sup> Here I draw from Joseph Raz's normal justification of authority. See *The Morality of Freedom* (Oxford: Clarendon, 1986), Chapters 2-4.

states of affairs: fair equality of opportunity might be frustrated, some citizens might not reach the ‘social basis for self-respect’<sup>82</sup> required by justice, and so on. These results might be the outcome of social prejudices that are implicit in the background culture; not ‘from public policies of explicit distinction and exclusion, but rather from widespread but relatively unnoticed assumptions embedded in institutional rules, the material infrastructure of social action, and everyday habits of comportment and interaction’<sup>83</sup>. Even though (some of) these patterns might not part of the basic structure, justice cannot ignore them. Justice requires that we find public ways to neutralise these effects. In part, the argument of this thesis provides an answer to such concerns.

*Automatic Behaviour and Situated Rationality.*

There are two further problems that afflict the public/private distinction. In contrast with the comments above, which were normative, both of these are naturalistic. They emanate from recent developments in Social and Cognitive Psychology.

**Automatic Behaviour and the Public/Private Overlap.** There is a dominant understanding of human rational agency is one in which perceptions are linked to actions indirectly<sup>84</sup>. A person perceives something and through rational processes he arrives at beliefs about the world. Later those beliefs, through rational processes, direct actions. It is assumed that the relation between perceptions and actions is

---

<sup>82</sup> Rawls argues that ‘perhaps the most important primary good is self-respect’ Rawls, *A Theory of Justice*, 386.

<sup>83</sup> Young, ‘Taking the Basic Structure Seriously’, 96. Cf. Cohen, *If You’re an Egalitarian How Come You’re so Rich?*: ‘While the law may play a large role in securing [the social basis of self-respect] to people vulnerable to racism, legally unregulable racist attitudes also have an enormous negative impact on how much of that good primary good they get’, 211, n. 19.

<sup>84</sup> Susan Hurley calls this feature ‘decoupling’. See ‘The Public Ecology of Responsibility’.

always mediated by reasoning. Perceptions form beliefs, which are evaluated according to the agent's goals.

There is some evidence however, suggesting that this picture is not accurate. Recent research indicates that a significant part of our behaviour is not 'decoupled', but automatically caused by features of the environment. In other words, in many cases perception causes behaviour directly, without rational mediation<sup>85</sup>. For instance, merely thinking about performing an action increases the likelihood of you actually performing that action<sup>86</sup>. Engaging in some activity automatically triggers goals associated with that pursuit. And the mere perception of a piece of behaviour increases the likelihood of an agent acting likewise without the intention of doing so<sup>87</sup>. This phenomenon, the 'chameleon effect', shows that 'the perception of another's behaviour (be it facial expression, body posture, mannerism, etc) increases the tendency for the perceiver to behave in a similar manner, and that this is an entirely passive and nonconscious phenomenon'<sup>88</sup>.

The idea behind this 'ideomotor principle' –that 'every representation of movement awakes in some degree the movement it represents'<sup>89</sup>— is that 'perception is for doing'. It is likely that perceptive systems evolved in order to make responses to the environment efficient, rather than to understand the world.

---

<sup>85</sup> Cf. John Bargh and Erin Williams, 'The Automaticity of Social Life', *Current Directions in Psychological Science* 15 (2006), 1-4; John Bargh and Tanya Chartrand, 'The Unbearable Automaticity of Being', *American Psychologist*, 54 (1999), 462-79.

<sup>86</sup> Ap Dijksterhuis and John Bargh, 'The Perception-Behavior Expressway: Automatic Effects of Social Perception on Social Behavior', *Advances in Experimental Social Psychology*, 33 (2001), 1-40; Tanya Chartrand and John Bargh, 'The Chameleon Effect: The Perception-Behavior Link and Social Interaction' *Journal of Personality and Social Psychology*, 76,(1999), 893-910; John Bargh Mark Chen and Lara Burrows, 'Automaticity of Social Behavior: Direct Effects on Trait Construct and Stereotype Activation on Action', *Journal of Personality and Social Psychology*, 71 (1996), 230-44; Hurley, 'Bypassing Conscious Control'.

<sup>87</sup> See Hurley 'Bypassing Conscious Control', 310-2.

<sup>88</sup> Chartrand and Bargh, 'The Chameleon Effect', 897.

<sup>89</sup> Hurley 'Bypassing Conscious Control', 310.

Having an 'expressway' between perception and action produces *more effective* responses than requiring rational intermediation. In evolution, animals developed perception systems that lead directly to reactions. For instance, frogs have a perceptual system that allows them to perceive and hunt small objects. Every time a frog perceives a small object moving irregularly on the pond's surface, the frog reacts and attacks it<sup>90</sup>. Human beings, in contrast, do not usually react in such fashion, because we have evolved ways of inhibiting some of those simple reactions<sup>91</sup>. However, this does not mean that 'old modules do not suddenly cease to exist; rather than that some new function is added'<sup>92</sup>. It is not only behaviour which can be caused automatically; there is also evidence showing processes of belief formation can be caused directly by things we perceive. Experiments show that, under certain circumstances, people tend to believe information without subjecting it to rational scrutiny<sup>93</sup>.

These examples of automatic behaviour challenge the public/private dichotomy in a fundamental way by casting doubts on some factual premises implicit in this distinction. In essence, the idea is that people's actions are not completely under their control and, seemingly, they are not as able to react in accordance with what is required in different realms as the distinction assumes. Imagine that Tom is a highly sexist person at home and at the local private club, but

---

<sup>90</sup> Dijksterhuis and Bargh, 'The Perception-Behavior Expressway', 3-4.

<sup>91</sup> An explanation of why this phenomenon happens is that 'perception and action share neurological systems. This means that the translation of perception into corresponding action is a consequence of the way we are wired'. Dijksterhuis and Bargh, 'The Perception-Behavior Expressway', 6, 7-8; and Hurley, 'Bypassing Conscious Control', 312-3 for more neurophysiologic evidence

<sup>92</sup> Dijksterhuis and Bargh, 'The Perception-Behavior Expressway', 5.

<sup>93</sup> Daniel Gilbert, 'How Mental Systems Believe' *American Psychologist*, 46 (1991), 107-19; Daniel Gilbert, 'The Assessment of Man: Representation and the Control of Belief', in Wegner, D., and J. Pennebaker, (eds.) *Handbook of Mental Control* (Englewood Cliffs, NJ: Prentice Hall, 1993). Cf. Chapter 5.

when he is at his local council discussing employment policies he tries to respond in a non-sexist way. Given what we have just reviewed regarding about automatic behaviour, it is likely that some sexist responses will be triggered without his awareness. In this sense, the distinction between a person's public and private identities becomes dubious.

**Rethinking Rationality.** There is also another naturalistic issue worth exploring. Recent developments in theories of rationality suggest that rationality has a public, rather than a private origin as it has been traditionally thought.

A view sometimes associated with liberalism holds that rationality has fundamentally a private dimension<sup>94</sup>. This is a common assumption in theories of rational choice and expected utility<sup>95</sup>. Criticisms have been made of some of these assumptions. It has been shown that people often fail to reason according to the rules of probability and Bayesian rules<sup>96</sup>. For instance, people 'make estimates by starting from an initial value that is adjusted to yield the answer'<sup>97</sup>, even if the initial value is arbitrary. In this sense people's judgements get 'anchored' to information that might not be (and often is not) relevant. Another common failure is the potentially variant responses elicited by different formulations of the same problem. People tend to value a medical treatment less, for instance, when it is

---

<sup>94</sup> Cf. Charles Taylor, *Philosophy and the Human Sciences* (Cambridge University Press, 1985); Michael Sandel, *Liberalism and the Limits of Justice* (Cambridge: Cambridge University Press, 1998).

<sup>95</sup> Cf. Hurley, 'The Public Ecology of Responsibility'.

<sup>96</sup> Gerd Gigerenzer, *Adaptive Thinking: Rationality in the Real World* (Oxford, Oxford University Press, 2000), 61.

<sup>97</sup> Daniel Kahneman and Amos Tversky, 'Judgement Under Uncertainty: Heuristics and Biases', in Kahneman, D., P. Slovic, and A. Tversky (eds.) *Judgement Under Uncertainty: Heuristics and Biases* (Cambridge: Cambridge University Press, 1982), 14.



described in terms of mortality than when it is described in terms of the number of people saved, even though the two options are logically similar<sup>98</sup>.

As a response to these problems, the ‘simple heuristics’ research program proposes that we should revise our conception of rationality (rather than dismissing failures to apply the rules of probability consistently as examples of irrationality)<sup>99</sup>. The aim of this program is to find ‘fast and frugal’ computational procedures that are reliable in a given environment, rather than having complicated and demanding, general-domain, rules. Gigerenzer coins the term ‘ecological rationality’ to describe how fast and frugal reasoning can match the structure of environments and yield reasonable and reliable outcomes. The key to their success is that they are domain-specific<sup>100</sup>; they work in some environments but not in others. In contrast, probability rules cannot make this trade-off; they must work in every domain<sup>101</sup>. Ecological rationality explains how systems with limited mental resources and limited time can make good decisions and inferences in particular environments.

Some authors argue that the mind is constituted by different autonomous cognitive modules<sup>102</sup>:

[a]ccording to the massive modularity hypothesis [...] the mind is to a large extent made up of a variety of domain- or task-specific cognitive mechanisms or ‘modules. [...] A cognitive module is an autonomous mind-brain device

---

<sup>98</sup> Daniel Kahneman and Amos Tversky, ‘Choices, Values and Frames’ in Kahneman, D., and A. Tversky (eds.) *Choices, Values and Frames* (Cambridge: Cambridge University Press, 2000).

<sup>99</sup> See especially the collection of essays in Gerd Gigerenzer and Peter Todd, *Simple Heuristics that Make Us Smart* (New York: Oxford University Press, 1999). For criticisms to this program see Kim Sterelny, *Thought in a Hostile World: The Evolution of Human Cognition* (Oxford: Blackwell, 2003).

<sup>100</sup> Gigerenzer, *Adaptive Thinking*, Ch. 10.

<sup>101</sup> Gerd Gigerenzer and Peter Todd, ‘Fast and Frugal Heuristics: The Adaptive Toolbox’, in *Simple Heuristics that Make Us Smart*, 18-9.

<sup>102</sup> Cf. John Tooby and Leda Cosmides, ‘The Psychological Foundations of Culture’, in J. Bakow, L. Cosmides, and J. Tooby, (eds.) *The Adapted Mind* (New York: Oxford University Press, 1992).

characterized by specific inputs from which it derives specific outputs through its own procedures. A module is autonomous not only in the way it functions but also in its phylogenetic and ontogenetic development, which are distinct from that of other models, and also in its failures, which can be quite diagnostic<sup>103</sup>.

The rationale behind this hypothesis is that it seems likely the human mind evolved in order to solve specific problems. In this picture, the mind resembles an adaptive tool-box; it has different modules for solving different problems. These adaptations are domain-specific rather than domain general. For instance, Leda Cosmides and John Tooby argue that people have a domain-specific adaptation to detect cheating in social exchanges. People often fail to apply deductive rules such the ‘modus ponens’ (during experiments using the Wason task<sup>104</sup>). However, if the task is framed as involving social contracts and cheating they tend to do significantly better<sup>105</sup>. Although the simple heuristics and the massive modularity hypothesis are different, they seem to be compatible; cognitive modules might be ecological in the same sense that fast and frugal heuristics are; they evolved in order to solve specific problems related to particular environments<sup>106</sup>.

---

<sup>103</sup> Dan Sperber and Lawrence Hirschfeld, ‘Culture and Modularity’ in Carruthers, P., S. Laurence, and S. Stich, (eds.) *The Innate Mind II: Culture and Cognition* (Oxford: Oxford University Press, 2006), 157. Phylogenesis refers to the sequence of events involved in the evolutionary development of a species or a taxonomic group. Ontogenesis refers to the process of an individual organism developing from a simpler to a more complex level.

<sup>104</sup> I will return to the ‘Wason task’ in Chapter 2, section 6F.

<sup>105</sup> Leda Cosmides and John Tooby, (1992) ‘Cognitive Adaptations for Social Exchanges’, in *The Adapted Mind*, 163-228. But their claims are controversial see Nick Chater and Mike Oaksford, ‘Human Rationality and the Psychology of Reasoning: Where Do We Go From Here?’ *British Journal of Psychology* 92 (2001), 193-216.

<sup>106</sup> Cf. Peter Carruthers, ‘Simple Heuristics Meet Massive Modularity’, in Carruthers, P. Stephen Laurence, and Stephen Stich, (eds.) *The Innate Mind: Culture and Cognition* (Oxford: Oxford University Press, 2006).181-98.

Despite the appeal of these hypotheses, there are difficulties regarding their flexibility in social environments. Solving the sorts of problems that nature presents to individuals can be relatively easy; environments are stable and can be exploited by specific heuristics. But social environments are much more dynamic and unreliable; they change according to the actions of other agents and present problems of deception, free-riding, and so on. In contrast with natural environments, simple heuristics can perform badly if they cannot cope with the changeability of social environments<sup>107</sup>. Some have argued that having general domain rules function better in solving social problems<sup>108</sup>. This suggests that formal rules might emerge due to demands of social ecology; a race in which one competes with others might explain, for instance, the emergence of processes such as imitation and mind-reading<sup>109</sup>. These processes, combined with general domain rules, are sufficiently flexible to cope with the difficulties of social life. If this view is plausible, then, rationality has a public, rather than a private origin. The traditional picture of rationality is reversed; publicity comes first, privacy later.

## V

## Racial Thinking, Cognition and the Costs of Speech

---

<sup>107</sup> Sterelny, *Thought in a Hostile World*.

<sup>108</sup> Dan Sperber, 'Modularity and Relevance: How Can a Massively Modular Mind be Flexible and Context-Sensitive?' in Carruthers, Peter, Laurence, Stephen and Stich, Stephen (eds.) *The Innate Mind I: Structure and Contents*, (Oxford: Oxford University Press, 2005), 53-68.

<sup>109</sup> Susan Hurley suggests the idea of 'social heuristics'. See her 'Social Heuristics that Make Us Smarter', *Philosophical Psychology*, 18 (2005), 585-612.

In this last section, I will comment briefly on the issue of racist speech and the reasons this thesis focuses on it. Racist speech and racial categories seem especially troublesome in contemporary politics because they build upon pervasive features of our society. First, despite many efforts, racial minorities still suffer massive disadvantages in society. Incidents of racial abuse and racial crime have increased in recent years, and the possibility of substantive integration between races looks grim<sup>110</sup>. Moreover, it seems that racial thinking is not only a product of history, but that it is linked in a more fundamental way to our cognitive architecture.

### *Categorising Race*

Lawrence Hirschfeld argues that human beings have a domain-specific module for categorising ‘human kinds’, which is independent from the categorisation of both non-living things and non-human animals<sup>111</sup>. He argues that

Humans [...] form knowledge of human types on the basis of outward appearance. Two sorts of appearance, gender and race, have been found to be particularly salient [...]. Like other natural categories, gender and racial categories seem to capture nonobvious similarities that can be recruited to extend knowledge. However, in contrast to the benign way their natural categories promote inference, these cognitive simplifications –usually called

---

<sup>110</sup> See the collection of essays Fred Pincus and Ehrlich Howard (eds.) *Race and Ethnic Conflict* (Boulder: Westview Press, 1994); see also Andrew Hacker, *Two Nations: Black and White, Separate, Hostile, Unequal* (New York: Ballantine Books, 1995)

An example of how racial integration is being threatened by right-wing judges in the U. S, see *Parents Involved in Community Schools v. Seattle School District No. 1 et. al.* 551 U.S. \_\_\_\_2007.

<sup>111</sup> Lawrence Hirschfeld, *Race in the Making: Cognition, Culture, and the Child's Construction of Human Kinds* (Cambridge, MA: MIT Press, 1996). The idea of nonobvious similarities means that there are alleged correlations between different observable and non-observable features. For instance, racist thinking attributes moral and psychological dispositions to people according to their skin colour.

stereotypes- create patterns of belief that have undesirable social and political effects<sup>112</sup>.

In contrast with ‘historical’ explanations of racial thinking, Hirschfeld argues that racial categories are present in many different societies across diverse locations and times. These observations suggest that the existence of racial categories is not *entirely* caused by socio-historical factors emerging from the last few centuries (although since the 1700s there have been many attempts to systematise racial thinking into a scientific body of knowledge<sup>113</sup>); rather they may have deeper roots.

It is significant that both adults and young children construct racial categories with significant ease, and they seem to do so naturally. Hirschfeld conducted a number of experiments to test whether young children construct racial categories on their own<sup>114</sup>. He showed drawings to young children (aged between 3 and 7) of adults: some black individuals, some white; some male, some female. A third distinctive feature was that a number were dressed as doctors and the others as police officers. Lastly, the drawings also varied in body build, some individuals being stouter than others. (Notice that all the features are visual perceptions, and they all have stereotype associations). The children were presented with drawings including one baby and a couple (a male and a female) of adults. The babies shared one feature with each adult. For instance, if there was a black male baby dressed as a doctor one adult was a white doctor, and the other was a black police officer. In this case, the children were asked to say which adult the baby would become. In the second task, the children were presented with pairs of adults who shared one feature

---

<sup>112</sup> Ibid. 9.

<sup>113</sup> For a brief history of racist thought see Ali Rattansi, *Racism: A very short introduction* (Oxford: Oxford University Press, 2006).

<sup>114</sup> The experiments are reported and discussed in Hirschfeld, *Race in the Making*, Ch. 4.

with the baby. For instance, a stout white baby was showed a couple of stout black adults and a couple of thin white adults. Children were asked to say which couple were the baby's parents. The results showed that children were more likely to make associations based on race than any other characteristics<sup>115</sup>. It is likely that they think that race is an inherent feature that remains over time (from being a baby to becoming an adult), and also that it is inherited (black couples have black children and white couples have white children). In contrast, inheritance and identity do not appear to be prize significant associations in correlation with hair colour or texture. More interestingly still is the fact that when the drawings in experiments displayed non-human animals and objects (such as cars), colour was not consistently the dominant association.

Another experiment was designed to investigate whether children essentialise race. Young children were told a story about a baby (in some stories the baby was black, in others it was white) who was switched at birth to another home (the black baby to a white home and vice versa). Children were then shown drawings of a black and a white adult, and they were asked to say which one the baby had become. They almost always made the claim that it was the adult of the same colour. In most cases, children 'naturalise' race; they believe it cannot be changed through nurture. These experiments show that children have a tendency to identify race as a feature that defines a person's identity<sup>116</sup>; it is natural and inherited. In a sense, race is used as describe the kinds of people there are in the world<sup>117</sup>.

---

<sup>115</sup> Race was chosen over profession, which was preferred over body built. See *ibid.* 97.

<sup>116</sup> 'Identity' not in a moral sense, but as the feature that makes an entity to be the same over time.

<sup>117</sup> Hirschfeld, *Race in the Making*, 106.

The ease with which people construct racial categories is somewhat puzzling. Children in the experiments do not ‘discover’ race, they construct it. Racial categories cannot be explained by the phenomenon of ‘race’ because we know that ‘[h]umans are not discriminable into discrete, self-evident biological kinds’<sup>118</sup>. Simply, ‘race’ has no biological basis at all. This does not mean that ‘there are not biologically grounded differences in human external anatomy [but] that races as socially defined do not (even loosely) capture interesting clusters of these differences’<sup>119</sup>. As Naomi Zack explains:

The genes for bone structure, skin shade, and hair texture, which are associated with social and racial membership do not all get inherited together. There is greater variation within any race of these ‘racial’ traits than between or among any of the recognized races, as groups. Within the human species, as a whole, the genetic variations that occur randomly –that is, between any two people- constitute 0.2% of all human genetic material. Of that 0.2%, or 1/1500, 85% occurs locally, or between any two individuals who happen to be neighbors, 7% occurs within races, and 8% occurs between races. Thus, the amount of human genetic difference due to difference in race is 8% of 0.2% or less than 1/6000 [...]. Of course, that 1/6000, although small could be important. But there is no evidence that it is. No racial essence has ever been identified. There are no general genes for race, such that, once identified, their presence could be used to predict more specific, or secondary, racial characteristics. None of the physical characteristics associated with racial difference in human talent, function or skill<sup>120</sup>.

---

<sup>118</sup> Ibid. 3.

<sup>119</sup> Ibid. 4.

<sup>120</sup> Naomi Zack, ‘Race and Racial Discrimination’. in LaFollette, H. (ed.) *The Oxford Handbook of*

Racial thinking, the widely shared belief that human beings can be partitioned into natural groups, is, according to Hirschfeld, parasitic on a domain-specific competency for perceiving human kinds. However, he is not suggesting that we have a 'race' module. This 'race' module could not exist because during human evolution populations (at least those which were in contact) were not as physically diverse as we are now. His suggestion is, rather, that race categories exploit the tendency we have to construct human kinds and to associate those kinds with intrinsic and essential properties. As he argues, '[r]ace is [...] a unique sort of belief, in significant measure unlike any other commonsense notion. This uniqueness is due in part to the interaction of historical and cultural particulars [...]. Nevertheless, [...] race is an extraordinarily widely encountered notion whose recurrence across varied cultural and historical landscapes derives from deep-rooted psychological processes'<sup>121</sup>.

### *Racial Thinking and Hate Speech*

It is important to note that racial thinking enables the possibility for a person engaging in racism. Although they are not the same, they are related: for racism exploits racial thinking by ascribing moral and political significance to racial categories. If this view is plausible, then the problems of racialism (the tendency to think along racial lines), and racism in particular, are more troublesome than we might at first think. In particular, it suggests that they will be difficult to eradicate. We have a vulnerability to racial thinking, which will not fade away easily or naturally. Although racial categories are not innate and racism is not inescapable,

---

*Practical Ethics* (Oxford: Oxford University Press, 2003), 259.

<sup>121</sup> Hirschfeld, *Race in the Making*, 12.



we are prone to them. In a sense, we have a blind-spot to our racialised perceptions of human beings. Just as children have susceptibility to certain diseases and need to be treated with vaccines, it seems that we might have a similar susceptibility to certain ideas, and we too might need some treatment to deliver protection.

This thesis elaborates an argument which claims that restrictions on racist and gender-based speech are compatible with the liberal account of political morality explored in the third section of this chapter. The argument provides an account of personal autonomy which highlights the need to control exposure to racist speech as a means of protection against our susceptibility to racial and racist thinking.

Although there is plausibility in Hirschfeld's theory, it is not decisive for my thesis. It does accommodate the intuition that race and racism have a central place in our society, and it offers a reasonable explanation for this. Moreover, it is coherent with the massive-modularity hypothesis and compatible with developments in socio-biology. However, the argument of this thesis is independent of the validity of Hirschfeld's cognitive theory of race. It is sufficient for my purposes that racial stereotypes and discrete forms of racism are widespread on our societies, even if they are caused by socio-historical factors alone. On the other hand, if Hirschfeld is right, then we have an even stronger reason to be deeply concerned with the effects and endurance of racial and racist thinking.

However, some might object that the phenomena of automatic behaviour, racial stereotypes, and cognitive biases (what I will call 'mental contamination'<sup>122</sup>) upon which the argument is built proves too much. According to this objection, if

---

<sup>122</sup> See Chapter 2.

the argument if generalised to spheres other than hate speech it would imply an unreasonable amount of restrictions on speech.

The argument I defend, however, is aimed at racist speech and might not be extended to other spheres. Although I acknowledge that the psychological and cognitive biases that sustain the argument have a more general scope, the argument here holds that race and gender are special cases. First, if Hirschfeld's theory is right, then there is a direct link between human cognition and racial thinking, a link that makes protection against it more urgent than in other domains. Second, race and gender inequalities are significantly larger in our societies than most (if not all) others. An important portion of these are (partly) produced by automatic processing as well as by actions in the 'private' realm. Reducing these inequalities in a manner compatible with justice is therefore imperative. Third, citizens have an important interest in living autonomously and justly. The argument I develop here claims that although mental contamination affects us in many different areas, we need to *categorise* the importance of those areas. It seems more important to reduce the extent of racist and sexist attitudes (both those cultivated autonomously and non-autonomously) than, for instance, attitudes regarding our tastes in cinema or food. I suggest that we begin by considering cases we regard as more important. Perhaps later, we should extend the argument to cover other areas, perhaps not. The decision, it seems, must be based on an assessment of each case. Such evaluation, however, must wait for another time and should not be inferred from this writing: here we are solely concerned with racist and sexist speech.

## CHAPTER 2

### AUTONOMY, MENTAL CONTAMINATION AND FREE SPEECH<sup>1</sup>

A distinctive feature of liberalism is its claim that governments must respect, in some sense, their citizens' autonomy. One way in which this idea is construed, for instance, holds that special justification is required for interfering with people's actions. In this sense, autonomy places a constraint on government conduct. As we have seen in the last chapter, the harm principle provides one such justification: government can legitimately interfere with my actions in order to prevent harm to third parties. Nevertheless, respecting autonomy prohibits certain justifications for governmental interference. For instance, the anti-paternalism argument holds that governments cannot coerce a person into behaving in certain ways for her own good: attempts to ban food with high contents of fat or sugar in order to encourage a healthier life style will be met with suspicion by most liberals. The argument against such restrictions is that autonomous agency requires that agents make these decisions for themselves<sup>2</sup>.

---

<sup>1</sup> Some of the material of this chapter appears in Andrés Moles, 'Autonomy, Free Speech, and Automatic Behaviour', *Res Publica* 13 (2007), 53-75.

<sup>2</sup> The anti-paternalistic principle is not absolute. There is some room for paternalistic policies but they require strong justification, in particular, they must be compatible with autonomy. I will discuss some aspects of paternalism in Chapter 3.

Some liberals extend this argument, claiming that governments have a duty not only to protect autonomy but also to promote it. Institutions must be set up in a manner that fosters citizens' autonomy. On this view, governments have a positive duty to facilitate spaces in which autonomy-enhancing practices are executed. Liberal perfectionists, for instance claim that the government can endorse autonomy-based programs justifying its decisions on the value of autonomy<sup>3</sup>. Anti-perfectionist liberals, despite being more ambiguous, might claim that promoting citizens' autonomy might be a legitimate goal under certain conditions and within certain limits<sup>4</sup>.

Both interpretations of the relation between autonomy and government have been employed to defend freedom of expression and association. As we have seen, the autonomy defence of speech cites positive reasons to defend it beyond the harm principle. One of the problems with such defences though is that the conception of autonomy invoked is seldom defined, and, moreover, it is sometimes employed as if it were an uncontested notion<sup>5</sup>. In order to assess the success of such defences it is necessary to get a clear understanding of the relation between autonomy and free speech. In this chapter I shall present a series of challenges that a conception of autonomy must answer in order to provide a sufficient basis for free speech.

The chapter is divided into three main sections. First, I will present a 'standard', procedural model of autonomy that is commonly used in political theory. In brief, the model holds that in order to be autonomous our mental states must be

---

<sup>3</sup> This does not mean, of course, that they are committed to the view that autonomy always trumps other values.

<sup>4</sup> In Chapter 4 I defend a version of anti-perfectionism which is, nevertheless, autonomy-based.

<sup>5</sup> Cf. Brison, 'The Autonomy Defense of Free Speech'. In this article she examines six notions of autonomy used by different defences of free speech. According to her, all six of them fail.

obtained through the right processes and have the right structure. Autonomy, here, is an ideal of personhood which can be achieved primarily through independence and rational reflection. In the second section I develop several challenges that might thwart the ideal. Some of the challenges are normative and are familiar in discussions in moral and political philosophy. Other challenges emanate from research in experimental social and cognitive psychology. As we discussed in the last chapter, psychologists have found that a large amount of human behaviour occurs automatically or non-consciously. Unfortunately, these findings seem to have been neglected by current studies in moral and political philosophy<sup>6</sup>. In the light of this automatic behaviour, autonomy seems to be impossible, or at least considerably rarer than we may suppose. How can we reconcile the idea that a large amount of our action occurs automatically with the claim that we must govern ourselves? In the third section, I will relate the challenges presented in the second with the conditions described in the first. This chapter will leave us better equipped to assess the autonomy defence of free speech; a task which will be undertaken in the next chapters.

I will defend a model of autonomy that is ecological in nature. This means that it abandons the requirements of classical rationality. This account is sensitive to the structure and history of our mental states, and pays particular attention to the processes of critical reflection, to the development of deliberative capacities, and to the way in which environments influence those processes and capacities. Moreover, it acknowledges that *meaningful* available options are necessary and requires mental

---

<sup>6</sup> An important exception is Hurley, 'Bypassing Conscious Control: Unconscious Imitation, Media Violence, and Freedom of Speech'.

contamination to be neutralised in sensitive areas. Although it does not embrace substantive constraints, it need not to be content-neutral.

## I

### Conditions for autonomy

In this section I reconstruct a model of autonomy which is roughly assumed in discussions regarding freedoms of speech and association, and more generally in political philosophy. Rather than offering an exegetical work on certain theories of autonomy I draw ideas from different authors. The goal is not to support any particular account, but, to a certain extent, to provide an overview of a familiar set of assumptions about the content of autonomy. The strategy here is not very dissimilar to Gilbert Ryle's picture of the 'official doctrine' of the mind or to John Tooby and Leda Cosmides' portrait of the 'standard social science model'<sup>7</sup>.

The preoccupation with autonomy derives, as Isaiah Berlin famously states, from the wish on the part of the individual to be his own master. I wish my life and decisions to depend on myself, not on external forces of whatever kind. I wish to be the instrument of my own, not of other men's acts of will. I wish to be subject, not an object; to be moved by reasons, by conscious purposes, which are my own, not by causes that affect me, as it were, from outside. I wish

---

<sup>7</sup> Gilbert Ryle, *The Concept of Mind* (London: Penguin, 1963); Tooby and Cosmides, 'The Psychological Foundations of Culture'. Although the strategy I follow here is similar to theirs, unlike Ryle I do not claim that there is an 'official version of autonomy'.

to be somebody, not nobody; a doer [...] [capable] of playing a human role, that is of conceiving goals and policies of my own and realising them<sup>8</sup>.

There are some background mental capacities and abilities that individuals require in order to reach the *capacity* for autonomy<sup>9</sup>. Broadly conceived, these include the ability to make rational choices, the ‘ability to comprehend the means required to realise [...] goals, the mental faculties necessary to plan actions, etc’<sup>10</sup>. Notice also that autonomy is not an all-or-nothing concept; people might have a greater or lesser degree of it and they can exercise their capacities to a greater or lesser extent. There is some debate, into which I will not digress, regarding the scope of autonomy. What are the units to which autonomy applies? Does it encompass whole lives, or periods of lives? Is a person who leads a non-autonomous life in his youth, but who later, in his adulthood, makes autonomous choices autonomous *tout court*? Do we need to make these sorts of judgements, or can we apply autonomy to parts of lives? Can a person be autonomous with respect to one aspect of his life (for instance, his professional career) but fail tragically with respect to autonomy in another (say, his intimate relationships)? Do we need to balance these spheres in order to reach a conclusion, or are we to say that he is autonomous only in some aspects and not in others? I will not try to address these questions here. Instead I will explore some of the conditions necessary for autonomy.

---

<sup>8</sup> Isaiah Berlin, ‘Two Concepts of Liberty’ in his *Liberty* (New York: Oxford University Press, 2002), 178. In this paragraph, Berlin is developing the idea of ‘positive freedom’ and not autonomy. Although these notions are not the same they are related, and Berlin’s words express also the idea of autonomy.

<sup>9</sup> Cf. Joel Feinberg, *The Moral Limits of the Criminal Law: Harm to Self* (Oxford: Oxford University Press, 1986), 28-31.

<sup>10</sup> Raz, *The Morality of Freedom*, 373.

### 1) Identification

The condition of identification requires that an individual's actions, desires, and higher-order mental states have the 'right' relation. Minimally they must satisfy some conditions of consistency: they must be neither contradictory nor intransitive. Similarly, the relationship between convictions and actions must be of the right kind; the convictions must explain actions qua reasons for actions, and they must not lead to overt irrationality, such as persistent weakness of the will or similar phenomena<sup>11</sup>. These minimal conditions constitute what Jon Elster calls the 'thin theory of rationality'<sup>12</sup>. Another aspect of the thin theory claims that plans must also be consistent; they must be logically coherent and they must have the correct means/goals structure<sup>13</sup> in the sense that 'there should be a possible world in which it [the plan] is realised'<sup>14</sup>. Moreover, consistent plans must harbour the property of being deliberately realisable, and not only by chance or essentially as a by-product<sup>15</sup>.

It is also supposed that first-order and second-order desires must be consistent<sup>16</sup>. In order to qualify as an autonomous agent it is not sufficient that I have certain desires; I must wish that I have those desires<sup>17</sup>. Think about Tom, a hard

---

<sup>11</sup> For a classical statement of the relations between reasons and actions see Donald Davidson, *Essays on Actions and Events* (Oxford: Oxford University Press, 2001) See especially Essay 1.

<sup>12</sup> Jon Elster, *Sour Grapes: Studies in the Subversion of Rationality* (Cambridge: Cambridge University Press, 1983). Chapter 1.

<sup>13</sup> Consistency in this sense does not require optimality or maximization. For a criticism of the idea that rationality requires optimality see *ibid.* 12-4; and Gigerenzer and Todd, *Simple Heuristics that Make Us Smart*.

<sup>14</sup> Elster, *Sour Grapes*. 11.

<sup>15</sup> *Ibid.* 11.

<sup>16</sup> Cf. John Christman, 'Autonomy in Moral and Political Philosophy', in Zalta E. (ed.), *The Stanford Encyclopedia of Philosophy*, (2003), <http://www.plato.stanford.edu/archives/fall2003/entries/autonomy-moral/> [accessed 1<sup>st</sup> December 2004]; Sarah Buss, 'Personal Autonomy', in Zalta, E. (ed.) *The Stanford Encyclopedia of Philosophy*, (2002), <http://www.plato.stanford.edu/archives/win2002/entries/personal-autonomy/> [accessed 5<sup>th</sup> January 2004].

<sup>17</sup> A classic exposition of this idea is elaborated in Harry Frankfurt, 'Freedom of the Will and the Concept of a Person', *The Journal of Philosophy* 68 (1971), 5-20. Notice, though, that in this article Frankfurt discusses the idea of free will, not the concept of autonomy.



working student who spends most of his Sundays reading rather than watching football at the local pub. Every Sunday morning Tom knows that he can either stay at home reading or go to the pub with his friends. Although he would like to watch the matches, usually Tom decides to stay at home because he wants to get ready for next week's seminars. Moreover, Tom also thinks that he would like to be a hard working student who reads on Sundays instead of watching TV. Although identification is too weak to be sufficient for autonomy, it, nevertheless, seems to be a necessary condition (insofar as one is moved to act by a desire that she also wishes to hold).

It is also important to note that this condition is completely internalist: autonomy consists solely in relations between mental states. One of the corollaries of this condition is that it does not render the origin and causal history of our mental states relevant for autonomy. For it says nothing about how such mental states came about. Consider again Tom. This model of autonomy claims that he is autonomous regardless of the causal processes which caused his mental states. Imagine that Tom is hard working as a result of his mother telling him that the only way in which he can earn the respect of others is by studying in every single moment of his spare time. After many years of living with his mother, Tom is convinced that the source of his worth as a human being depends on being well prepared for his seminars. So, he studies on Sundays and he is satisfied with the kind of person he is. Under the model of autonomy I have been describing Tom's obsessiveness is irrelevant to the question of whether or not he qualifies as an autonomous being. It is this sense in which this model is completely internalist: autonomy is only comprised of relations

between mental states. The next condition of autonomy is aimed at addressing this shortcoming.

A second problem of this condition is concerned with the meaning of 'identification'. In particular it is not clear what it requires. Two interpretations are possible. On the one hand, a weak version claims that in order to identify with my higher order mental states I merely need to *acknowledge* that fact that I have them<sup>18</sup>. However, this requirement does not seem too promising. What does my acknowledgment of having a mental state add to my having it? Tom could acknowledge that he is an obsessive student, he could even accept that he has got a problem, and, according to this interpretation, he would still satisfy this condition. On the other hand, under a more stringent interpretation, the identification clause would require *approval* of my mental states. But approval seems to be too strong a condition for autonomy, 'in that I will not identify with many of my own imperfections [...] in the sense I do not approve of them all told'<sup>19</sup>. The fact that I am not perfect according to my own standards cannot be enough to make me non-autonomous. Perhaps I regret that I like cheesy pop-music, and would rather that my musical tastes were more refined. Nevertheless, it does not seem true that I do not identify with myself, and therefore am not autonomous. This strong version needs to be qualified. It seems more reasonable to suggest that so long as I do not deeply regret having some trait, or I do not feel dismay at certain aspects, the identification condition is satisfied<sup>20</sup>.

---

<sup>18</sup> Christman, 'Autonomy in Moral and Political Philosophy'.

<sup>19</sup> John Christman, 'Procedural Autonomy and Liberal Legitimacy', in Taylor, J. (ed.) *Personal Autonomy: New Essays on Personal Autonomy and Its Roles in Contemporary Moral Philosophy*. (Cambridge: Cambridge University Press, 2005), 279-80.

<sup>20</sup> Christman replaces the condition of 'identification' with a 'non-alienation' condition. Cf. Ibid. 279.

## 2) *Historical Amendments*

Some authors have tried to solve the problems of the first condition by proposing an amendment: in order to be autonomous an agent's mental states must (also) have been derived in the 'right' way<sup>21</sup>. Accordingly, John Christman writes: 'the key element of autonomy is [...] the agent's acceptance or rejection of the *processes* of desire formation or the factors that give rise to that formation'<sup>22</sup>.

Despite the difficulties of defining which are the right ways that mental states need to be formed, there are some intuitive ideas that help us to make sense of this condition. The most important aspect here is the concern *only* with the causal history of those mental states, not their content. Consider Elster's distinction between autonomy<sup>23</sup> and ethical goodness (in regard to desires), and between judgment and truth (regarding beliefs)<sup>24</sup>. He argues that in order to construct a substantive view of rationality and autonomy we must be concerned with the causal processes by which our mental states are brought about. Rightly formed beliefs are to be contrasted with the available salient evidence rather than with the truth. In this sense autonomously held beliefs are independent of the truth; they are dependent only on how they were formed. What this shows is that one can autonomously and rationally hold false

---

<sup>21</sup> Although in a strict sense the only feature described thus far that is completely internal is the first one, the other features are similarly internal insofar as they are concerned only with internal mental states: their structure, the causal processes by which they are shaped, and the way in which they must be revised, adopted or rejected. So it seems that there are two different ways of understanding internalism here: the first (stronger) one implies that only references to internal states, structures and processes matter (as in the condition of identification); the second (weaker) sense is internalist in that it does not make references to actual present conditions of the world, but only to those that have affected internal mental processes. Condition 5 in particular focuses on the external conditions for autonomy.

<sup>22</sup> John Christman, 'Autonomy and Personal History', *Canadian Journal of Philosophy* 21 (1991), 2. Emphasis added.

<sup>23</sup> Because his understanding of autonomy comprises only desires shaped in the right way, his theory covers much less than the version pictured here. Cf. Elster, *Sour Grapes*, 2, 20.

<sup>24</sup> *Ibid.* 20.

beliefs provided that the evidence supports them in the right way. For example I can falsely, but autonomously, believe that the train to Edinburgh departs at noon because the current schedule says so, when in fact the departure time has been suddenly changed to 12:30. The fact that my belief is false does not affect the way we assess the processes by which it came about. Similarly, Elster argues that autonomously held preferences and desires are independent of their ethical content<sup>25</sup>. A person can have non-moral preferences and desires that have been shaped by the correct causal processes, and presumably even immoral preferences<sup>26</sup>. The task seems to be, then, to search for ‘drives’ which distort the autonomous processes of preference and desire formation. Following Elster’s definition, ‘drives must be conceived as non-conscious psychic forces that are geared to the search for short-term pleasure, as opposed to the conscious desires that may forego short-term pleasure to achieve some longer-term gain’<sup>27</sup>. Due to the difficulties of determining which processes are causally correct, it seems that autonomy is what remains once we remove drives that misshape preferences, such as adaptation and counteradaptation. These phenomena will be discussed at length in the second section of the chapter.

### *3) Critical Reflection*

This condition links the two already reviewed. It gives normative force to the processes by which desire and belief are formed, and also assesses how identification must take place. For it says that the critical rational reflection must play *a causal*

---

<sup>25</sup> Ibid. 23.

<sup>26</sup> This distinction is further articulated by the condition of content neutrality that I will discuss below.

<sup>27</sup> Elster, *Sour Grapes*, 25

*role* in the manner our mental states are formed and structured. According to this condition, in order to be autonomous one needs to be able to conceive of a plan for her life and to formulate a conception of the good. Moreover, it is not sufficient that autonomous agents hold such a conception; they must also have the capacity to critically revise their convictions in the light of pertinent information<sup>28</sup>.

There seem to be three different ways of understanding this condition. First, one could argue that critical endorsement of a life-plan needs to be done just once. After careful deliberation, being 20 years old, Ernesto decided that a particular conception of the good (say, living as a hippie community in Mid-Wales) is best for him. He endorses the values, beliefs and goals that such a conception of the good demands. Forty years later, Ernesto still behaves, endorses and approves of his way of life. However, he has not subjected his beliefs to critical scrutiny since he was 20 years old. Imagine that in those 40 years things have changed enough that if he thought about, say, the meaning of his being a hippie nowadays, he would not endorse those beliefs and values anymore. But because he has not thought about them, he still believes that his hippie commune in Mid Wales is the best available way of life. Although he satisfies many of the conditions so far stated it seems that his critical endorsement is somewhat defective. Autonomy cannot be guaranteed by one comprehensive act of critical reflection, it needs to be revised and updated as circumstances change.

Accordingly, under a second interpretation, critical deliberation is understood as a continuous exercise. Strictly speaking this requirement is impossible. It implies that an agent needs to critically deliberate about her conception of the good at every

---

<sup>28</sup> What counts as 'pertinent' information is an open question.

moment in her life. Every single decision, even the most trivial action needs to be contrasted and critically assessed in order to be autonomous. Autonomy so understood leads to paralysis.

Richard Arneson<sup>29</sup> suggests that there is a third, more plausible way of understanding the condition of critical deliberation. According to him, it should be understood 'dispositionally'.

To live an autonomous life an agent must decide a plan of life through critical reflection and in the process of carrying it out, remain disposed to subject the plan of critical review if disturbing or unanticipated evidence indicates the needs for such review<sup>30</sup>.

Arneson's view, however, also has problems. To the point, the agent himself needs to set the level at which revision is needed, and that level needs to be set autonomously. Autonomy, then, can only be achieved by an agent who autonomously knows when and what will trigger the need for a critical review. Although the person might have the disposition to change his beliefs if 'disturbing or unanticipated evidence' suggest he should do so, he still needs to know what would qualify as 'disturbing' and 'unanticipated'. The problem, in essence, seems to be that even the dispositional understanding 'require[s] perfect rationality and unlimited capacities'<sup>31</sup>.

---

<sup>29</sup> Arneson, Richard 'Autonomy and Preference Formation', in Coleman, J., and A. Buchanan, (eds.) *In Harm's Way: Essays in Honour of Joel Feinberg* (New York: Cambridge University Press, 1994).

<sup>30</sup> Ibid. 94.

<sup>31</sup> Ibid 48. For a criticism of these assumptions see Gerd Gigerenzer and Peter Todd 'Fast and Frugal Heuristics', in Gigerenzer and Todd, *Simple Heuristics that Make Us Smart*. See also section II below.

In order to make autonomy possible, critical reflection must be ‘effective’ in an important sense. It must make a difference to what agents do: agents must be in a situation in which they can respond to the reasons derived from their critical reflection and scrutiny. Otherwise, reflection is wasted.

#### 4) *Content Neutrality*

An important feature of autonomy requires that it be content neutral. The reason is that autonomy ‘should not contain conditions that allow the imposition of a particular value conception [...] into the lives of otherwise competent, authentic adults’<sup>32</sup>. According to authors who support neutral accounts of autonomy, appealing to neutrality makes autonomy, in some sense, valuable<sup>33</sup>. Imagine a man who accepts his desires and goals, has not been subject to any wrong processes of belief or desire formation, and is free from ‘illegitimate external influences’<sup>34</sup>. His life plan, however seems to be a waste of time: his only goal is to watch reality shows on TV. He might be a ‘coach potato’, but an autonomous one nevertheless. Defenders of neutrality argue that even if we disagree with his ‘freely’ chosen life style we cannot conclude that he lacks autonomy. Moreover, the only means to protect a plurality of ways of life and respect for agents is by adhering to neutrality<sup>35</sup>.

---

<sup>32</sup> Christman, ‘Autonomy in Moral and Political Philosophy’.

<sup>33</sup> This is by no means true of all writers. For instance, Raz’s account of autonomy is compatible with perfectionism. He argues that the value of autonomy depends ‘many morally acceptable options be available to the person’. Raz, *The Morality of Freedom*, 378 ff.

<sup>34</sup> Catriona Mackenzie and Natalie Stoljar, ‘Introduction: Autonomy Reconfigured’, in Mackenzie, C., and N. Stoljar (eds.) *Relational Autonomy* (New York: Oxford University Press, 2000), 16. Unlike John Christman, they are skeptical about this assumption. For arguments rejecting the idea that autonomy must be content-neutral see Natalie Stoljar, ‘Autonomy and the Feminist Intuition’, in Mackenzie and Stoljar (eds.) *Relational Autonomy*, 94-111.

<sup>35</sup> The problem of political neutrality will be discussed in Chapter 5.

This condition is controversial, and reaching a conclusion regarding its appeal is beyond the remit of this chapter. It is important to note, however, how it is embodied in many defences of free speech. Normally only content-neutral restrictions on speech are deemed acceptable, such as regulations related to the time and manner of speech. However there is a very strong presumption against regulations aimed at subject matter and/or viewpoint. The neutrality condition of autonomy is the basis for this presumption: if autonomy were not content-neutral, then autonomy-based justifications could not be either. For if autonomy were consistent with the ‘imposition’ of some substantive values, then a coherent defence of free speech would need to be consistent with those impositions<sup>36</sup>.

### *5) Adequacy of Choice*

A fifth point to be discussed relates to the external conditions of autonomy. According to Joseph Raz, autonomy requires a varied, adequate range of significant options. In order to be autonomous an agent ‘must not only be given a choice but he must be given an adequate range of choices’<sup>37</sup>. Consider Raz’s example of the man in the pit: A man who falls into a pit and remains there for the rest of his life. In the pit there is food enough to feed him and shelter to protect him on cold nights. Can he be a master of himself? According to Raz he cannot. The range of available options is trivial and his choices are somewhat meaningless: ‘His options are all short-term and negligible in their significance and effects’<sup>38</sup>.

---

<sup>36</sup> The presumption against content regulations and its relation with autonomy is the subject of Chapter 4.

<sup>37</sup> Raz, *The Morality of Freedom*, 373.

<sup>38</sup> *Ibid.* 374



However, we should not think that trivial choices *per se* are irrelevant to autonomy. Consider the ‘hounded woman’ case who shares an island with a fierce carnivorous animal that perpetually hunts her<sup>39</sup>. Someone whose options are all life-or-death choices cannot achieve autonomy either. Living an autonomous life requires the ability to form a life-plan, which, in turn, requires both significant and trivial options. The extremes illustrated by the hounded woman and the man in the pit show two ways in which the absence or ‘tragedy’ of choice can thwart autonomy.

How does this condition qualify the value and meaning of options? It seems that they are specified in an objective agent-independent sense. These examples suggest that living in a pit or being hounded are circumstances incompatible with autonomy independently of whether the person in the situation believes this to be the case. If this is so, then a conflict arises with the condition of neutrality: while the latter claims that conditions for autonomy must not make *any* substantive claims, the former argues that there are situations so dreary as to make autonomy impossible. A possible means of reconciling these views would be to specify a substantive threshold below which autonomy is unachievable while withdrawing further substantive judgment above that line. How acceptable this position is for a defender of content neutrality is an issue to be seen.

### 6) *Classical Rationality*

Some versions of autonomy seem to make certain assumptions about the nature and character of human rationality, often resembling those of rational choice and

---

<sup>39</sup> Ibid. 374.

expected utility theories<sup>40</sup>. According to these approaches, human rationality can be best conceptualised as a decision making device connecting goals and means, where the former are taken as given. This view of rationality, so described, is consistent with the condition of content-neutrality. Rationality demands that an agent finds the best means to achieve whatever ends he desires: he must ponder all the available options and compare how they ‘score’. The rational action is the most efficient way to achieve a given end. Explaining a rational action ‘[w]ould’, according to Elster, ‘show that the action is the (unique) best way of satisfying the full set of agent’s desires, given the (uniquely) best beliefs the agent could form, relatively to the (uniquely determined) optimal amount of evidence’<sup>41</sup>. Or, in the case of decisions under uncertainty, a rational decision between options A and B an agent must

compute his personal expected utility for [A], he [has] to determine all the possible consequences that [A] [can] bring, attach quantitative probabilities to each of these consequences, estimate the subjective utility of each consequence, multiply each utility by its associated probability and finally add all these numbers up. The same procedure [has] to [be] repeated for the alternative [B]. Finally, he [has] to choose the alternative with the higher total expected utility<sup>42</sup>.

For instance, recall Richard Arneson’s remarks on critical reflection. He suggests that an autonomous agent must be predisposed to revise his plans and commitments when ‘disturbing or unanticipated evidence’<sup>43</sup> advise the need to do so. Similarly, the

---

<sup>40</sup> Hurley, ‘The Public Ecology of Responsibility’.

<sup>41</sup> Jon Elster, ‘The Nature and Scope of Rational Choice Explanation’, in LePore, E., and B. McLaughlin (eds.) *Actions and Events: Perspectives on the Philosophy of Donald Davidson* (Oxford: Blackwell, 1985), 71.

<sup>42</sup> Gigerenzer and Todd, ‘Fast and Frugal Heuristics’, 9.

<sup>43</sup> Arneson, ‘Autonomy and Preference Formation’, 49.

processes that lead to one's formation of plans and goals must involve 'full information and perfect rationality'<sup>44</sup>.

This conception of rationality might be segregated into two different variants. One, purely behavioural, argues that agents are expected to behave *as if* they were maximising expected utility. This view does not make any claims about the processes by which actions are caused. On the other hand, a different approach suggests that rationality encompasses the processes by which actions are brought about. Here, the processes of decision-making must themselves be rational in some sense. In general, these processes must satisfy some criteria of reasoning, such as being consistent with the rules of logical deduction. This criterion implies that rationality is thought to be general in its application and scope.

Lastly, this view of rationality seems to suggest that mental states such as perceptions, beliefs, desires and intentions are connected through rational processes. These mental states are mediated by processes of reasoning in a familiar way: my perception of red, my forming a belief about the object being red, my desire to reach that red object and my intention to move my arm towards it are all mediated by 'acts of will', which must be, in some sense, rational.

### *Conditions for Autonomy: A Summary*

Before I present some challenges that might thwart the possibility of autonomy, I will sum up its associated conditions. The first condition states that autonomy requires mental states to be coherent in a usual way: they must be transitive and consistent. Moreover, there must be some identification between higher- and lower-

---

<sup>44</sup> Ibid. 48.

order states, such that if a person has a desire X, she must also accept or wish she had that desire. Second, mental states must have also the proper history; they must be the outcome of the 'right' processes of desire and belief formation. In contrast with the first condition, which focused on the content of mental states, this condition focuses on the processes by which mental states come to be what they are. The third condition links the previous two: it holds that rational processes must have a causal role in the formation of mental states and that their content must be subjected to critical scrutiny whenever circumstances require this. This condition is here understood as a disposition towards critical reflection. The fourth condition holds that autonomy is to be gauged in a content-neutral fashion; this is motivated by the idea that autonomy should be independent of ethical considerations regarding the worthiness of controversial views of the good life. The fifth clause holds that autonomy requires an available range of meaningful options. Although this condition is perhaps in some tension with content-neutrality, it seems plausible that a life with very limited choice or a life in which every choice is life threatening cannot be an autonomous one. Finally, I specified some assumptions about rationality that seemingly underlie many conceptions of autonomy. These assumptions are embodied by classical theories of rational choice and expected utility theory. Paradigmatically, they assume that rationality is a general-domain, maximising, decision-making device linking given goals with the most effective means to achieve them. Another assumption is that mental states are mediated by reasoning and acts of will. With these conditions explored we must also consider various issues which threaten the possibility of autonomy prevailing. It is to this issue I shall now turn

## II

## Challenges to Autonomy

*A) Adaptive Preference Formation*

One of the important challenges to autonomy is the tendency for an individual to adapt her preferences and desires to the set of *feasible* options and opportunities in order to avoid cognitive dissonance<sup>45</sup>. This process ‘runs behind their back’ insofar as they are not aware that their preferences are being adapted. Imagine Beto, a relatively good young football player from a town in a remote village in a Latin American country. One day, scouts from Arsenal FC offer him the chance to participate in a pre-selection process at The Emirates Stadium. If successful, Beto could sign a contract for Arsenal’s youth team for 5 years (let’s assume that playing for Arsenal is much better than playing for his local team ‘Atlético Huixquilucan’). However, Beto cannot afford the flight to London. When his friends ask him why he didn’t attend the pre-selection match, Beto sincerely tells them that, after all, he didn’t want to play for Arsenal, the weather in London is very bad, Arsenal are an awful team, and, besides, he would really miss his mother’s cooking<sup>46</sup>. According to the structural model Beto’s choice would be autonomous, for he has both the desire to keep playing for his local team, and the desire to have that desire. However we might object that something is wrong here<sup>47</sup>. The processes by which Beto has

---

<sup>45</sup> Elster, *Sour Grapes*.

<sup>46</sup> We must assume that he is telling the truth; he has become convinced about these reasons.

<sup>47</sup> In order to establish that a process of adaptive preference has taken place we need to assume other things: we need to presuppose that if Beto had had enough money to go to London, he would have gone and would not have formed the desire not to travel to London. Another possibility, even more complicated, is that even having the money he just could not picture himself playing abroad. In this case, we must presuppose that his preferences are adapted towards the meagre expectations he has got.

achieved coherence between his different mental states are awkward in the sense that his refusal to play for Arsenal was caused by his limited (in this case economic) means for satisfying a desire. So it is not the case that his higher order desires or goals determined his actions (and lower order desires) but, rather, his circumstances have determined his character *with no awareness* on Beto's part. There are many different situations in which individuals adapt their preferences and desires in this fashion. In Amartya Sen's words it is possible that '[t]he hopeless beggar, the precarious landless labourer, the dominated housewife, the hardened unemployed or the overexhausted coolie' could qualify as autonomous under the structural view. Surely, however, they would not pass the historic test<sup>48</sup>.

The opposite phenomenon to 'sour grapes' is what Elster calls 'counteradaptive preferences'. Here, preferences are also shaped by the feasible set but in the opposite direction. This phenomenon is captured by the maxim, 'the grass is greener on the other side'. Imagine Dave, a football player who, last year, was playing for an English team and dreamed about playing in the Spanish League. This year he is hired by a Spanish team, but now he dreams of returning to play in England. Similar to the scenario in 'sour grapes' his preferences are shaped 'behind his back' and not through learning<sup>49</sup>. Counterfactually if he were to play in England again he would return to his hope of playing in Spain. These phenomena are important for autonomy because individuals' desires and preferences are being shaped by drives (with no awareness) rather than by intentional desires.

However, not every adaptation threatens the ideal of autonomy. Sometimes individuals can autonomously reduce the gap between what they can achieve and

---

<sup>48</sup> Amartya Sen, *On Ethics and Economics* (Oxford: Blackwell, 1987), 45.

<sup>49</sup> Elster, *Sour Grapes*. 111.

what they wish to achieve. For instance, contrast ‘wrong’ adaptation processes with ‘character planning’. As Elster writes,

In both cases [sour grapes and character planning] the process [of adaptation] begins with a state of tension between what you can and what you might like to do. If the escape from this tension takes place by some causal mechanism of dissonance reduction, we are dealing with sour grapes; if it is engineered by conscious ‘strategies of liberation’, with character planning<sup>50</sup>.

The crucial difference is that in the former the process of adaptation is purely causal (drive shaped), while in the latter it is also intentional (produced through meta-preferences)<sup>51</sup>.

### *B) Abnormal Belief Formation*

Beliefs can also be misshaped by a process of ‘wishful thinking’. In this case, beliefs are changed in order to conform with desires and feasible options. In similarity to adaptive preferences, the ‘function’ of wishful thinking is to reduce cognitive dissonance. In its pure form, wishful thinking operates under the maxim ‘believe what you would like to be the case’. I believe that the town in which I live is the best town to reside in, or the University for which I work is the best in the world. Those can be cases of wishful thinking, provided that the belief in question has been caused

---

<sup>50</sup> Ibid. 117.

<sup>51</sup> Sometimes character planning might also conflict with autonomy; one person might become non autonomous by over-planning the way he wants to be. A strong desire to be a certain kind of person might conflict with critical reflection. Think for example in Tolstoy’s picture of Prince Steven Arkadievich Oblonsky: ‘Oblonsky’s tendency and opinions were not his by deliberate choice: they came of themselves, just as he did not choose the fashion of his hats or coats but wore those of the current style. Living in a certain social set, and having a desire, such as generally develops with maturity, for some kind of mental activity, he was obliged to hold views, just as he was obliged to have a hat. If he had a reason for preferring Liberalism to Conservatism of many in his set, it was not that he considered Liberalism more reasonable, but because it suited his manner better’. Leo Tolstoy, *Anna Karenina* (Oxford: Oxford University Press, 1995), 6.

(merely) by the desire that the world is thus rather than by salient evidence or good reasons. Wishful thinking reduces autonomy insofar as beliefs are formed subconsciously or by the wrong processes.

It seems also possible to find a process of abnormal belief formation similar to the sour grapes mechanism<sup>52</sup>. According to this phenomenon, the processes by which an individual conceives her life-plans or conception of the good might be threatened by her deprived circumstances. In this case the mechanism operates at the level of convictions rather than at that of desires. Someone can fail to appreciate salient evidence because the circumstances in which he lives make it difficult to form certain rational beliefs. Imagine someone who due to living in an extremely violent society<sup>53</sup>, believes that violence is 'natural' (in the sense that there is nothing wrong with it). His failure to recognise the wrongness of violence might be explained by the fact that the 'normal' processes of belief formation have been distorted by the violent environment in which he lives<sup>54</sup>.

### *C) Oppressive Socialisation*

The problem of 'oppressive socialisation' arises because a theory of autonomy needs to avoid individuals repeating and reproducing certain conditions into which they have been socialised. Social environments and a community's shared values (be it the family, nation, clan, social class) can be causally effective in conditioning an individuals' conception of the good. Somehow, the upbringing generates the standpoint from which critical reflection and the assessment of one's opportunities

---

<sup>52</sup> Indeed Elster notes that some versions of the Sour Grapes fable involve misshapen beliefs. Elster, *Sour Grapes*, 123 n. 45.

<sup>53</sup> Something like the society pictured in the movie *City of God*.

<sup>54</sup> Cf. Gilbert 'How Mental Systems Believe'. especially 111-2.



and choices are carried out. Values and practices inculcated during childhood have a great impact in the development of autonomy: some of them facilitate its developments, whereas others hinder it. Through oppressive socialisation people might find themselves merely reproducing the conditions under which they were socialised, and, furthermore, they might also identify themselves with such practices. There are dramatic examples of oppressive socialisation in which manipulation and brainwashing are common; women educated to be dominated by males, people educated into making huge sacrifices for a political party or religious leader, and so on<sup>55</sup>.

Besides those well known examples there are less dramatic cases in which people are introduced to practices and comprehensive views non-autonomously<sup>56</sup>. Think about Alice, a competent woman who enters into an unequal marriage. Due to her traditional and conservative Catholic upbringing, she does not feel uncomfortable living a life which consists mainly in looking after children, attending social meetings with her husband, and gossiping with friends. Although her upbringing was relatively traditional, we cannot say that she was brainwashed or manipulated and she certainly was not coerced by anybody to enter into this marriage. Moreover, at the time she married she also had the options of attending a university or pursuing a career in ballet dancing. In other words, she did have a variety of meaningful options. Finally, she now thinks that her life is fine: she identifies with the kind of life she is living and so on. There seems to be something awkward in the way Alice shaped her life; she is merely replicating the values of this

---

<sup>55</sup> Think about educational policies such those described in George Orwell's novel *1984*.

<sup>56</sup> See, for instance, Clayton, *Justice and Legitimacy in Upbringing*.

form of Catholicism<sup>57</sup>. Now compare Alice with Ana, another Catholic woman whose life is substantively similar to Alice's. She also is confined to looking after the children, attending social meetings in Church, and gossiping. However Ana was not raised in a Catholic family. Her upbringing was 'free' in the sense that she was not introduced to any particular conception of the good. But when she reached 22 years old, she converted into Catholicism. Does it make any difference? There seem to be two different ways of assessing these examples. On the one hand we could say that Alice is not autonomous due to the distorted processes through which she came to affirm the values she does. This view is consistent with the condition of content-neutrality insofar as it does not oppose the alleged oppressiveness of Catholicism; only the processes that led Alice to endorse those values. Alice's autonomy is threatened because her 'oppressive' upbringing is the standpoint from which she evaluates and judges her life. From this perspective, the problem is the way in which she has come to endorse those values; namely through an oppressive, or at least a non-autonomy-fostering socialisation. On the other hand, a second strategy can attempt to disqualify both of Alice and Ana from being autonomous on the grounds that they voluntarily entered in an unequal marriage. In order to defend this strategy one needs to argue that oppressive relations are incompatible with autonomy even if 'autonomously' chosen. On this view, autonomy is incompatible with certain oppressive values regardless of the way in which the person came to affirm them. Autonomy here is not content-neutral, but substantive<sup>58</sup>. The problem

---

<sup>57</sup> There are two different ways to read Alice's case: one holds, consistently with content-neutrality, that there is nothing autonomously-diminishing in affirming Catholic oppressive values. On the second view autonomy is incompatible with certain oppressive values regardless of the way in which the person came to affirm them.

<sup>58</sup> See Mackenzie and Stoljar 'Introduction: Autonomy Reconfigured' for a defence of this view. A critical assessment comes in John Christman, 'Relational Autonomy, Liberal Individualism, and the

with this second view is that it somewhat confuses the question of autonomy with ethical assessment of our lives. In order to rule out cases of oppressive socialisation, we must find a procedure to ensure that critical reflection does not ‘simply replicate the oppressive social conditions that autonomous living is meant to stand against’<sup>59</sup>.

One could conclude that because Alice is less autonomous than Ana, the latter should bear the costs arising from her choices. However, this claim is mistaken. Andrew Mason rightly claims that women living in a sexist society should not bear the full costs of their actions (even if, like Ana’s, their decisions are autonomous)<sup>60</sup>. Mason argues that insofar as some choices are made against a background in which sexist norms are ‘widely accepted and transmitted in a variety of different ways through a complex system of gender socialisation’<sup>61</sup>, it would be unfair to make women bear the full costs of her choices, ‘even if their decision is fully autonomous’<sup>62</sup>. The reason is not merely that preferences and desires are shaped by socialisation, but that some of the norms into which a person is socialised are unjust. The implication of this point, however, extends beyond the aims of this chapter. Here the concern is the *concept* of autonomy, rather than in which circumstances autonomy is relevant for the distribution of burdens and benefits in society.

#### *D) Automatic and Unconscious Processes*

---

Social Constitution of Selves’, *Philosophical Studies*, 117 (2004), 143-64; Christman, ‘Procedural Autonomy and Liberal Legitimacy’.

<sup>59</sup> Christman, ‘Relational Autonomy, Liberal Individualism, and the Social Constitution of Selves’, 154.

<sup>60</sup> Andrew Mason. *Levelling the Playing Field: The Idea of Equal Opportunity and its Place in Egalitarian Thought* (Oxford: Oxford University Press, 2006), 182-8.

<sup>61</sup> Ibid. 185.

<sup>62</sup> Ibid. 185.

As discussed in the previous chapter, it has been shown by several experiments in Social and Cognitive Psychology that much behaviour is automatic and non-conscious. Distinct disciplines have converged on the conclusion that the automaticity of behaviour constitutes a salient ingredient of the way human beings function<sup>63</sup>. As Selten writes: 'Much of human behaviour is automatised in the sense that it is not connected to any conscious deliberation'<sup>64</sup>.

Automatic processes have been defined as mental acts which have one or more of the following elements: they are unconscious, unintentional or effortless<sup>65</sup>. 'An automatic mental phenomenon occurs reflexively whenever certain triggering conditions are in place; when those conditions are present, the process runs autonomously, independently of conscious guidance'<sup>66</sup>. Some of the conditions necessary for automatic processes

require only the presence of the triggering environmental event; it does not matter where the current focus of conscious attention is, what the individual was recently thinking, or what the individual's current intentions or goals are. [...]

What it means for a psychological process to be automatic, therefore is that it happens when its set of preconditions are in place without needing any conscious choice to occur, or guidance from that point on<sup>67</sup>.

---

<sup>63</sup> See the collection of essays in Ran Hassin, James Uleman and John Bargh, *The New Unconscious* (Oxford: Oxford University Press, 2005); Bargh and Chartrand 'The Unbearable Automaticity of Being'; Bargh, Chen, and Burrows, 'Automaticity of Social Behavior'; Chartrand and Bargh, 'The Chameleon Effect'; Timothy Wilson, *Strangers to Ourselves: Discovering the Adaptive Unconscious* (New York: Belknap, 2002).

<sup>64</sup> Reinhard Selten, 'What is Bounded Rationality', in Gigerenzer, G., and R. Selten (eds.) *Bounded Rationality: The Adaptive Toolbox* (Cambridge, MA: MIT Press, 2001), 16.

<sup>65</sup> Wilson, *Strangers to Ourselves*, 222 n.4. Cf. Bargh and Chartrand define a conscious process as those 'mental acts of which we are aware, that we intend (i.e. that we start by an act of will), that require effort, and that we can control (i.e. we can stop them and go on to something else we choose'. Bargh and Chartrand 'The Unbearable Automaticity of Being', 463.

<sup>66</sup> John Bargh, 'The Automaticity of Everyday Life', *Advances in Social Cognition*, 10 (1997), 3.

<sup>67</sup> *Ibid.* 3.

There seem to be at least two major forms of automatic mental processes: skill acquisition and unintentional automatic response. Skill acquisition is automatic in the sense that, once some degree of mastery has been achieved, the performance of some activity becomes mechanical. Think, for example, of the process of learning a foreign language or driving a car. This kind of behaviour begins with an act of will. Then, as the processes become more efficient over time, agents are able to perform the behaviour without conscious guidance: '[t]hese [are] intentional but effortless mental processes'<sup>68</sup>.

In contrast there is another kind of automatic processing that in addition does not require an act of will to be prompted. It is this kind of automaticity which occupies me in this thesis. Among the mechanisms that operate at an automatic level, one, which has received a great amount of attention in academic literature, is the so-called 'ideomotor principle'<sup>69</sup>. The idea behind this is that the mere perception of some behaviour increases the likelihood of an agent performing that same behaviour unintentionally<sup>70</sup>. Of import here is that 'the influence of perception on behavioural tendencies is automatic in that it is passive, unintentional and non-conscious'<sup>71</sup>. This principle is not restricted to perceptual experiences. The 'representation of an action's goal can elicit movements that would be means to that end; it has effects

---

<sup>68</sup> Bargh and Chartrand 'The Unbearable Automaticity of Being', 463. Skill acquisition is perfectly compatible with the standard notion of autonomy. In a sense, this kind of behaviour is kept in virtual control: individuals can make an effort to bring it back to consciousness. Philip Pettit, *A Theory of Freedom: From the Psychology to the Politics of Agency* (Cambridge: Polity, 2001)

<sup>69</sup> Bargh, 'Bypassing the Will'; John Bargh, 'The Automaticity of Everyday Life', *Advances in Social Cognition*, 10 (1997), 1-61; Dijksterhuis and Bargh, 'The Perception-Behavior Expressway'.

<sup>70</sup> Hurley, 'Bypassing Conscious Control', 310.

<sup>71</sup> Bargh, Chen, and Burrows, 'Automaticity of Social Behavior', 233.

even when movements do not break through overtly. [...] [M]erely imagining a skilled performance, in sport or music, improves your own performance'<sup>72</sup>.

The ideomotor principle can be described as having two steps. First, there is a link between environmental stimuli and perception. Then there is a second step between perception and behaviour. In contrast with the standard view of the relationship between perception and action, which supposes mental mediation between different mental states, the ideomotor principle suggests that perceptual traits activate behavioural tendencies directly. Contrast the following pictures of the relation between perception and behaviour:

Environment → Perception → Mental Representation → Act of will → Action

According to this view, behaviour always requires an element of 'volition'. The automatic behaviour perspective proposes a picture that dispenses the necessity of will. Its suggestion is this:

Environment → Perception → Mental Representation → Action (Behaviour).

In other words, 'perceiving an action activates the mental representation of this action, which in turns leads to performance of the action'<sup>73</sup>.

An important phenomenon related to the ideomotor principle is the so-called 'Chameleon Effect'<sup>74</sup>, which holds that individuals tend to imitate whatever other people are doing. In a series of experiments, Tanya Chartrand and John Bargh

---

<sup>72</sup> Hurley, 'Bypassing Conscious Control', 310.

<sup>73</sup> Dijksterhuis and Bargh, 'The Perception-Behavior Expressway'. 8.

<sup>74</sup> Chartrand and Bargh, 'The Chameleon Effect': Hurley, 'Bypassing Conscious Control'.

suggest that imitative tendencies are easily triggered. In one of the experiments two individuals were sitting side-by-side both looking at a desk with pictures. They were asked to focus all their attention on the pictures (not on the other person). The results show that individuals tended to move their foot when sitting by someone who was also moving his foot, and tended to rub their face when the other person did likewise<sup>75</sup>. Similarly, when seated by someone who was smiling the experimenters tended to smile, whereas when seated by a non-smiling individual they tended not to. Surprisingly, when the participants were asked if they felt that the other person was friendly or not, their answers were vague. In the majority of cases they said they did not pay attention to the other person. They also reported not attempting to be friendly or unfriendly themselves. It is important to note that, according to their own reports, in both experiments, participants were unaware of either, the other person's behaviour or their own<sup>76</sup>.

A related group of phenomena do not depend on perceiving behaviour, but rely on activating concepts within an agent's mind. The idea here is that environmental features can activate concepts automatically, increasing the likelihood of behaving according to those concepts. It is important to see that concept activation is also a form of the ideomotor principle. Concepts trigger associated activities, and perceiving associated traits can activate those concepts<sup>77</sup>. Not surprisingly, automatic

---

<sup>75</sup> Chartrand and Bargh, 'The Chameleon Effect', 897-900.

<sup>76</sup> Ibid. 900-4.

<sup>77</sup> It is not difficult to speculate as to why these tendencies evolved. The ideomotor principle does not require as many mental resources as a principle in which perception and action are completely independent. Imitation can be a very successful strategy for learning new unusual tricks. For instance, Hurley, 'Bypassing Conscious Control' refers to the importance of imitation in Michael Tomasello, *The Cultural Origins of Human Cognition* (Cambridge: Harvard University Press, 1999). On the other hand, stereotyping and generalising might prove more successful for structuring new information than using all new information as if it were the first time it was dealt with. See Gigerenzer and Selten, *Bounded Rationality*; Gigerenzer and Todd, *Simple Heuristics that Make Us Smart*.

and implicit elements<sup>78</sup> also figure prominently in social perception and social cognition. For instance, in a series of experiments Bargh and his colleagues primed some students with different concepts<sup>79</sup>. The students were not aware that they were being primed –they believed that they were performing a different task. In experiment 1, the students were primed with traits of rudeness or politeness by the use of a scrambled sentence test. After performing this task they were taken into a room in which a conversation was taking place. They were given the option of interrupting the conversation or waiting for it to end. The results show that individuals primed with words related to politeness tended to wait longer for the conversation to end (i.e. they behaved more politely) than people who were primed with the concept of ‘rudeness’<sup>80</sup>.

An important component of implicit social cognition are social stereotypes.

A stereotype is a socially shared set of beliefs about traits that are characteristic of members of a social category. [...] Stereotypes guide judgement and action to the extent that a person acts toward another as if the other possesses traits included in the stereotype. [...] Implicit stereotypes are introspectively unidentified (or inaccurately identified) traces of past experience that mediate attributions of qualities to members of a social category’<sup>81</sup>.

Because people are exposed to the content of social stereotypes from early childhood (according to Patricia Devine<sup>82</sup> stereotypes are established in children as young as 5 years old) it is predictable that stereotypes can be automatically triggered, as with

---

<sup>78</sup> Although different dichotomies such as ‘explicit-implicit’, ‘aware-unaware’, ‘conscious-unconscious’, ‘automatic-controlled’ are used to describe more or less the same processes, they tend to emphasise different aspects of overlapping phenomena.

<sup>79</sup> Bargh, Chen, and Burrows, ‘Automaticity of Social Behavior’.

<sup>80</sup> Ibid. 233-6.

<sup>81</sup> Greenwald and Banaji, ‘Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes’, 14-5.

<sup>82</sup> Patricia Devine, ‘Stereotypes and Prejudice: Their Automatic and Controlled Components’, *Journal of Personality and Social Psychology*, 56 (1989), 5-18.



other cognitive traits. This does not mean that people approve or endorse the content of the stereotype<sup>83</sup>. But, because stereotypes are learnt in early childhood, the cognitive structures of proposition rejection have not been fully developed. This makes it more difficult for individuals to put aside the automatic reactions to the stereotype in question. In a second experiment, John Bargh and others used the same scrambled sentence test employed in the 'rude/polite' experiment. This time the stereotype of the elderly was activated. After completing the task, students who were primed with words related to the elderly stereotype walked slower (slowness being a trait associated with the elderly people stereotype) than those not primed<sup>84</sup>.

A large number of similar experiments have been completed, and the evidence for these automatic phenomena is strong<sup>85</sup>. From them we can draw some significant conclusions. First, it is important to notice that these processes operate at many levels; they trigger goal-oriented behaviour as well as merely instrumental, bodily behaviour<sup>86</sup>. Second, the ideomotor principle and the chameleon effect operate continuously; they are the default state of human beings<sup>87</sup>. Third, it is the environment that directly causes the behaviour<sup>88</sup>. It does not matter whether individuals are aware that they are being primed; they usually do not accept that their behaviour is being caused by external stimuli rather than by their own will<sup>89</sup>. And fourth, the cost of controlling those automatic processes is high. It requires a great amount of attention and effort. Although people can control the processes by

---

<sup>83</sup> Ibid. 5.

<sup>84</sup> Bargh, Chen, and Burrows, 'Automaticity of Social Behavior', 236-8.

<sup>85</sup> Cf. Bargh, 'Bypassing the Will' for a survey of convergent literature.

<sup>86</sup> Hurley, 'Bypassing Conscious Control', 311.

<sup>87</sup> Ibid. 311.

<sup>88</sup> Bargh, 'Bypassing the Will'.

<sup>89</sup> Bargh and Chartrand 'The Unbearable Automaticity of Being'; Bargh, 'Bypassing the Will'; Hurley, 'Bypassing Conscious Control'.

focusing on it not happening, this is only possible ‘when there is sufficient attentional capacity available for the act of control’<sup>90</sup>. When students were asked to walk fast after being primed with the stereotypes they were able to, thus overcoming the priming effect. It is very easy, however, due to our limited attentional capacity, to lose control. More important perhaps is the fact that people do not accept that such processes influence them. Automatic behaviour occurs, and it happens much more than we suppose. It increases the likelihood of meaningless movements (like rubbing my face, move my foot while sitting), but it also activates goal-oriented behaviour. Moreover, automatic behaviour influences social relations: people can act more or less politely, be more or less rude, or react to certain social groups with no awareness of their doing so.

### *E) Mental Contamination*

Mental contamination is defined as the ‘process whereby a person ends up with an unwanted judgment, emotion or behavior because of mental processing that is unconscious or uncontrollable’. ‘Unwanted’ here means ‘that the judgment maker would prefer not to be influenced by the mental process in question’<sup>91</sup>. Mental contamination is related to how stimuli trigger processes in a person’s mind, usually with no awareness and necessarily without her approval. In this respect, mental contamination is linked to the condition of identification discussed above. An important feature of this definition is that it is *subjectively* defined; contamination occurs when a person responds in ways *she* would not like to.

---

<sup>90</sup> Bargh, Chen, and Burrows, ‘Automaticity of Social Behavior’, 232.

<sup>91</sup> Timothy Wilson and Nancy Brekke, ‘Mental Contamination and Mental Correction: Unwanted Influences on Judgments and Evaluations’, *Psychological Bulletin*, 116 (1994), 119.

One of the dynamics which facilitates mental contamination relates to our mind's opacity: due to our limited introspective capacities we are not very good at discovering the causes of our reactions and in many respects we are mistaken about ourselves. For instance, 'when [people] form an evaluation of someone, what they experience is usually the final product [...], not the mental processes that produced this product'<sup>92</sup>. When mental contamination occurs, it may leave some clues. However, they are usually not definitive and, moreover, they can be misleading. There are two main causes of mental contamination: on the one hand, it can be the by-product of automatic responses; on the other, it can be due to source confusion.

Recall that a significant part of our behaviour is automatic (it is effortless and unconscious). These processes might lead to contamination when they cause responses we would rather not have. Moreover, because they occur outside of one's awareness, these processes are extremely difficult to control. How can I control my mimicking someone when I am not aware that I am doing so? For this reason, automatic behaviour's contaminant potential is elevated. Similarly, it is not easy avoiding feelings or beliefs through sheer acts of will<sup>93</sup>. Even if behaviour is easier to control than mental states, as the literature on automatic behaviour suggests, it is likely that much will escape our control. We usually do not have spare mental resources to control everything we do.

A second source of mental contamination concerns source confusion.

---

<sup>92</sup> Wilson and Brekke, 'Mental Contamination and Mental Correction', 121.

<sup>93</sup> Cf. Daniel Wegner, 'You Can't Always Think What You Want: Problems in the Suppression of Unwanted Thoughts', *Advances in Experimental Social Psychology*, 25 (1992), 193-225; Daniel Gilbert, Romin Tatarodi and Patrick Malone, 'You Can't Not Believe Everything You Read', *Journal of Personality and Social Psychology*, 65 (1993), 221-33. See my discussion of this problem regarding free speech in Chapter 5.

In everyday life, people's responses are almost always multidetermined; their feelings for a loved one spring from many sources, and their evaluation of a job candidate is based on more than the candidate's many attributes. A considerable amount of research suggests, however, that people are not skilled at dissecting a judgment (recognizing the exact contribution of everything that caused it) [...]. They recognize the thoughts and feelings that result but not the exact recipes that produced them. [...] Because of this source confusion –the inability to recognize the exact contribution of all of the influences on one's judgments– it is difficult to take steps to avoid being affected by unwanted agents<sup>94</sup>.

For instance, individuals tend to make mistakes in the attribution of the cause of their emotions and/or mental states; they can believe their arousal is caused by an attractive experimenter for instance, when in fact it was caused by fear of crossing a bridge<sup>95</sup>.

An important and well-known source of confusion is the 'halo effect', which is characterised as a situation in which a subject's learning that an unfamiliar person possesses a characteristic B tends to produce a 'positive or negative attitude' (depending on whether B is judged positively or negatively) toward that person. That attitude then is likely to be generalised to some other specific attribute A of that person. 'The attitude toward B is said to operate implicitly when the subject does not notice that B is influencing the judgment of A'<sup>96</sup>. For instance, essays written by a physically attractive student tend to get better marks than essays written by students who are less attractive<sup>97</sup>. Similarly, 'attractive males and females are judged to be kinder, more interesting, more sociable, happier, stronger, of better character, and

---

<sup>94</sup> Wilson and Brekke, 'Mental Contamination and Mental Correction', 129.

<sup>95</sup> Ibid. 129. They list more examples of this kind.

<sup>96</sup> Greenwald and Banaji, 'Implicit Social Cognition', 9.

<sup>97</sup> Wilson and Brekke, 'Mental Contamination and Mental Correction', 129

more likely to hold prestigious jobs'<sup>98</sup>. In another example unrelated to physical attractiveness, a male person was presented to different groups of observers as a student, demonstrator, lecturer, senior lecturer and professor. Later, when the observers had to remember how the man looked, the description of his height was a function of his status, the student being the shorter and the professor the taller<sup>99</sup>.

How can people avoid mental contamination, and how likely are they to be successful? Wilson and Brekke suggest that four steps are required. First, people must be aware that the biasing element will occur. Imagine that a teacher is marking the essays of two students. One of them is written by a physically attractive student, the other is written by a non attractive one. The marker gives the first student a 65%, while she gives to the second student 63% (suppose that in unbiased evaluation both essays get 64%). The teacher believes that the marks are fair and reflect only the quality of the written work. Following the 'halo effect', it is likely that the teacher has been contaminated; an objective evaluation (of the quality of the essay) has been influenced by a subjective evaluation (how much she likes/dislikes the student). In order to avoid contamination, she must know that the 'halo effect' has occurred, or is likely to occur.

Second, people must be motivated to neutralise the bias. In the example above, the teacher must concede that she is actually susceptible to being biased. This step is not always taken since some people reject that they might be biased in these ways because it violates our common sense about ourselves. While the teacher reads about these phenomena, she would probably say "hold on, I believe that the 'halo effect' happens to some people, but surely not to me. I'm being fair!". In order to

---

<sup>98</sup> Greenwald and Banaji, 'Implicit Social Cognition', 9

<sup>99</sup> *ibid.* 9.

prove the extent to which people are motivated to correct their biases Wilson and others<sup>100</sup> designed the following experiment. They provided some data, published by the impartial magazine *Consumer Report* (CR), about two brands of condoms (A and B) to a group of female students. According to the *Consumer Report* brand A was of better quality than Brand B. They also produced two anonymous reports by students' of their university. The students' report (SR) recommended brand B over brand A. Participants could choose to read neither, one or both of the reports and later they had to choose which brand they would select in their private lives. Some of the students (13%) said they would want to be more influenced by the SR than by CR, but only 3% wanted a friend to be more influenced by the SR. For the rest of participants, the SR satisfied the definition of contaminant (they did not want to be influenced by it). Interestingly, when they were offered the information, most people choose to see both reports (77%), and very few avoided the SR (12%). 'This suggests that although people recognized the invalidity of the [SR], they had faith in their ability to avoid being influenced by it in an undesirable way'<sup>101</sup>. As expected, when given both kinds of information, 31% preferred brand B (the one recommended by SR), and, consistently their choice was the product of a contaminated decision. To sum up the results, 13% wanted to be influenced by SR, 77% read both reports and 31% ended up choosing the brand recommended by SR.

Acknowledgment and motivation alone, however, are not sufficient to avoid contamination. Third, then, we also need to know the direction and magnitude of the bias. In the example of the teacher, suppose that she thinks she might be influenced by the 'halo effect' and she wants to neutralise it. However, she does not know how

---

<sup>100</sup> Cf. discussion in Wilson and Brekke, 'Mental Contamination and Mental Correction', 124-5.

<sup>101</sup> Ibid. 125.

much biased she is. To be sure, she could take 5 points away from the attractive student, or give 5 extra points to the unattractive one. But then her marks would be biased again, not by the 'halo effect' but due to an overestimation of the effects of her previous bias. It seems that people tend to underestimate their own susceptibility to contamination: people seem to believe that information is more contaminating in the hand of others than in their own hands. Significantly, though, they also usually do not know how to calculate the correct adjustment necessary to counteract the halo effect.

Fourth, people need to take effective steps and have the ability to adjust their responses in a non-contaminated way. One of the problems besetting this is that people tend to overestimate their capacity for mental control. Otherwise, why did participants in the experiment above choose to see both reports, even when they did not want to be influenced by one of them? Another experiment's results are similar. Participants were to hire someone. Even though many of them did not want to be influenced by the applicant's gender and they agreed that knowing the applicant's gender might influence them more than they would wish it to, 87% still chose to know the gender of the person (only 5% declined the knowledge).

One possible solution to this problem is to forewarn people about the contamination potential of certain sources<sup>102</sup>. This strategy's success is limited though: critically it depends on people's attitudes towards themselves. If I believe that I can exert control over my mind (as we normally *feel* we can), then being forewarned about the dangers of some influences will make no significant

---

<sup>102</sup> This strategy will be discussed with detail in Chapter 5 in relation to speech. The idea that 'more, and better speech' can undo the costs of free speech is a familiar example.

difference. Even if a person wants to neutralise the bias, it might prove just too costly; lots of mental resources are required for ‘policing’.

In today’s world, we are exposed to great amount of physical pollutants, such as asbestos, automobile emissions, and hazardous waste. We usually take action to avoid contact with these materials. Following the analogy, mental contamination might not be different. An appropriate response may suggest controlling exposure to the contaminant. This requires taking steps before contamination occurs, therefore saving mental resources for other purposes (rather than employing them in attempting to correct the effects of contamination<sup>103</sup>). Among other problems this strategy may have (it might cause people to confirm their biases, it might be more useful only against external, rather than internal, influences), one is of particular importance: it seems to be too demanding insofar as it asks people to be ‘ever vigilant, ready to shut their eyes and cover their ears whenever they suspect that potentially unwanted information was in the vicinity’<sup>104</sup>.

A possible solution is to delegate others to control the things to which we are exposed. The design of democratic and public procedures for controlling exposure to contaminant material might be critical in protecting individuals’ autonomy. Notice though that contamination occurs at different levels, the scope of which ranges from the trivial (what flavour of ice cream to choose) to the meaningful (behaving aggressively to people because we have just watched a violent movie, treating people differently due to their physical attributes). It would be impossible to control every source of contamination; trying to do so assumes unrealistic mental capacities.

---

<sup>103</sup> This is what we do with physical pollution; we avoid being exposed to it, rather than correcting its effects.

<sup>104</sup> Wilson and Brekke. ‘Mental Contamination and Mental Correction’, 135.



Autonomy as non-contamination needs to categorise the risks and weights of different types of contamination. Controlling exposure to the serious sources of biasing might be a necessary condition for autonomy. Autonomy seems to require that social relations are sensitive to contamination, and that the exposure to its sources be more or less socially controlled.

### *F) Revised Rationality*

There is much evidence now suggesting that people fail to satisfy the normative criteria of classical rationality. As explained in Chapter 1, people usually fail to apply the rules of logic and probabilistic reasoning when presented with problems. Rather, they make decisions based on rules of thumb and heuristics<sup>105</sup>. While some authors retain the normative criteria of decision theory, and argue that humans are beset by many kinds of irrationality, other have abandoned this model and instead sought a revised account. One of these revisions is the idea of ‘ecological rationality’<sup>106</sup>.

One of the main contrasts between classical and ecological rationality relates to the ideal of maximisation: whereas the former allocates it a fundamental role, the latter excludes it altogether. In this respect the ecological model is a form of bounded rationality. It is even more modest than satisficing rationality.

To make the contrast clearer, imagine someone who is deciding for which degree she will apply. There are many different options to choose from: medicine,

---

<sup>105</sup> See Chapter 1; classic studies appear in Kahneman, Slovic, and Tversky (eds.) *Judgment under Uncertainty: Heuristics and Biases*; Kahneman and Tversky (eds.) *Choices, Values and Frames*; Thomas Gilovich, Dale Griffin and Daniel Kahneman, (eds.) *Heuristics and Biases: The Psychology of Intuitive Judgment* (Cambridge: Cambridge University Press, 2002).

<sup>106</sup> Gigerenzer, *Adaptive Thinking: Rationality in the Real World*; Gigerenzer and Selten (eds.) *Bounded Rationality*; Gigerenzer and Todd, *Simple Heuristics that Make Us Smart*.

politics, economics, philosophy, sociology, and so on. Each department has some 'open days' to provide information about the degrees, careers etc. An individual deciding which degree to choose can follow (roughly speaking) one of two strategies. The first, which is consistent with classic, maximising rationality, consists in visiting every open day, weighing every piece of information, assigning numerical value to each one according to its expected utility and so forth<sup>107</sup>. Strategy two is more simplistic: it consists in fixing 'an adjustable aspiration level and ending the search for alternatives as soon as one is encountered that exceeds the aspirational level'<sup>108</sup>. This strategy is especially suitable for making decisions when time is scarce or when uncertainty is high.

Although this idea is attractive, according to Gigerenzer and Todd, it sometimes still demands a large amount of computational capacities. For example, it might be that even setting the aspiration level is hard to achieve, or 'calculating how the option compares with the aspiration level' might demand a great amount of attention and capacities<sup>109</sup>. So, in order to cope with these problems their proposal, 'ecological rationality', develops the idea of 'fast and frugal heuristics', which 'employ a minimum of time, knowledge, and computation to make adaptive choices in real environments'<sup>110</sup>, sets easy stopping rules in searching for information and exploit the environmental structure to yield adaptive decisions.

Ecological rationality highlights the fact that rational decisions are taken in specific environments. Whereas classical rationality focuses on internal logical

---

<sup>107</sup> There is also option 1'. This is similar to strategy one, but here individuals stop attending open days when attending them outweighs the benefit of the new information given in the open days. This strategy refers to 'optimisation under constraint'.

<sup>108</sup> Gigerenzer and Todd 'Fast and Frugal Heuristics: The Adaptive Toolbox', 13.

<sup>109</sup> Ibid. 14.

<sup>110</sup> Ibid. 14.

standards, such as internal coherence, transitivity, etc., ecological rationality requires being ‘adapted to the structure of environment’<sup>111</sup>. The life of real organisms consists in dealing with changing, challenging environments. These organisms must attempt to make rational decisions; that is, decisions that allow them to survive and reproduce<sup>112</sup>. It is in this sense in which the standards of rationality move from internal coherence to environmental fitness. The idea is to match decision-making strategies with the structure of the environment<sup>113</sup>: they exploit the recurrence of some environmental structures in order to reach ‘better’ results than unbounded versions of rationality. An example of fast and frugal heuristics is the ‘recognition heuristic’. Imagine that you are asked which city has a larger population, San Diego or San Antonio. The recognition heuristic tells you that ‘if you recognize one object and not the other, then infer that the recognized object has the higher value on the target variable; if you do not recognize either object, then guess’<sup>114</sup>.

Another problem with classical versions of rationality is their assumption that rational processes are general-domain rules for decision. It is in this respect in which classical rationality is not ecological, the rules are assumed to work in any environment. Ecological rationality, by contrast, assumes rationality as being essentially an evolutionary adaptation for solving specific problems, rather than an ability to make abstract computations. Adaptations ‘are mechanisms or systems of properties crafted by natural selection to solve specific problems posed by the regularities of the physical, chemical, developmental, ecological, demographic,

---

<sup>111</sup> Ibid. 12.

<sup>112</sup> Ibid. 18.

<sup>113</sup> Gerd Gigerenzer, ‘The Adaptive Toolbox’, in Gigerenzer and Selten, *Bounded Rationality*.

<sup>114</sup> Valerie Chase, Ralph Hertwig, and Gerd Gigerenzer ‘Visions of Rationality’, *Trends in Cognitive Sciences*, 2 (1998), 208. Cf. Daniel Goldstein and Gerd Gigerenzer, ‘Models of Ecological Rationality: The Recognition Heuristic’, *Psychological Review*, 109 (2002), 75-90.

social and informational environments encountered by ancestral populations during the course of a species' or population's evolution'<sup>115</sup>. In order to be successful, adaptations tend to be domain-specific rather than general because specificity allows them to respond more efficiently to environmental pressures<sup>116</sup>. Rather than the laws of formal logic, individuals tend to employ different heuristics which allow them to solve problems successfully.

Contrast the different approaches of general rules and domain specificity. According to the former, there are only a few all-purpose rules (essentially those of formal logic) which must solve every problem we encounter. Whether we are trying to engage in social exchange, choosing habitat, or finding a mate, we should make the decisions in accordance with these general rules. It is merely a matter of applying them regardless of the specific content of the problem. On the other hand, the adaptive approach suggests that rationality resembles a 'toolbox', where there are different rules and strategies specific to particular problems. The advantage of the adaptive toolbox seems to be that it is able to respond more efficiently to the regularities of the environment. Because organisms face situations that are not altogether new, they can employ the same strategies that have proved to be successful in the past, rather than devising them from scratch. It is important, however, to bare in mind that the specific tools need to be flexible and simple so as to cope with environmental variations<sup>117</sup>.

---

<sup>115</sup> Tooby and Cosmides, 'The Psychological Foundations of Culture', 62.

<sup>116</sup> This seems to be the evolutionary reason that explains the existence of the fast and frugal heuristics discussed in the section above.

<sup>117</sup> It is interesting that artificial intelligence works in terms of domain specificity rather than with all-purpose tools. There is software specific for text working, a different one for sound editing, etc.

There is some empirical evidence that supports the existence of the adaptive toolbox<sup>118</sup>. Think about an animal whose food usually appears scattered around at random. This animal will develop a heuristic mechanism for finding food which will consist in looking around randomly for it. In contrast, another animal's food appears in hidden places that can be discovered by learning and inference from cues. This animal will use a different heuristic, employing memory and inferences between food and cues<sup>119</sup>. In another example, people are very good at recognising other people's faces but not so good at recognising hands (although hands are as unique as faces), and also better at recognising people's faces than horses' faces. Recall Lawrence Hirschfeld's theory of race cognition, he suggests that we have a domain-specific module to conceptualise human kinds, and that module is distinct from concerns of natural kinds or living things<sup>120</sup>.

To sum up, the ecological notion of rationality is a version derived from the idea of bounded rationality. As such it abandons completely the ideal of maximisation. It has two dimensions. On the one hand it acknowledges the mind's limits in terms of mental and computational abilities and the time constraints which affect decision-making. On the other hand, it pays particular attention to the structure of environments within which agents interact. According to this theory, there are heuristics and strategies that are suited to solving specific adaptation problems and which have evolved by natural selection. This view of rationality also holds that the

---

<sup>118</sup> Gigerenzer, *Adaptive Thinking*; Gigerenzer and Selten, (eds.) *Bounded Rationality*; Wilson, *Strangers to Ourselves*.

<sup>119</sup> Gigerenzer and Todd 'Fast and Frugal Heuristics: The Adaptive Toolbox'. 13.

<sup>120</sup> Chapter 1; Hirschfeld, *Race in the Making*.

criteria for assessing the rationality of behaviour is not internal consistency, but rather its fit with reality and 'ecological' success<sup>121</sup>.

### III

#### Relations between Conditions and Challenges

In this section I will relate briefly the condition of autonomy with the aforementioned different challenges. Although the discussion is not exhaustive, the goal of this section is to provide an overview of how the conditions relate to each of the threats discussed above.

This table illustrates the relations between conditions and challenges:

Threats to Autonomy							
Conditions for autonomy		A	B	C	D	E	F
	1	—	—	—	—	X	—
	2	X	X	X	X	X	X?
	3	X	X	X	X	—	X?
	4	—	—	X	—	—	X
	5	X	X	—	—	—	X?
	6	X	X	—	X	—	X

1) Identification, 2) Right Processes of Formation, 3) Critical Reflection, 4) Content-Neutrality, 5) Adequate Options, 6) Classical Rationality. A) Adaptive Preferences Formation, B) Abnormal Belief Formation, C) Oppressive Socialisation, D) Automatic Behaviour, E) Mental Contamination, F) Ecological Rationality.

<sup>121</sup> Cf. 'The Standard Social Sciences Model requires an impossible psychology [...] A psychological architecture that consisted in nothing but equipotential, general-purpose, content-independent, or content-free mechanisms could not successfully perform the tasks the human mind is known to perform or solve the adaptive problems humans evolved to solve [...]. The alternative view is that the human psychological architecture contains many evolved mechanisms that are specialized for solving evolutionary long-enduring adaptive problems and that these mechanisms have content-specialized representational formats, procedures, cues and so on. These richly content-sensitive evolved mechanisms tend to impose certain types of content and conceptual organization on human mental life and, hence, strongly shape the nature of human social life and what is culturally transmitted across generations' Tooby and Cosmides 'The Psychological Foundations of Culture', 34.

*Identification*

**1E)** Identification encounters an obvious challenge from mental contamination. Recall that the latter is the ‘process whereby a person has an *unwanted response* because of mental processing that is unconscious or uncontrollable’<sup>122</sup>. A person whose responses are contaminated cannot identify with them. The teacher who wants to be fair but who nevertheless (unconsciously) gives a higher mark to a student because he likes her (the ‘halo effect’) cannot identify himself with his automatic response. The very definition of mental contamination is incompatible with the identification condition. Even though the definition of mental contamination is worded only in terms of products, it also applies to the outcomes of processes people reject. Imagine that after watching a violent movie in the cinema you respond aggressively to the bus driver. The aggressive response was caused automatically by the ‘chameleon effect’<sup>123</sup>. It is not only that you regret harbouring the *process* which prompted you to react aggressively, but also that you regret the *response* itself<sup>124</sup>. Mental contamination then needs not be restricted only to processes; it can also encompass their outcomes. In both cases it conflicts with identification.

*Right Causal History*

**2A)** Adaptive preferences are primarily and foremost in conflict with the condition of ‘right’ causal history. Autonomy requires our desires and preferences to be shaped by reasons and by the right mental processes. However, as stated in the last section,

---

<sup>122</sup> Wilson and Brekke, ‘Mental Contamination and Mental Correction’, 117. Emphasis added.

<sup>123</sup> For a discussion on the links between violence in media and automatic behaviour see Hurley, ‘Bypassing Conscious Control’.

<sup>124</sup> This does not mean that both go necessarily together; there might be cases in which you regret having the process but not the outcome (someone manipulates you into exercising more often than you normally do, and you feel quite happy about the result, nevertheless you still reject being manipulated).

individuals sometimes adapt their preferences in order to avoid cognitive dissonance; a ‘painful’ or ‘distressing’ psychic phenomenon consisting in having a set of mental states which does not cohere with a different set of mental states. For instance, in the case of sour grapes, cognitive dissonance results when someone’s desires are not likely to be satisfied due to beliefs regarding external circumstances. The mechanism is triggered in order to avoid the frustration involved in knowing that certain things one desires will not happen. It is fundamental to remember, however, that what threatens autonomy is not the existence of cognitive dissonance, but the mechanism by which the mind reduces it: that is, the processes by which our desires and preferences are shaped. This mechanism is purely causal as opposed to being operated by the agent’s intentions<sup>125</sup>.

Contrast this phenomenon with instances of character planning, in which a person deliberately decides to abandon desires that he could not fulfil. Although their phenomenology might be difficult to pinpoint (people can always rationalise their adaptations), in the case of sour grapes preference change occurs due to drives, in the case of character planning it is caused by meta-preferences<sup>126</sup>. A means of perceiving the difference is that adaptive preferences usually downgrade the inaccessible options. In the case of Beto, he probably would say: ‘Anyway, life in England surely is horrible!’. Whereas character planning tends to upgrade the accessible options: if I am not entirely happy in the University for which I work, I could try to reduce the cognitive dissonance by focusing on the (few) positive elements of my situation.

---

<sup>125</sup> Elster, *Sour Grapes*, 110. Strictly speaking this is not true. Intentions are also causally effective in bringing about actions. The distinction means that intentional causes play no role in reducing cognitive dissonance.

<sup>126</sup> Ibid. 117-9.



Another process which, in terms of results, resembles adaptive preference formation is precommitment. In this case an agent deliberately shapes the set of available options in order to make his preferred choice within the set to coincide with his preferred option in a wider range of conceivable possibilities. Ulysses' strategies are an illustration of these mechanisms<sup>127</sup>. Someone can pay the whole dental treatment in advance in order to make it very costly for him to withdraw. Notice that both character planning and precommitment also reduce cognitive dissonance, but they do so in a rational, autonomy-friendly fashion.

**2B)** Similarly, poor belief formation and wishful thinking violate the condition of mental states been brought about in the 'right' way. For instance, in the case of wishful thinking, a person's beliefs are shaped by the desire for something to be the case, not by salient evidence. In order to be autonomous and respond rationally to the world our beliefs need to be shaped by external inputs, or more specifically reasons, not merely by our desires.

**2C)** It is not clear whether oppressive socialisation threatens the right processes of belief and preference formation, but there seems to be a *prima facie* contradiction between them. While the right history is purely procedural and content-neutral, oppressive socialisation is usually cast in substantive terms. This leads to a dilemma, we can either endorse a substantive account of autonomy, rejecting content-neutrality<sup>128</sup>, or we could deny that oppressive socialisation is a

---

<sup>127</sup> Cf. *ibid.* 114-5. Cf. Jon Elster, *Ulysses and the Sirens* (Cambridge: Cambridge University Press, 1984).

<sup>128</sup> Paul Benson, 'Autonomy and Oppressive Socialization'. *Social Theory and Practice*, 17 (1991),

problem for autonomy. A possible solution may be to formulate a procedural yet content-neutral sense in which some forms of socialisation thwart autonomy. We could describe some forms of upbringing as autonomy enhancing: we must educate children to appreciate evidence, to respond to it in the right fashion, and to form rational preferences, not to react only on whim.

**2D)** Although automatic responses are an essential part of our cognitive architecture, they can, nevertheless, threaten the history of our preferences and beliefs. Sometimes we form goals and manifest preferences which have not been produced through rational processing, but only through automatic associations. Although in many cases automatic and rational processes might overlap (as when we automatically categorise the thing in front of us as a ‘living’ rather than as a ‘nonliving thing’), sometimes they go awry; we form social and racial stereotypes<sup>129</sup>, we make judgments based on the halo effect, and so on. It is this sense in which external influences can trigger automatic responses that have not been thought through in the right way.

**2F)** It is unclear whether ecological rationality can provide an answer to which processes of belief and desire formation should be considered ‘right’. If rationality is relative to specific environments, then perhaps processes of formation are also relative. It could be that a determinate process A is only adequate in environment B. If this is the case, we might need to describe the right way in which

---

385-408.

<sup>129</sup> The relation between racial stereotypes, mental contamination and autonomy will become clearer in the following chapters.

beliefs and desires are formed not in general terms, but domain specifically. These are hard questions, and it seems that there is still important research to be done regarding the relationship between these processes and different features of the environment<sup>130</sup>.

### *Critical Reflection*

**3A)** Adaptive preferences can challenge the condition of critical reflection in an important way, as the phenomenon suggests agents cannot revise and assess their life plans in a normatively significant fashion if ‘some’<sup>131</sup> of their desires have been adapted by these wrong processes. This is so because the very standards of critical reflection might themselves have been distorted. For instance, if an agent revises his conception of the good by accommodating adapted goals and desires, then this critical revision would have no normative force whatsoever. When Beto decides that playing for Arsenal is not attractive and he revises his life plan accordingly, it is already defective insofar as it is the outcome of (or has been contaminated by) adapted preferences. If the goals that prompt and lead to critical reflection are themselves adapted the exercise of reflection is normatively empty. For autonomy it is important to have reflection that is free from such flawed processes.

Two remarks are of note here. First, usually agents *ignore* that their preferences have been adapted in this manner. Think about Beto’s case; he believes (probably through a process of rationalisation) that his not attending his trial is

---

<sup>130</sup> See the collection of essays in Carruthers, Laurence, and Stich, (eds.) *The Innate Mind I: Structure and Contents*; Carruthers, Laurence, and Stich, (eds.) *The Innate Mind II: Culture and Cognition*.

<sup>131</sup> The word ‘some’ here refers to both quantity and quality of the preferences adapted. While having adapted preferences about lollipops’ flavours might not be very worrying, being subject to those processes regarding the whole domain of gastronomy and nutrition certainly is.

caused by the reasons he provides (London is grey and cold, he will miss his mother's culinary talents, and so on), when in fact his desires have been shaped by his lacking the economic resources to travel. In this example, he is not aware that his preferences have been shaped to conform to his circumstances. In this respect adaptive preference formation is an automatic response, and it is easy to see why it threatens autonomy. However, and this is the second remark, things can be more complicated. Imagine that we explain to Beto what has affected his thinking, and he answers that that he does not regret having this response, because somehow the fact that he (falsely) believes that his decision is due to *his* reasons is more comfortable than acknowledging that his not travelling in London is due to poverty<sup>132</sup>.

Some have argued that eliminating of adaptive preferences requires an 'ability to do otherwise'<sup>133</sup>. An interpretation holds that people should be able to actually act in a different fashion. This interpretation, however, is problematic insofar it opens us to 'Frankfurt-style cases'<sup>134</sup>. Imagine that Pepe (the priest in Beto's town) does not want Beto to go to London. In order to prevent it, Pepe implants a device in Beto's brain, such that if Beto decides to go the device will make him reject the offer on the grounds that London is cold and grey, that he will miss his mother, etc. In the end, there is no need for the device to work, Beto decides not to show up and misses the chance to play for Arsenal. Frankfurt's intuition

---

<sup>132</sup> In this case, his adapted preferences would not count as a form of mental contamination. But we need to ask, whether this response is not itself the outcome of an adapted preference, which obviously opens the door to a regression ad infinitum. Nevertheless, I think most people would regret having their preferences adapted. People have an interest in responding to the right reasons, not just to reduce cognitive dissonance.

<sup>133</sup> Elster, *Sour Grapes*, 128. This condition has been the source of many important debates regarding responsibility. I cannot do justice to these controversies in this thesis.

<sup>134</sup> Harry Frankfurt, 'Alternate Possibilities and Moral Responsibility', *The Journal of Philosophy*, 66 (1969), 829-39.

shows that the fact that Beto could not have acted otherwise makes no difference in our assessment of Beto's action<sup>135</sup>.

So, it does not seem to matter whether Beto *could* have acted otherwise; the relevant question is whether he *would* have. But, turning the choice into a hypothetical one does not seem to solve the problem either, as hypothetical choice can also be the victim of processes of adaptation<sup>136</sup>. Imagine we want to know whether Maria's preferences are adapted. A requirement of hypothetical choice suggests that if she would choose to have the preferences she has and to live the lifestyle she does, then her preferences are not adapted. However, a weakness of this requirement is that hypothetical choices can also be the product of adaptation. Her circumstances might be such that she cannot even imagine, or respond, to other, very different, ways of life<sup>137</sup>.

**3B)** Abnormal processes of belief formation threaten critical reflection in a very similar fashion. Reflection requires that agents are able to evaluate relevant evidence, but most of the processes that ill-shape beliefs preclude agents from executing exactly those evaluations. Recall Ernesto, the hippie from Mid-Wales. Imagine that when he is revising his life plans a significant part of his beliefs are the product of wishful thinking, so that the outcome of his critical exercise merely mirrors the way he wishes the world to be. In terms of autonomy his revising his life

---

<sup>135</sup> Cf. Susan Hurley, *Justice, Luck, Knowledge* (Cambridge, MA: Harvard University Press, 2003), Chapters 1-2.

<sup>136</sup> Cf. *ibid.* 28-30.

<sup>137</sup> Some authors try to eliminate the problem of adaptive preferences by using 'informed preferences' instead; preferences that people would have if they had perfect information and were not influenced by illegitimate processes. However, these preferences are problematic, as they require far too much in terms of mental resources. Cf. Mozaffar Qizilbash, 'Well-Being, Adaptation and Human Limitations'. *Royal Institute of Philosophy Supplements*, 81 (2006), 83-110.

plan is meaningless; a wrong process has constrained him since the beginning. Importantly, autonomy is concerned with acting for reasons in the sense that our actions and responses are motivated by the latter rather than by delusion and self-deception. In difficult circumstances we can plan our character, and try to use rational strategies to reduce cognitive dissonance, but ‘sour grapes’ and ‘wishful thinking’ do not allow us to be autonomous in terms of fulfilling the critical reflection condition.

**3C)** How does oppressive socialisation challenge critical reflection? First it is clear that individuals are socialised during childhood and the values inculcated in them by nurture (non-critically unless this is encouraged) form the standpoint from which future critical judgement and reflection will be undertaken. This does not mean that individuals cannot step back and criticise the values they were taught, but it does suggest that the value system inculcated during childhood has a special place in people’s conception of the good. Imagine Amy, a woman who has had a very traditional upbringing. As an adult she is expected to get married and give up her career to become a housewife. After deep thinking, she realises that she cannot embrace those values. She decides to get married, but also to keep her job. Her decision is not easy, she feels guilty for disappointing her parents and her community. What this example shows is that values inculcated during childhood have an impact on how adults lead their lives<sup>138</sup>.

Oppressive socialisation poses an important problem for autonomy because of its connection to critical reflection. Autonomy seems to require that both the

---

<sup>138</sup> Cf. Cohen, *If You’re an Egalitarian How Come You’re so Rich?*, Ch. 1.

outcomes and standpoint of critical reflection are exercised freely or rationally. In the case of oppressive socialisation, however, the standpoints are defective<sup>139</sup>. What is required, then, is a starting point (a basic socialisation) that fosters, rather than inhibits, the capacities for critical judgement. The task is to identify what sorts of upbringing are compatible with autonomy and what kind of socialisation children can legitimately engage with. Although this question has no easy answer it seems that socialisation needs to be targeted, at least towards the development of those capacities that are needed for critical reflection and for neutralising distorting influences, such as adaptive preferences, wishful thinking and mental contamination.

There might be two different approaches to addressing this problem. On the one hand we could try to set up a purely procedural method of upbringing. Under this strategy, we should focus only on cultivating *capacities* for autonomy, such as the ability to link means and ends, make congruent plans, resolve internal conflicts in an intelligible way, etc. Notice that this strategy says nothing about the content of education: it claims only that socialisation is acceptable (regarding autonomy) insofar as it inculcates the capacity for autonomy.

On the other hand, a more robust strategy might be that socialisation should include more stringent requirements. For instance, we could also demand that socialisation satisfies certain criteria of justice, so that children can learn and develop a sense of it<sup>140</sup>. Or, even more controversially, we could argue that upbringing should satisfy constraints of public reason<sup>141</sup>. These are important questions, but I

---

<sup>139</sup> Cf. Christman, 'Autonomy and Personal History', 19-20.

<sup>140</sup> Okin, *Justice. Gender and the Family*.

<sup>141</sup> Clayton, *Justice and Legitimacy in Upbringing*.

shall leave the discussion here for a moment. I shall say a bit more about this in Chapter 6.

**3D)** Automatic responses can jeopardise critical reflection by interfering with the processes of reasoning. As we have seen, implicit associations and concept activation are difficult to control, and might make us react in ways that conflict with our reasons. Moreover, there are a number of cognitive failures that make critical reflection harder to achieve. For instance, as I shall explain in Chapter 5, we have a tendency to treat information we know is false as if it were true, and on occasions we act based on this information. An individual might find these responses ‘alienating’, in the sense that they are not the product of her values and beliefs but rather of mental processes that she does not control.

**3F)** It is not clear how ecological rationality threatens critical reflection. On the one hand, we usually think that critical reflection is subject to general domain constraints, such as internal consistency, transitivity, etc. However, ecological rationality does not require such conditions; it relies on fast and frugal heuristics and domain specific procedures. This means that, in an important sense, critical reflection outsources ecological rationality. On the other hand, it may well be the case that critical reflection needs to be relative to the structure of environments. A possibility is that rationality has two tiers: a lower-level which is automatic, ecological, domain-specific, and fast and frugal, and a higher-level which is intentional, costly, domain-general, and slow<sup>142</sup>. Although this seems an

---

<sup>142</sup> Hurley, ‘The Public Ecology of Responsibility’.



oversimplification, we need to know more regarding the architecture of rationality to provide more definitive answers<sup>143</sup>.

### *Content-Neutrality*

4C) The fundamental question in this section is whether there is a content-neutral specification of the fashion in which oppressive socialisation might thwart autonomy. A possible strategy is to follow a bias neutralising approach<sup>144</sup>. This approach holds that

we can often know that certain biasing influences tend to undermine knowledge, even when we make no politically controversial assumptions about the truth. Moreover, we do not need to know what precise relationship between the truth and belief makes for knowledge in order to know that certain factors tend to defeat knowledge<sup>145</sup>.

As noted above, some processes of poor belief formation, such as wishful thinking, have a negative impact in the search for knowledge by creating cognitive biases. Similarly under certain conditions people are less able to reflect upon the information they are given<sup>146</sup>. The manner in which it is framed can also bias their responses towards certain reactions<sup>147</sup>. Moreover, environmental conditions might trigger associations that do not contribute to knowledge such as reliance on racial and gender stereotypes<sup>148</sup>. This condition seems compatible with content-neutrality

---

<sup>143</sup> Chater and Oaksford, 'Human Rationality and the Psychology of Reasoning'.

<sup>144</sup> Susan Hurley, 'Cognitivism in Political Philosophy', in Crisp, R., and B. Hooker (eds.) *Well-Being and Morality: Essays in Honour of James Griffin* (Oxford: Clarendon, 2000), 177-208; Hurley, *Justice, Luck, and Knowledge*, Ch. 10.

<sup>145</sup> Hurley, *Justice, Luck, and Knowledge*, 257.

<sup>146</sup> Gilbert, 'How Mental Systems Believe'. Cf. also Chapter 4 in this thesis.

<sup>147</sup> Kahneman and Tversky 'Choices, Values and Frames'.

<sup>148</sup> Devine, 'Stereotypes and Prejudice'. See also Chapter 5 in this thesis.

because it does not make any claims about the truth; it merely specifies that certain influences will jeopardise the truth, regardless of its content.

A second feature of a cognitivist approach claims that ‘we should develop and use effectively certain capabilities of citizens. The cognitive capabilities of individuals to respond appropriately to relevant reasons are valuable both in their own right and to the social search for knowledge of that should be done’<sup>149</sup>. This feature is more controversial because it claims that certain capacities are more desirable than others, and it arranges institutions to promote these, but not other, capabilities<sup>150</sup>. Although this feature is not neutral towards the capabilities required for having an autonomous life, it is, nonetheless, neutral towards its contents<sup>151</sup>.

**4F)** Ecological rationality is domain-specific in terms of content, whereas autonomy requires content-neutrality. How are these two ideas related? The first thought is that, if ecological rationality cannot be detached from specific contents then it may be the case that autonomy cannot, after all, be content-neutral either. This suggests that we may have to distinguish between different domains, and that the possibility of neutrality might be tied to those particular domains. This distinction could imply that we must differentiate between two levels. First, we might ask whether neutrality is possible *between* domains. To which it seems that the most plausible answer is ‘no’. It is difficult to see how we could make ‘trade-offs’

---

<sup>149</sup> Hurley, *Justice, Luck, and Knowledge*, 257.

<sup>150</sup> Notice that Hurley believes that her bias-neutralising approach is not neutral; when applied to political institutions she claims it endorses a form of liberal perfectionism. Cf. *ibid.* 258.

<sup>151</sup> It is very likely that citizens who develop these capabilities will reject some forms of life. However, this is not a worry; neutrality requires that we do not impose some contents on autonomy, not that we give every form of life an equal chance. Cf. Chapter 4.

between them. If rationality is essentially linked to certain domains we might not be able to step back and be neutral towards them.

On the other hand, the second question is whether we might be neutral *within* different domains. Neutrality may be possible regarding some domains, but not others. The answer here may be determined by the structure of the domain itself<sup>152</sup>. This distinction somewhat resembles one that has become popular in discussions about free speech. Some people claim that although protection of speech need not be neutral regarding categories of speech, it must be neutral within those categories<sup>153</sup>. For instance, we do not need to be neutral between political and commercial speech, and can, therefore, give greater protection to the former, in virtue of some ‘domain’ specific features (political speech is more important vis-à-vis autonomy, or democracy, or truth, etc.). Nevertheless, neutrality might be required within those categories; we should not grant more protection to libertarians than to Marxists, for instance. The point here is not to defend the strategy of categorisation, but merely to illustrate how there might be some links between domain-specificity in ecological rationality and discussions about content neutrality in free speech<sup>154</sup>.

### *Adequate Options*

**5A)** Adaptive preferences do not shape the conditions of choice directly. Nevertheless, they can affect the manner by which agents assess their options and choices. For example, Beto sees the option of playing at the Emirates Stadium as less

---

<sup>152</sup> These answers are tentative. On the one hand we still do not know enough about ecological rationality, on the other, we are too familiar with the classical conception of rationality, making the shift hard. Nevertheless, asking the questions might help us to clarify some ideas, even if we do not have answers yet.

<sup>153</sup> Scanlon, 'Freedom Of Expression and Categories of Expression', 84-112.

<sup>154</sup> I will say more about the presumption about content-based regulations on Chapter 5.

attractive than staying in his hometown, and Ernesto thinks that living a non-hippie life would conflict with his conception of the good. More importantly, the distortions these processes cause have important consequences in the options agents face later. For instance, after failing to attend the trials (due to his adaptive preferences) Beto will not have the option of visiting London and encountering the choices it offers, and Ernesto will not have the choice of experiencing an urban life.

**5B)** Poor belief formation affects the right options requirement in a very similar fashion to adaptive preferences. We act on beliefs, and we take them as reasons for actions. Those actions, then, affect the options we have in the future. When we form beliefs in the wrong way we do not act on right reasons, and the options we encounter are not the right options. Autonomy does not (and cannot) require that we assess those consequences at every juncture, but it does demand that when we make important decisions, we do so by assessing the options in a ‘right’ way. That is, at least the assessment should not be distorted by adaptive preferences and ill-shaped beliefs (as the paragraph (3A) on critical reflection above discusses)<sup>155</sup>.

**5F)** Does an ecological account of rationality threaten the idea that autonomy requires the right options? Ecological rationality is relative to environments, and to the options it offers. Maybe this means that autonomy would inherit some ecological relativity. If so, then we might need to judge the options agents encounter not in general, but according to certain environments. And it may be the case that some

---

<sup>155</sup> Cf. Mill, *On Liberty and Other Essays*; Elizabeth Anderson, 'John Stuart Mill and Experiments in Living'. *Ethics*, 102 (1991), 4-26.

environments offer options that are plainly incompatible with autonomy, while other environments are richer and more autonomy-friendly.

### *Optimising Rationality*

6A) In general, classical rationality is concerned primarily with finding the best means for given goals. Adaptive preferences are beyond the reach of this approach, because it lacks the conceptual tools to assess the causal history of ends<sup>156</sup>. Two persons with equally satisfied preferences will score equally in terms of rationality, even if one of them holds adapted preferences but not the other. Similarly, this assumption fails to make sense of how people can alter their preferences. It cannot distinguish between changes in preferences that are due to meta-preferences, or other rational processes, and adaptations caused by drives.

Not accommodating for the causal history of preferences into account has well-known implications in political theory: a natural reading of this assumption produces a form of welfarism (be it utilitarian or not), which assumes that the most important variable in social choice is the rate at which people satisfy their preferences. This view, however, has some perverse incentives, such as giving more value to a state of affairs in which everyone harbours adapted preferences, rather than one in which open opportunity creates expectations that lead some people to frustration<sup>157</sup>. Another well-known problem is that sometimes it recommends giving more resources to people with expensive tastes (however formed) than to people with cheaper ones<sup>158</sup>.

---

<sup>156</sup> Elster, *Sour Grapes*, 30, 121.

<sup>157</sup> Ibid, 136.

<sup>158</sup> There is a vast amount of literature on the topic. Cf. Dworkin, *Sovereign Virtue*, Chapter 1; Ronald

**6B)** Poor belief formation might thwart the processes by which agents are to decide the best way to realise their goals. Classical rationality is fairly demanding in this sense, for it requires agents to maximise their expected utility functions. If someone forms beliefs in a poor fashion, then his actions will not be the most appropriate means to pursue his goals. For instance, rationalisation might work this way in cases of cognitive dissonance. An agent can alter the perception of the situation he is in, forming a belief that might not be instrumentally effective in achieving his goals. If someone does not attain the promotion he is hoping for, he might think that there is a conspiracy against him. By believing this, he might miss future chances of achieving the promotion he wishes<sup>159</sup>.

**6D)** Automatic behaviour presents a challenge to this conception of rationality by highlighting that much significant behaviour is not under instrumental control. Moreover, automatic processes affect goals, not only means for action. Classical rationality works under the assumption that agents control the processes which lead to their actions, by choosing the best means to achieve their goals. But, by contrast, phenomena such as implicit cognition, the chameleon effect, and concept activation show that some of these processes are not controlled.

---

Dworkin, 'Sovereign Virtue Revisited', *Ethics*, 113 (2002), 106-43; G. A. Cohen, 'On the Currency of Egalitarian Justice', *Ethics*, 99 (1989) 906-44; G. A. Cohen, 'Expensive Tastes Ride Again', in Burley, J. (ed.) *Dworkin and His Critics* (Oxford: Blackwell, 2004), 3-29.

<sup>159</sup> He could also reduce the cognitive dissonance by adapting his preferences. He might say that, after all, he did not fancy the new post.

6F) Ecological rationality is a response to the failures of classical rationality, so its relation is straightforward. The contrast should now be clear. The latter argues that agents' rational behaviour consists in maximising their utility function<sup>160</sup>. Maximisation functions on three levels: the action must be the best possible mean to achieve the purposed end, the belief for the action must be the best belief given the available evidence, and the evidence must be the best possible given the constraints (primarily time and space). And ideally, an explanation according to rational choice theory '[w]ould show that the action is the (unique) best way of satisfying the full set of agent's desires, given the (uniquely) best beliefs the agent could form, relatively to the (uniquely determined) optimal amount of evidence'<sup>161</sup>.

As a result of much evidence showing that agents fail to satisfy these conditions, ecological rationality abandons the ideals of optimisation and internal consistency. Instead, rationality is defined as relying on fast and frugal heuristics that exploit cues in the environment. This form of rationality can be extended to incorporate more than single individuals; it can function at the level of different units<sup>162</sup>. It essentially holds that rationality and environments are embedded.

---

<sup>160</sup> The conditions defining 'preference ordering' and 'utility function' are the following: *Reflexivity*: For any bundle A, A=A. This condition demands that each bundle of goods is always as good as itself; *Completeness*: For any two bundles, A and B, either A>B, or B>A, or A=B. This means that every available good is subject to be ordered and compared with another available good; *Transitivity*: if A>B, and B>C, then A>C; *Continuity*: if A>B, and A undergoes a very small change, then A>B and ~(B>A); Imagine that someone prefers to eat a sandwich while watching a football match (A), over having the same sandwich while watching a Rugby Union match (B). His preferences are continuous if he still prefers A, even if the lettuce is taken from the sandwich. When someone's preferences satisfy (i-iii), we can say that he have a consistent and complete preference ordering, when additionally his preferences satisfy 'iv' as well, they can be represented by a utility function. The utility function permits us to assign numbers to the preferences and, thus to say that while choosing, he is maximizing his utility function. Shaun Hargreaves Heap, et. al. *The Theory of Choice: A Critical Guide* (Oxford: Blackwell, 1992), 6, 368.

<sup>161</sup> Elster, 'The Nature and Scope of Rational Choice Explanation', 71.

<sup>162</sup> Hurley, 'Social Heuristics that Make Us Smarter'.

Another set of problems that plague classical rationality relate to the use of general rules of logical inference. Wason tasks have been used to show that people systematically fail to make simple logical operations of the form ‘if P then Q’. In the standard form of the experiment, ‘the rule concerns cards, which have a number on one side and a letter on the other’. A classical rule is ‘if there is a vowel on one side ( $p$ ), then there is an even number on the other side ( $q$ )’. Four cards are presented to the participants, so that they can only see one side. The visible faces show an ‘A’ ( $p$  card), a ‘K’ ( $not-p$  card), a 2 ( $q$  card) and a 7 ( $not-q$  card). ‘Participants then select those cards they must turn over to determine whether the rule is true or false’. Normal results are: ‘ $p$  and  $q$  cards (46%),  $p$  card only (33%),  $p$ ,  $q$ , and  $not-q$  card (7%), and  $p$  and  $not-q$  card (4%)’<sup>163</sup>. Obviously (or not, given the results of the tests), the right result is turning over the  $p$  and the  $not-q$  cards, but it is usually only 4% of people who get this right.

These results have been interpreted in a variety of ways. Among others is the idea that there are domain-specific cognitive modules<sup>164</sup>. However, when the content of the Wason task is presented in different terms, participants tend to perform better. This is remarkable when the propositional content is related to social contracts in which cheating and free riding is possible. Instead of pairing vowels with even numbers Leda Cosmides and John Tooby used the rule ‘if a person is drinking beer, then he must be over 20 years old’<sup>165</sup>. With the different content, about 75% of participants picked the correct cards. These results suggest that people are good at

---

<sup>163</sup> Chater and Oaksford, ‘Human Rationality and the Psychology of Reasoning’, 195.

<sup>164</sup> See my remarks in Chapter 1.

<sup>165</sup> Cosmides and Tooby ‘Cognitive Adaptations for Social Exchange’. 183. Cf. also discussion in Gigerenzer, *Adaptive Thinking*, Ch. 10. But for alternative explanation see Chater and Oaksford, ‘Human Rationality and the Psychology of Reasoning’ 206-7.



making deductions of this type when facing certain problems in particular environments. In the case of Cosmides and Tooby's theory, people are good at detecting cheating because they have a cognitive module that is content- and domain specific.

## Conclusions

### *Toward Ecological Autonomy*

What kind of autonomy remains once all these challenges have been considered? Clearly we cannot ignore the fact that many of our activities are non-autonomous. The aims of sections two and three has been to note how general automatic behaviours and other forms of biases are, and to highlight how deeply they threaten the ideal of autonomy. There seem to three possible responses to this problem. First, we could abandon the idea of autonomy. We could argue that, because many of our actions are non-autonomous, we should say that autonomy is impossible, or merely an illusion. Or, we could say that it is so difficult that we should not regard it any longer as an ideal of political morality.

Second, we could revise and adjust some of the conditions for autonomy in light of this information. For instance, we could argue that autonomy requires a different conception of rationality, one that is ecological rather than general-domain and purely internalist. Or perhaps, we could suspend the condition of content-neutrality. These revisions, however, still maintain the fundamental intuition about autonomy; the idea of a self-governing agent. Conditions other than content-neutrality and the assumptions of classical rationality seem sufficient for that.

Third, consistently with the second response, we can think about ways to minimise our propensity to be beset by processes of mental contamination (and forms of irrationality). These include controlling certain features of the environment that trigger those reactions and maintaining the social conditions which facilitate the development of autonomy. I suggest that a mixture of the second and third options is the most attractive response to the challenges described above. In this regard, my suggestion is to abandon the assumptions of classical rationality, and embrace a more ecological, yet not substantive, account. The content-neutral clause should be suspended temporarily, until we understand exactly how it relates to the structure of environments. Regarding the first three conditions, I suggest that we need to take positive steps to neutralise unwanted influences which might thwart them. In order to do so, it is important to recognise that autonomy is a function of two variables. On the one hand, people's deliberative capacities need to be developed and exercised in an appropriate possible fashion. On the other hand, social environments must be sensitive to the way in which they influence people. We should not rely on people's capacity for mental control. We know we are not very good at it. In this respect the condition of 'right options' has considerable force. Social environments need to be carefully assessed regarding the impact they have on autonomy.

### *Ecological Autonomy and Free Speech*

After considering the conditions of autonomy and some challenges which threaten it, we can begin to see why the autonomy defence of free speech is so strong. First, free speech can be an important source for avoiding and rebutting adaptive preferences and wishful thinking. Being exposed to other people's expression can lead us to plan

our characters better and to evaluate evidence in a successful way, thus overcoming self-deception and delusion. Expression is an important means of fostering critical reflection and deliberation regarding one's life plans and conception of the good. Moreover, free speech might itself be a component of the right options we must hold in order to lead an autonomous life.

Nevertheless, we cannot neglect the costs of speech. As noted in Chapter 1, if there is a principle of speech, it must recognise that it goes beyond harm and provide reasons for protection independently. Speech can be highly contaminating: we can be misled into believing false things, it can trigger automatic and unwanted processes, we can be manipulated, and our processes of reflection can be impaired by information we do not want to receive. Much information can also saturate our deliberative capacities, triggering collective biases and cognitive failures, from which, perhaps with less speech available, we would not suffer.

Some might object that the conception of autonomy developed in this chapter is too 'thick', and that a 'thinner' one could function better, or at least equally well with less work. This objection, however, is not convincing. Remember that we need to find positive reasons for justifying the free speech beyond the scope of the harm principle: a thin conception fails to provide the extra justification necessary. Autonomy understood as the absence of coercion, for instance, provides little ground for protecting free speech.

Imagine that we want to decide whether tobacco advertisement on TV should be protected. We can distinguish between the different parties involved and consider how their autonomy is affected or promoted by the regulation in question<sup>166</sup>. First,

---

<sup>166</sup> I will employ and discuss this strategy at length in Chapters 3 and 4.

from the point of view of an audience, it is not clear that autonomy (as non-coercion) gives a definite answer. For, in what sense is an audience more or less coerced by the possibility of watching tobacco ads on TV? Regarding audience's autonomy regulations on advertisement (or its protection) does not seem to make a difference.

Second, restrictions on advertisement certainly affect would-be speakers. It is clear that those who wish to advertise their products on TV are coerced (and to that extent less autonomous) if not allowed to. Nevertheless, this response fails to capture some of our intuitions about both coercion and free speech. It is plausible that some degree of coercion is acceptable for the sake of speech<sup>167</sup>. Imagine that TV companies decide to broadcast only chat shows on the basis that they maximise profits. The government might, for instance, legitimately force them to broadcast some programs on political campaigns, on the grounds that citizens have an autonomy-based interest in being informed on political controversies (say, in order to develop their sense of justice)<sup>168</sup>. Notice, though, that audiences are neither more nor less coerced when the regulation is in place than when it is not. If autonomy (as absence of coercion) were the basis of free speech, then the government would be wronging TV companies. Moreover, sometimes it is acceptable to impose some coercion in order to warrant fair access to speech<sup>169</sup>. For instance, governments might force broadcasters to give some airtime to unpopular or disfavoured views, which would not have access to speech otherwise.

---

<sup>167</sup> Cf. Fiss, *The Irony of Free Speech*, in which he describes how the state (through coercion) can be the 'friend of speech'.

<sup>168</sup> Cf. Dworkin, *Is Democracy Possible Here?*, 150-154.

<sup>169</sup> Dworkin, *Sovereign Virtue*, Chapter 10; Scanlon, 'Freedom of Expression and Categories of Expression', 103-4.

These remarks suggest that coercion is not a bad thing *per se*. It is acceptable when it brings about certain outcomes (for instance, a robust debate on important political issues, or the dissemination of controversial lifestyles) and when it is justified in certain ways (for instance through public, democratic procedures, or by reasons no one could reasonably reject). If these remarks are plausible, then autonomy as non-coercion cannot be the benchmark of free speech. These remarks suggest, though, that the initial appeal of the absence of coercion is due to a thicker notion of autonomy, one that is captured by the conditions discussed in this chapter.

With those qualifications and conditions addressed let us now turn to two influential defences of free speech. In the next chapter I will discuss Thomas Scanlon's 'Millian Principle'. In Chapter 4, I will discuss Ronald Dworkin's defence.

## CHAPTER 3

### THE MILLIAN PRINCIPLE AND AUDIENCES' INTERESTS

A common strategy for defending free speech involves first distinguishing the different interests that free expression serves. Given that speech has benefits on some persons and costs on others, '[a]ny decision [...] must be made by considering how [it] would affect these interests by weighing the interests against one another and by considering their significance for the affected parties'<sup>1</sup>. Imagine that an individual wishes to protest against nuclear weapons at 3:00AM in a residential area. According to this strategy, we should identify a) the interest this person has in speech, b) the interests people who live in the area have in sleeping undisturbed, and c) reaching a balance. In this example, it is reasonable to suppose that the protestation could occur equally well at a different time, or at a different place, in order to protect the interest of residents. In the case of autonomy-based defences of speech, the relevant interests involved are all autonomy- related.

In this chapter I explore the interests of audiences in speech. I will consider the interests of speakers in the next chapter. It will suffice to say here that a speaker has an expressive interest in communicating her views to a wide audience: she might

---

<sup>1</sup> McKinnon, 'Should We Tolerate Holocaust Denial?', 16.

want to bear witness to matters of justice, or about important features of her conceptions the good. Besides, she also has deliberative interests in speech, articulating a voice might be a valuable way to exercise critical reflection.

Although, by definition, audiences do not have expressive interests, being exposed to speech serves their autonomy in several ways. First, an environment in which speech is protected provides people with a variety of alternative views, from which they can choose the most convincing. Second, people have an interest in being presented with (or at least having access to) different evidence and valuable information to make up their minds about certain issues. In order to take a stance on the merits of the Labour Party (before an election, for instance), I need access to its manifesto, to its leaders' speeches, and to information on its position regarding policies I consider important. I also need to witness the responses its representatives give when questioned on controversial issues. I may even need access to writings on political and moral theory to compare and contrast Labour policies with the viewpoints I have reasons to endorse. This is true not only for important political decisions but also for everyday choices, such as what brand of clothes I wear and which Internet supplier I pay for their service. Third, free speech serves people insofar as exposure to expression sometimes triggers critical reflection regarding their own commitments and beliefs, inducing them to reassess and critically examine their convictions. The condition of critical reflection, discussed in Chapter 2, demands that people be predisposed to evaluate their views when available evidence requires this. And it is likely that this requirement is better satisfied within an environment in which free speech prevails than one in which it does not. This seems to be what John Stuart Mill had in mind when he wrote: '... if opponents of all

important truths do not exist, it is indispensable to imagine them, and supply them with the strongest arguments which the most skilful devil's advocate can conjure up'<sup>2</sup>.

In addition to the interests of audiences we must also consider those of bystanders. First, bystanders have an interest in not suffering the undesirable effects of some acts of expression<sup>3</sup>: people protesting against war in the streets, for instance, produce noise, traffic jams, pollution (in the forms of leaflets and graffiti), and so on. Second, speech can have undesirable effects on members of the audience who change their behaviour: as a result of much campaigning, consumer behaviour can alter to make a particular brand of junk food go bankrupt, or some people might be convinced that a social group has certain undesirable features and change their attitudes towards its members accordingly<sup>4</sup>. Generally speaking, the first type of bystanders' interests is addressed through regulation regarding the 'time, place and manner of speech'. Governments minimise the noise and pollution of protestations by planning them in non-residential zones, scheduling them for weekends to keep traffic jams and commuter disruption to a minimum, loud music might be prohibited in the early hours, and so on<sup>5</sup>. The second kind of bystander interests is harder to minimise, and their protection is more controversial. As we shall see in this chapter, Thomas Scanlon produces an argument against restrictions of speech if they are justified on the basis that some people might be harmed because a speaker convinces an audience to act on certain reasons.

---

<sup>2</sup> Mill, *On Liberty and other Essays*, 43.

<sup>3</sup> Scanlon, 'Freedom of Expression and Categories of Expression' 92-3.

<sup>4</sup> An example of this are the existence of social and gender stereotypes. See my discussion of this topic below in section II, and also in Chapters 1, 2, 5 and 6.

<sup>5</sup> I discuss the difference between content-based and content-neutral regulations in Chapter 5.



Although often the interests of speakers and audiences converge, occasionally they might conflict: someone lies about someone else, thus harming his reputation, or some people are misled into believing that brand A is better than brand B, and so on. In cases such as these, we need to balance the interests at stake and judge how interests *overall* are best served<sup>6</sup>. There are some controversies about the relative weight of different interests. For instance, Larry Alexander believes that issues of free speech are primarily concerned with the interests of audiences<sup>7</sup>, while Ronald Dworkin believes that speakers' interests are enough to warrant its protection<sup>8</sup>.

In this chapter I consider a very influential argument produced by Thomas Scanlon<sup>9</sup>. His position mainly targets the interests of audiences in protecting speech<sup>10</sup>. Although Scanlon himself later revised and rejected this argument, it remains a classical autonomy-based defence of speech<sup>11</sup>. In this article, Scanlon defends speech by establishing 'more stringent constraints than the minimal principle of liberty [i.e. the harm principle] by positing that the government is barred certain reasons for prohibiting speech'<sup>12</sup>. In this respect, Scanlon's principle is a proper principle of free speech: it provides positive reasons for allowing agents to

---

<sup>6</sup> Sometimes, the strategy for defending speech gives a clear and straightforward answer: people who hold a minimalist view would discount the costs of speech as almost non-existent, or always redeemable. Maximalists would hold that the benefits of speech always trumps its costs, however big they are.

<sup>7</sup> Larry Alexander, *Is There a Right to Freedom of Expression?* (Cambridge: Cambridge University Press, 2004), 8.

<sup>8</sup> Dworkin, *Sovereign Virtue*, 160-1.

<sup>9</sup> Scanlon, 'A Theory of Freedom of Expression', 6-25.

<sup>10</sup> Notice, though, that in this argument Scanlon does not distinguish between the different interests involved.

<sup>11</sup> Cf. David Strauss, 'Persuasion, Autonomy and Freedom of Expression', *Columbia Law Review*, 91 (1991), 334-71; Thomas Nagel *Concealment and Exposure* (New York: Oxford University Press, 2002), 43.

<sup>12</sup> Greenawalt, *Speech, Crime, and the Uses of Language*, 9.

bring about harm to third parties through speech, harm which, had it been produced by other means would certainly warrant governmental regulation.

Section 1 of this chapter presents Scanlon's principle of free speech. His argument is divided into two parts: one attending to the role of responsibility of speakers in bringing about harm to third parties through acts of expression, and the other appealing to an account of the proper relation between autonomous citizens and governments. After the presentation of his 'Millian Principle', I will offer some criticisms that, presumably, led him to revise, and later reject, this principle<sup>13</sup>. These criticisms form sections 2 and 3. The former includes criticisms to Scanlon's first argument. The latter discusses his second argument and contrasts it with Rawls's account of citizens' interests. In the last section I conclude that audiences' autonomy can be better protected by restricting speech in certain ways.

## I

### The Millian Principle

Scanlon's Millian Principle holds that

There are certain harms which although they would not occur but for certain acts of expression, nonetheless cannot be taken as a part of a justification for legal restrictions on these acts. These acts are (a) harms to certain individuals which consist in their coming to have false beliefs as a result of those acts of expression; (b) harmful consequences of acts performed as a result of those acts of expression, when the connection between the acts of expression and the

---

<sup>13</sup> Scanlon, 'Freedom of Expression and Categories of Expression', 84-112; Thomas Scanlon, 'Content Regulation Reconsidered', in *The Difficulty of Tolerance* (New York: Cambridge University Press, 2003), 151-68.

subsequent harmful acts consist merely in the fact that the act of expression led the agents to believe (or increased their tendency to believe) those acts to be worth performing<sup>14</sup>.

The principle is based on two different, albeit interrelated and not clearly distinguished arguments<sup>15</sup>. One is based on a notion of agency and responsibility. The other is based on a contractualist theory.

### *The Discontinuous Responsibility Argument*

According to Scanlon, the reason speech should be protected beyond the harm principle is that there are actions whose harmful consequences are not imputable to those to whom the restriction of speech would apply. A person should not be prevented from communicating some thoughts because she is not responsible for the harm other people cause when motivated by those ideas. The first part of the Millian Principle, then, denies the possibility of restricting speech in cases of being misled by others' acts of expression, or acting upon those misleading reasons. The rationale is that my being an adult 'in full possession of [my] faculties'<sup>16</sup> implies that I have the actual capacity of rationally assessing the content of others' expression. Regarding myself as a rational autonomous agent requires that I bear full responsibility for my beliefs. And, at the same time, it seems that my being responsible for them means that other people cannot also be (even partly) so.

In Scanlon's view, responsibility for beliefs cannot be shared between different people. There seems to be a gap between the casual history of someone's

---

<sup>14</sup> Scanlon, 'A Theory of Freedom of Expression', 14.

<sup>15</sup> Brison, 'The Autonomy Defense of Free Speech', 328; Robert Amdur, 'Scanlon on Freedom of Expression', *Philosophy and Public Affairs*, 9 (1980), 209.

<sup>16</sup> Scanlon, 'A Theory of Freedom of Expression', 13.

beliefs and her responsibility for holding them. Scanlon's argument suggests that the only person responsible for one's beliefs is the agent herself, independently of the processes by which she came to possess them (their causal history). It is assumed that either the agent is responsible for her beliefs or somebody else is (namely, those whose acts of expression resulted in the agent's holding those beliefs). In this view, it is not possible for both to be responsible for them<sup>17</sup>. If one is to count as a rational autonomous agent (in the relevant sense) then she must be the only person responsible. The first part of the Millian Principle is a consequence of this notion of personal responsibility: the agent herself is the only one that can be held responsible for her beliefs and, moreover, she seems to be the only one responsible for the processes that led her to form such beliefs.

The second part of the Millian Principle addresses the harmful actions motivated by those beliefs communicated through acts of expression. Scanlon assumes as a baseline the difference between reasons and means for actions. Accordingly, underlying his theory of free speech is the difference 'between expression which moves others to act by pointing out what they take to be good reasons for action and expression which gives rise to action by others in other ways, e. g. by providing them with the means to do what they wanted to do anyway'<sup>18</sup>. This difference limits the realm of the principle of free speech insofar as it only protects reasons for actions, not means.

---

<sup>17</sup> Scanlon does not hold the bucket-theory of responsibility, according to which there is a fixed amount of responsibility to be distributed. If person A is held fully responsible for an act X, then there is no more responsibility available for person B. Cf. Nozick, *Anarchy, State, and Utopia*, 130. See also Chapter 1 in this thesis.

<sup>18</sup> Scanlon, 'A Theory of Freedom of Expression', 13.

It is important to note that Scanlon does not provide further explanation regarding how to understand the difference between means and reasons. Consider the following description of an action: 'I flip the switch, turn on the light, and illuminate the room. Unbeknownst to me I also alert a prowler to the fact that I am home'<sup>19</sup>. This example demonstrates a clear ambiguity in how we perceive means and ends in relation to descriptions of actions. Essentially, qualifying as an end or a means is not an absolute distinction, but a 'perspectival' one; it is description related.

In the traditional dichotomy, reasons and beliefs are instrumental to desires and intentions<sup>20</sup>. If Scanlon has this picture in mind, then the Millian Principle would only protect expression regarding the latter, leaving speech about reasons and beliefs unprotected. But it is unlikely that he would accept this consequence because he argues that reasons are primitive, they are not instrumental or reducible to other mental states<sup>21</sup>.

According to Scanlon, 'our normal views about legal responsibility'<sup>22</sup> suggest that individuals might be held responsible when they provide the means for committing, say, a crime, whilst they cannot be considered responsible just for providing reasons about the worthiness of such an act. According to Scanlon's argument, then, if I convince you that looting the British Museum is a good idea because the works of art contained there belong to everybody, I could not be held legally responsible for your crime. On the other hand if I (additionally or only) provide you with means (information about security plans, narcotics to poison the

---

<sup>19</sup> Davidson, *Essays on Actions and Events*, 4.

<sup>20</sup> Cf. Chapter 2 on classical rationality.

<sup>21</sup> Scanlon, *What We Owe to Each Other*, 7. Nevertheless, it is unclear whether he had reached this conclusion at the time he developed the Millian Principle.

<sup>22</sup> Scanlon, 'A Theory of Freedom of Expression', 13.

guards, and hidden passages to escape) then we could both be condemned for 'your' ('our'?) crime.

The argument relies on the assumption that reasons (but not means) are filtered, so to speak, by the autonomous agent's judgement. In a short, Scanlon claims that

a person who acts on reasons he has acquired from another's act of expression acts on what he has come to believe and has judged to be a sufficient basis for action. The contribution to the genesis of his action made by the act of expression is, so to speak, superseded by the agent's own judgement<sup>23</sup>.

This could be named the 'discontinuous responsibility argument'. It holds that although there is a harmful consequence which is causally connected to an act of expression, attribution of responsibility cannot follow the causal connection all the way back because it is mediated by the agent's judgement. Although Scanlon restricts this form of responsibility to free speech, its usage can be generalised to other realms. People might argue that the environmental harms produced by big SUVs are attributable only to consumers of those cars, and not to the producers. Similarly, corporations selling junk food should not be held responsible for the health damages suffered by people who eat their products, the consumers are responsible for their choices. Moreover, this notion of responsibility can be used to defend the right to buy and bear arms by claiming that gun providers are not

---

<sup>23</sup> Ibid. 13.

responsible for the uses their products are put (murder included); it is *only* the agent (who fires the gun) who is responsible for this.<sup>24</sup>

### *Citizen's Autonomy*

The second component of the Millian Principle appeals to a different aspect of autonomy. Whilst the first is based on an account of personal autonomy derived from the concepts of agency and responsibility, the second derives governmental constraints based on the view that citizens' autonomy is incompatible with the state's right to prohibit the expression of reasons. This argument claims that the relation between autonomous citizens and a legitimate government must be compatible with the view that citizens can recognise the government's authority 'while still regarding themselves as equal, autonomous, rational agents'<sup>25</sup>. It is motivated by a concern for the interests of audiences in speech. Whereas the argument in the previous section suggests that speakers cannot be held responsible for the harms bystanders might suffer from audiences, this one claims that the latter group have an autonomy-based interest in speech which warrants its protection.

Scanlon argues that individuals who regard themselves as equally autonomous, rational agents would not, under a hypothetical contract, grant the state the right to decide which kind of information should be restricted. According to Scanlon, giving the government the right to impose restrictions on speech (contrary to the Millian Principle) would be tantamount to waiving the presumption of autonomy, insofar as 'an autonomous person cannot accept without independent

---

<sup>24</sup> I owe this last example to Hurley, 'The Public Ecology of Responsibility'.

<sup>25</sup> Scanlon, 'A Theory of Freedom of Expression', 14-5.

consideration the judgement of others as to what he should believe or what he should do',<sup>26</sup>.

There are, however, circumstances under which citizens do bequeath responsibility for regulating what they can express to the state. In a trivial case, individuals are rightly prevented from shouting 'fire' in a theatre. According to Scanlon, such circumstances are cases of permissible paternalism. Without digressing into his discussion of this subject, he argues that there are exceptions to the Millian Principle; namely when its grounds are undermined. When audiences lack the capacity of judgement (and therefore are presumably not autonomous) restrictions on expression are acceptable. As the example of shouting 'fire' demonstrates, the justification is that individuals would not react in a rational, judicious way. In such situations acts of expression seem not to be 'superseded by the agent's own judgement',<sup>27</sup>. This example is highly illustrative. Forms of expression that do not fall within the remit of autonomy are not considered 'speech' in the technical sense necessary for a free speech principle to be invoked. They become subject to the same regulations of the harm principle. Shouting 'fire' in a theatre is not a protected act of speech because it is not covered by autonomy; and it is not so because people tend to react non-autonomously. According to Scanlon, in rare cases of foreseen irrationality (presumably, when expression will not be judged with independent consideration) citizens would form 'voluntary agreements',<sup>28</sup>. These are examples of what Elster calls 'Ulysses strategies': individuals sometimes

---

<sup>26</sup> Ibid. 16.

<sup>27</sup> Ibid. 13.

<sup>28</sup> Ibid. 19



decide rationally to protect themselves from their own future irrationalities by limiting the set of feasible choices open to them<sup>29</sup>.

Even though citizens can occasionally employ such strategies, they would not, according to Scanlon, grant to the state the capacity of deciding what they ought to believe or hear as a part of its normal powers<sup>30</sup>. In fact the Millian Principle demands such agreements are to be reached only in very special cases. In defence of his principle Scanlon asks

could an autonomous individual regard the state as having, not as part of a special voluntary agreement with him, but as part of its normal powers qua state, the power to put such an arrangement [restrictions on speech] into effect without his consent whenever it (i.e. the legislative authority) judged to be advisable?<sup>31</sup>.

Clearly, Scanlon's answer is no. However it is important to highlight a distinction Scanlon appears not to have noticed. It seems clear that Scanlon's argument applies to considerations about the *content* of acts of expression, but it might be problematic regarding the *processes* triggered by them.

Scanlon's answer can be explained recalling his principle of contractualism:

An act is wrong if its performance under the circumstances would be disallowed by any system of rules for the general regulation of behaviour which no one could reasonably reject as a basis for informed, unforced general agreement<sup>32</sup>.

---

<sup>29</sup> Elster, *Ulysses and the Sirens*.

<sup>30</sup> Scanlon, 'A Theory of Freedom of Expression', 19.

<sup>31</sup> *Ibid.* 19.

<sup>32</sup> Scanlon, 'Content-Based Regulation Reconsidered', 132. Definitions of contractualism appear slightly modified several times in his later book *What We Owe to Each Other*, 170, 189.

Regarding the content of 'what to believe', Scanlon suggests that a principle that requires citizens to forgo the right to decide to which acts of expression they wish be exposed would be reasonably rejected. It would not be reasonable for citizens to waive the right to make independent judgments by weighing one's reasons and reaching one's conclusions in favour of the view that the state's judgments are always correct. There are exceptions to this general proviso, notably circumstances under which paternalism is legitimate. The argument regarding *processes* of judgment, decision-making and preference formation, however, might be different. It seems that intervention would not be reasonably rejected when addressing well-known cognitive defects of the sort discussed in the previous chapter. As we have seen, people suffer from biasing influences, such the halo effect, their judgments are distorted by framing, they get anchored with irrelevant information, they tend to believe propositions they know to be false, concepts can be activated triggering different automatic responses, and so on. If this is accepted, then it can be said that we are concerned here with a relatively weak form of paternalism insofar as it operates for the individual's interests (namely helping them to achieve independent judgment) by neutralising distorting processes and influences that function outside their control<sup>33</sup>. It is important to notice that although someone might reject the principle of weak paternalism, nevertheless the distinction between content and processes still holds, and they must be treated as different (but interrelated) questions.

Scanlon's view of rationality is in stark contrast to the ecological version described in the last chapter. In particular, individuals are *prima facie* autonomous

---

<sup>33</sup> Robert Goodin, *No Smoking: The Ethical Issues* (Chicago: University of Chicago Press, 1989), 23.

and rational, and therefore a legitimate government faces autonomy as a constraint on its powers. This account is not concerned with the social ecology of autonomy. Essentially, autonomy and rationality are considered to be a *status* that does not allow degrees. The Millian Principle was defended, Scanlon writes,

by showing it to be a consequence of a particular idea about the limits of legitimate political authority: namely that the legitimate powers of government are limited to those that can be defended on grounds compatible with the autonomy of its citizens- compatible, that is, with the idea that each citizen is sovereign in deciding what to believe and in weighing reasons for actions<sup>34</sup>.

## II

### Criticisms of the Millian Principle

This section focuses on three points. First, I show that the Millian Principle has some unreasonable consequences regarding paternalistic policies. Second, I discuss whether the distinction between means and reasons upon which Scanlon's principle is based, is tenable. Third, I assess the notion of autonomy assumed by the Millian Principle and conclude that it is not appealing.

#### *Paternalism*

Recall the argument for permissible paternalism that the Millian Principle allows: it states that only in cases of 'extremely brief'<sup>35</sup> severe diminished rationality or

---

<sup>34</sup> Scanlon, 'Freedom of Expression and Categories of Expression', 95.

<sup>35</sup> Scanlon, 'A Theory of Freedom of Expression', 20.

judgment can a government legitimately place restrictions on expression. From a citizen's point of view, paternalistic restrictions are legitimate only when she can foresee circumstances in which her agency is extremely diminished. In contrast, we sometimes restrict other forms of liberty while retaining the presumption of rationality: for instance, compulsory use of helmets and seat belts, restrictions on tobacco smoking and laws against false advertisement. Regardless of the controversies that such policies involve, generally they do not suppose that people cease to regard themselves as rational agents.

Two kinds of paternalism can be distinguished: a weak and a strong version. The former 'works within the individual's own theory of the good and merely imposes upon him a better means of achieving what after all are only his own ends'<sup>36</sup>. The latter, by contrast, works from outside his theory of the good, imposing upon him both, means and ends. As an example, think about a state in which the practice of religion is compulsory (atheists are obliged to attend religious services and lead a religious life), defended on the basis that people's souls must be saved from eternal damnation. This policy can be reasonably rejected by those who do not hold a religious viewpoint<sup>37</sup> as it cannot be accommodated within their own conception of the good. Notice that this policy might also be unacceptable to a religious person, on grounds that a religious life requires free endorsement, and it is incompatible with coercion<sup>38</sup>. This is an extreme case. An example of weaker paternalist policies is the compulsory use of seat belts in cars. We assume that it is in people's interests to be safe while on moving vehicles, and that sometimes, due to.

---

<sup>36</sup> Goodin, *No Smoking*, 23.

<sup>37</sup> It might also be unacceptable to those who hold one. Cf. Dworkin, *Is Democracy Possible Here?*, Chapter 3.

<sup>38</sup> Thanks to Andrew Walton for this point.

say, weakness of the will or laziness they do not fasten them. Mandatory seat belt legislation generates extra incentives overriding the momentary inconvenience of fastening them, thus making people more likely to do it<sup>39</sup>. The reason for accepting these regulations is that, in general, it is in people's own interest to be safe and that reasonable conceptions of the good value personal safety<sup>40</sup>. Notice, though, that we can accept these regulations without giving up the presumption of rationality; surely there is a big difference between laziness and weakness of the will and severe diminished mental capacities. Not all forms of paternalism need to be justified from a first person perspective. Some cases of acceptable paternalism might be justified from the third person. Rawls argues that when we do not know enough about people's settled preferences and interests, but we know that some forms of irrationality might occur (they need not be as acute as the Millian Principle requires, though), paternalistic policies should be guided by the principles of justice and an account of primary goods. This case is clear in the example of children, where we act on their behalf even though we do not follow their own conception of the good (because we do not think that they satisfy a threshold of autonomy)<sup>41</sup>.

The Millian Principle proposes restrictions that are too strong for paternalistic policies regarding expression. Take the case of the regulation of misleading or false advertisement, for example. Imagine an ad which says that a low-fat cheese has only 3% fat, while in fact it has 25%. We can accept laws forbidding this ad without endorsing the view that people are *severely* rationally diminished

---

<sup>39</sup> Cf. Lawrence Lessig, 'The Regulation of Social Meaning'. *The University of Chicago Law Review*, 62 (1995), 943-1045. He notes that governments can change the social meaning of some paternalistic policies making them look not paternalistic.

<sup>40</sup> Of course this is not the only manner in which compulsory seat belt use might be defended. One can appeal to the costs that accidents have on the NHS, for instance.

<sup>41</sup> Rawls, *A Theory of Justice*, 218-20.

when tempted to buy cheese. We can justify it on the grounds that the costs of deciding for oneself (reading independent consumers' magazines, consulting scientific research reports, etc) are too high in comparison with the costs of allowing the state to intervene on this issue. Assuming that it is in people's interest to be exposed to as much information as possible (even false or misleading information) seems to embrace a mistaken notion of our mental capacities: we do not have unlimited mental resources to judge every piece of information when making decisions on every issue. Instead, it seems reasonable to grant the government the power to filter out information which would not help us pursue our own ends.

There is a way to accommodate this worry within the Millian Principle, for Scanlon could argue that regulations on false advertisement can be accepted within if they concern only means for doing what people want to do (say, getting low-fat food), but not reasons (say, having a healthier life). As we will see in the next section this distinction is untenable. Assuming temporarily that it is viable, Scanlon still encounters a problem as the distinction between means and ends does not cover all cases in which we would accept regulations on advertisement. Think, for instance, of tobacco ads, which are regulated even though people need not regard their rational capacities as severely diminished. This example conflicts with to the Millian Principle<sup>42</sup>.

According to Scanlon, any plausible theory of paternalism must take into account a concern for the importance of choice, or, in his own words, the 'value of

---

<sup>42</sup> Although, again some might retort that tobacco ads concern only means, rather than reasons to something else, say getting pleasure in life.

choice'<sup>43</sup>. Choice, Scanlon argues, is valuable in a number of senses. Instrumentally, we want to choose the means by which we attain our ends. Moreover, sometimes we want to be able to choose certain things not because of the outcome of the choice itself but rather because what matters to us is the act of choosing. The outcome of the choice might have a different meaning if chosen by me than if by somebody else. Furthermore, choice might be valuable in a symbolic sense; 'in a situation in which people are normally expected to make choices of a certain sort for themselves, individuals have reason to value the opportunity to make these choices because not having or not exercising this opportunity would be seen as reflecting their judgment (their own or someone else's) that they are not competent or do not have standing normally accorded an adult member of society'<sup>44</sup>. The significance of choice and the relative weight that each of the above senses holds, varies according to circumstances, such as the agent's knowledge and self control, his aims and attachments, and the groups and society in which he lives<sup>45</sup>.

It is important to see how narrowly the Millian Principle construes the importance of choice, for it suggests citizens are only interested in *not being interfered* with when forming beliefs. On this reading it overlooks the various different senses in which choice is meaningful. Regarding free speech, then, we need a nuanced account of the relation between the interests that speech serves and the meaningfulness of choices. This account is necessary in order to understand the difference between bans on junk food, alcohol or tobacco advertisements on the one

---

<sup>43</sup> Scanlon defines the 'value of choice' by saying that 'what matters is the value of the opportunity to choose that the person is presented with'. See Scanlon, *What We Owe to Each Other*, 258.

<sup>44</sup> Ibid. 253.

<sup>45</sup> This worry opens the door to the idea that regulations on speech must be sensitive to different categories, and the way they relate to different the interests people have on speech. Cf. Scanlon, 'Freedom of Expression and Categories of Expression' and 'Content Regulation Reconsidered'.

hand, and bans on religious, political or artistic speech on the other. Unfortunately the Millian Principle offers no room for such concerns because it sets the conditions for justification too strictly.

### *Reasons and Means*

As we noted in the first section of this chapter, Scanlon thinks that the difference between reasons and means is central to understanding the remit of a principle of free speech. In his words, the important distinction is 'between expression which moves others to act by pointing out what they take to be good *reasons* for actions [on the one hand] and [on the other] expression that gives rise to action by others in other ways, e.g. by providing them with the *means* to do what they wanted to do anyway'<sup>46</sup>.

According to this distinction, arguments for restricting speech are illegitimate if they are focused on the 'reasons' for actions, but could be acceptable if focused on the 'means' for actions. This difference, Scanlon argues, is based on 'our normal views about legal responsibility'<sup>47</sup>.

As a general structure, the difference between means and ends seems to be less secure than Scanlon thinks. For instance, last night before going to bed I went to the kitchen, turned the light on, poured a glass of water, and scared a drunken man who was walking past my window. It seems that, with the exception of frightening the man, all the things I did were means to a further end (namely, to quench my thirst). But on a closer inspection we could also say that I went to the kitchen in order to turn on the light, which, in turn, was only a means to pouring the water from

---

<sup>46</sup> Scanlon, 'A Theory of Freedom of Expression', 13. Emphasis added.

<sup>47</sup> Ibid. 13.



the tap. We can describe my action in several ways, and the features we characterise as 'ends' or 'means' are always relative to the description itself. The distinction is perspective-dependent rather than absolute.

Another feature of the supposed means/ends dichotomy is that it is relative to other ends and other means. Recall the example discussed above of looting the British Museum. There the action's goal was to loot the museum because the pieces of art, supposedly, belong to everybody and the action's means were the poison recipe for the guards and the plans of the museum. Imagine, however, that the looting is only a means towards another goal, say, to improve global justice by recovering everybody's belongings (looting all of the world's museums). According to the Millian Principle, under the first description the reasons for the looting are protected, under the second description they are not.

Moreover, the generic 'means-ends' structure requires an ultimate goal towards which all other actions are instrumental. If individuals had such an ultimate end, then the Millian Principle would only protect expression related to that goal (all other reasons would be just means to it). But this is a problematic supposition. In general individuals have a plurality of ends and means, and they are not arranged in a perfectly consistent, transitive order<sup>48</sup>.

There is also a different approach by which the means/goals distinction, which the Millian Principle depends upon, can be undermined. Means and goals are related psychologically through automatic processing that bypasses autonomous

---

<sup>48</sup> For problems of this kind of presuppositions see criticisms to the Rational Choice Theory. Especially Elster, 'The Nature and Scope of Rational Choice Explanation'; Jon Elster, *Rational Choice Theory* (Oxford: Blackwell, 1986); Jon Elster, 'Some Unresolved Problems in the Theory of Rational Behaviour' *Acta Sociologica*, 36 (1993), 179-90; Hargreaves Heap, et. al. *The Theory of Choice*; Amartya Sen, 'Rational Fools: A Critique of the Behavioural Foundations of Economic Theory' in his *Choice, Welfare and Measurement* (Oxford: Blackwell, 1982). See also Charles Taylor, 'The Diversity of Goods' in his *Philosophical Papers*.

control. The idea, then, is that 'stimuli in the environment can directly activate a goal, which will then become operative and guide cognitive and behavioural processes within that environment, without any need for conscious decision'<sup>49</sup>. In particular, an active goal automatically triggers mental representations usually associated with attaining that goal, making them more accessible. Searching for something to eat activates concepts related to that end, such as restaurants and bodegas<sup>50</sup>. The relation also works in the opposite direction: an active concept that is usually a means towards an end can activate that goal. In an experiment, people primed with words related to success strived for longer to finish a task than people who were not primed<sup>51</sup>. Moreover, the source of activation is irrelevant to the goal activated; it does not matter whether the stimuli is conscious or not, goal-activation is likely to occur automatically<sup>52</sup>.

The Millian Principle is able to accommodate restrictions on speech only if those restrictions concern means to do what people have reason to do. It cannot manage a concern with people's goals. This restriction is untenable for two reasons. First, conceptually the means/goals distinction is not tenable because both components are description-dependent. Second, psychologically, means and ends are linked automatically, insofar as expression regarding one might activate the other directly, bypassing autonomous control.

### *Autonomy and Responsibility*

---

<sup>49</sup> Bargh, 'The Automaticity of Everyday Life', 30.

<sup>50</sup> John Bargh, 'What Have We Been Priming all these Years? On the Development, Mechanisms, and Ecology of Nonconscious Social Behavior', *European Journal of Social Psychology*, 36 (2006), 158.

<sup>51</sup> Bargh, 'The Automaticity of Everyday Life', 39.

<sup>52</sup> *Ibid.* 35.

Scanlon holds that his Millian Principle is supported 'by our normal views about legal responsibility'<sup>53</sup>. However, there is an oddity here as some of our moral intuitions seem to point in the opposite direction. It is not sufficient to demonstrate that our normal views of *legal* responsibility accord with Scanlon's theory, it is also necessary to show that the latter is consistent with our views on *moral* responsibility. At the very least, Scanlon must show that they do not conflict, or, if they do, which set of intuitions has priority.

Imagine these two examples. In the first, Owen convinces Carol to murder the president of the company for which she works. He does so using only rational, valid arguments. Carol is an autonomous lady, capable of rational judgment and very much able to balance reasons for action, critically comparing beliefs and so on. After a long discussion, Carol is convinced by Owen to kill the boss and some hours later she cuts his throat. According to Scanlon, we have no reasons to hold Owen responsible for the harm done by Carol. The second example is quite similar, but, here, Owen uses fallacious and deceptive arguments. He frames the arguments differently in order to exploit some of her cognitive biases and failures. Perhaps he appeals to emotional features or shows her violent movies<sup>54</sup>, or manipulates her in other ways.

How would the Millian Principle treat both cases? Remember that the second part of the Millian Principle holds that:

harmful consequences of acts performed as a result of those acts of expression, when 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

---

<sup>53</sup> Scanlon, 'A Theory of Freedom of Expression'. 13.

<sup>54</sup> Cf. Hurley, 'Bypassing Conscious Control'

believe (or *increased their tendency* to believe) those acts to be worth performing<sup>55</sup>.

The Millian Principle does not assess the manner in which acts of expression lead the agents to believe (or increases their tendencies to believe) in the worthiness of the expressed reasons. In this sense, for the Millian Principle the examples above are similar. It does not include a qualitative criterion to discern between autonomous or non-autonomous tendencies and influences. If this is tenable, then the reason for protecting speech loses its strength. Remember that a principle of free speech poses more stringent conditions than the harm principle; it claims that, even if speech brings about harm, the latter cannot be used as a reason to regulate expression because audiences are autonomous in respect to those acts of expression. An agent's ability to judge speech independently is the basis for the special free speech principle. However, if speech bypasses that ability (through mental contamination), then autonomy cannot be used as an exception to the harm principle<sup>56</sup>. Scanlon's failure to make the distinction between autonomous and non-autonomous tendencies of belief and preference formation render his Millian Principle an inadequate defence of free speech.

Consider, for instance, some forms of mental contamination associated with social stereotypes. Stereotyping is an automatic process by which traits that purportedly describe social groups are integrated<sup>57</sup>. Stereotypes seem to occupy a similar psychological function to schemas and categories; they help to us to

---

<sup>55</sup> Scanlon, 'A Theory of Freedom of Expression', 14. Emphasis added.

<sup>56</sup> Hurley, 'Bypassing Conscious Control', 321.

<sup>57</sup> Dijksterhuis and Bargh, 'The Perception-Behavior Expressway', 9.

complete missing information about a person or event and also to 'generate expectancies about what is going to happen next'<sup>58</sup>. As such, they form an important component of implicit knowledge furnishing people with more efficient information about groups. A beneficial consequence of stereotypes and categories is that they allow people to devote cognitive resources to other activities. However, there is a cost to this efficiency. In most cases, social stereotypes are inaccurate: they integrate concepts which have little to do with the group in question<sup>59</sup>, or, in some cases, represent merely the result of prejudice. The content of racial and gender stereotypes are well known in Western societies. For instance, traits associated with black people are poverty, aggressiveness, criminality, low intelligence, lack of education, laziness, sexual perversity, athleticism and rhythm among others<sup>60</sup>. People with both high and low levels of prejudice highlighted some of these traits, and all singled out aggressiveness<sup>61</sup>. For example, a classical experiment described by Devine shows how stereotypes are applied to assess situations. People were shown a short film in which a man behaved in an 'ambiguously aggressive way'. After the film they had to complete a report qualifying the man's behaviour from 'very aggressive' to 'not aggressive'. Not surprisingly, when the role was played by a black actor people tended to interpret his behaviour as being much more aggressive than when it was played a white actor<sup>62</sup>.

---

<sup>58</sup> Mark Chen and John Bargh, 'Nonconscious Behavioral Confirmation Processes: The Self-Fulfilling Consequences of Automatic Stereotype Activation', *Journal of Experimental Social Psychology*, 33 (1997), 541.

<sup>59</sup> Ibid. 542.

<sup>60</sup> These traits were the most prominent among 40 white students who participated in the experiments reported in Devine, 'Stereotypes and Prejudice', table 1. 8.

<sup>61</sup> The level of prejudice is measured according to the Modern Racism Scale. *ibid.* 7.

<sup>62</sup> Ibid. 9.

As with other concepts, social stereotypes can be automatically activated in a variety of ways. For instance, recall the experiments in which Bargh and others primed students with concepts related to the stereotype of the elderly. Participants were primed with concepts such as 'Florida', 'Bingo', 'Grey', 'Retired', and so on. After the experiment, those who were primed walked along the corridor slower than those who were not. The concepts associated with a stereotype also automatically activated other related traits<sup>63</sup>. In a subsequent experiment, Bargh and his colleagues decided to prime participants with the 'African American' stereotype. In this case the students were asked to perform a tedious and boring task on a computer. Without knowing, they were subliminally shown pictures of African American faces (the pictures were displayed just for a fraction of a second so the participants were not aware of witnessing them at all). After some time performing the task a message appeared in the computer screen telling the students that the computer had broken down. The experimenter then told the participants that they had to begin the task again. The result shows that people who were primed pictures of black people reacted more aggressively to the request than those who were not primed. This experiment is another example of the perception-behaviour link discussed in Chapters 1 and 2: a perception associated with aggression directly increases an individual's tendency to behave aggressively<sup>64</sup>.

Moreover, stereotypes can also generate self-fulfilling prophecies: imagine a person who interacts with a member of a stereotyped group. The first person has certain expectations about the second person's behaviour. Given these expectations, he behaves towards her as if the stereotype were true. Then she might respond in a

---

<sup>63</sup> Bargh, Chen and Burrows, 'Automaticity of Social Behavior', 236.

<sup>64</sup> Ibid. 238-9.

consistent way with his behaviour, thus confirming the original stereotype<sup>65</sup>. Notice that the personal expectations need not to be conscious. John Bargh and Mark Chen argue that automatic stereotype activation can produce self-fulfilling prophecies of the same kind as those produced when the expectations are conscious. In this experiment, they combined two paradigms: stereotype activation and behavioural confirmation. Pairs of participants played a game in which one had to get the other to guess each of several target words. This game tends to create some frustration in both players (one because he cannot understand the other's ineptitude at representing words, the other at the inability of his partner to guess the words given such good illustrations). In this experiment, the participants were in different rooms and communicated via headphones and microphones. Prior to playing the game, one member of the pair was primed with pictures of young black persons (as in the experiment described above). One member of the pairs was randomly assigned the role of 'perceiver' and the other the role of 'target'. Only half of the perceivers were primed with the black faces, the other half and all the targets were primed with white faces. After the game, each participant judged their partner on a variety of traits, including aggressiveness. Independent judges (unaware of the goals of the experiment) also listened to the audiotapes in a random order, and were asked to evaluate the person on hostility (among other traits). The 'audiotapes ratings showed that [...] both the perceiver and the target participants in the stereotype-primed condition were rated as being more hostile than their counterparts in the no-priming condition'. This confirms that subliminal priming automatically increased hostility. 'Finally, [...] perceiver participants primed with subliminal black faces rated their

---

<sup>65</sup> Chen and Bargh, 'Nonconscious Behavioral Confirmation Processes', 542.

interaction partners as being more hostile, compared to nonprimed perceivers' ratings of their partners'. This remark suggests that stereotypic confirmation occurred<sup>66</sup>. Social stereotypes have costly effects not only to the stereotyped targets (costs such as discrimination in employment opportunities and the elicitation of self-fulfilling prophecies<sup>67</sup>), but also for the stereotype holders. The existence of these sorts of social stereotypes presents a strong form of mental contamination for many people who reject the fact that they are victims to those processes (though not for those who endorse the contents of the stereotype).

### *Responsibility for What?*

The relation between acts of expression and resultant harmful consequences might also be problematic in another way. If we read again the example involving Owen and Carol, it is clear that Carol is responsible for the murder. It is also obvious that Owen is responsible for convincing her into that act. Nevertheless, Scanlon would argue that Owen cannot be held responsible. But it is unclear what Owen is not responsible for or why this is so. If, after committing the murder, Carol said 'I am not responsible, Owen made me do it' we would not be convinced. But, the relevant question here is if, after the murder, Owen said 'Carol is a rational adult, she made her own decision, and my hands are clean', would we be any more convinced?<sup>68</sup>. The important issue is not the positive claim that Owen could be held (partly) legally responsible for the crime, but the negative claim that he could not. This point returns

---

<sup>66</sup> John Bargh, 'The Cognitive Monster: The Case against the Controllability of Automatic Stereotype Effects', in Chaiken, S., and Y. Trope (eds.) *Dual-Process Theories in Social Psychology* (New York: Guildford Press, 1999), 374-5.

<sup>67</sup> Christian S. Wheeler, W. Blair G. Jarvis and Richard Petty, 'Think Unto Others: The Self-Destructive Impact of Negative Racial Stereotypes', *Journal of Experimental Social Psychology*, 37 (2001), 174.

<sup>68</sup> Cf. Amdur, 'Scanlon on Freedom of Expression', 296.



to the discontinuous responsibility argument, according to which, people are not responsible for the harm other people do based on the reasons communicated by the former. As I claimed above, if the harms caused by speech operate non-autonomously, the discontinuous responsibility argument falters.

The Millian Principle leaves another possibility open. It might be that we do not think Owen is responsible for Carol's murder. However that does not mean that we think that Owen is free from *any* wrong doing. In particular, he is guilty of convincing Carol to commit a murder. An appealing notion of responsibility needs to be sensitive to the fashion in which people acquire and respond to the reasons they act upon.

Elsewhere Scanlon distinguishes between two different senses of responsibility: 'substantive responsibility' and 'responsibility as attributability'<sup>69</sup>. According to the latter, being responsible means to be 'properly subject to praise or blame for having acted in such way', while, according to the former, being responsible is to say 'that the person cannot complain of the burdens or obligations that result' from her actions<sup>70</sup>. This difference implies, among other things, that, on occasions, one can be subject to blame for doing something wrong without bearing the consequences of such actions. For instance, despite the fact that drug addicts are subject to blame for their addiction, it does not follow that they are properly burdened with (all) the costs those addictions generate. Society still has a duty to help them with rehabilitation treatments, and unemployment benefits (if required) and so on.

---

<sup>69</sup> Scanlon, *What We Owe to Each Other*, 248-9,

<sup>70</sup> *Ibid.* 290.

Using this distinction, it seems that the discontinuous responsibility argument can be rephrased as follows: An individual X should not be held substantively responsible for the harms done by other people who act based on reasons expressed by X (note though that X may still be subject to blame). It seems that there is an ambiguity in this argument; from the claim that Owen cannot be held responsible for a murder, it does not follow that he should be free of any burdens.

It may be we consider that Owen is properly held substantively responsible for some wrongdoing (in this case, of convincing someone to murder her boss), and thus we might impose upon him some burden. Imagine a similar example: an inebriated man arrives at an off-license asking for a bottle of cognac. He gets the bottle, and after drinking half of it while driving knocks down a couple that are leaving the theatre. Although the shopkeeper cannot be held responsible for the murder of the theatregoers, it may be the case that he is held responsible for some wrongdoing, and accordingly we might agree to impose some burdens upon him (maybe a fine, a caution, or a licence cancellation)<sup>71</sup>. Returning to the Millian principle, then, the point of import is not whether Owen can be properly charged with murder, but why he should be left free of *any* burdens.

The discontinuous responsibility argument links (or more specifically, fails to link) speakers with bypassers<sup>72</sup> through audiences' autonomy. However, recall that features in the environment can trigger automatic responses that are beyond people's control. It seems that insofar as speakers participate in the social environment's

---

<sup>71</sup> Scanlon seems to suggest that the difference between legal and moral responsibility arises at this point. According to him, our intuitions about legal responsibility point towards the discontinuous responsibility argument. In contrast it may well be that we do not hold people responsible for wrongdoings based on pragmatic considerations such as efficacy or costs. The point then is that we would hold them *prima facie* responsible, but there are practical considerations that weigh more.

<sup>72</sup> Bypassers described as those who suffer the consequences of the change of attitudes that audiences experiment in virtue of speakers' acts of expression.

construction they are in a situation in which they could be held responsible for the ecology of those environments.

Following Amdur, we can also claim that, if the responsibility basis for the Millian Principle was adequate, then the principle would prove too much. If our normal views about responsibility were as Scanlon argues, then it is not clear why the government can legitimately restrict expression of means for actions at all. Remember that the means/goals distinction underlies the Millian Principle. It is difficult to see why Scanlon thinks that autonomy implies the capacity for weighing reasons but not means. The Millian Principle holds that when Owen argues about the worthiness of murdering the boss Carol's judgement supersedes the causal connection between her beliefs and Owen's acts of expression. However, it is not clear why when Owen gives Carol the recipe for the poison which will be used her judgement does not supersede Owen's act of expression as well<sup>73</sup>. What in the expression of reasons differentiates them so starkly from the expression of means? Remember that the difference between them is description dependent and, therefore, not absolute. If Amdur is right in this suggestion, then Scanlon should not accept this consequence of his principle: if the distinction between goals and means is not as Scanlon suggests then governments would need to protect every form of speech insofar as every expression would be superseded by an agents' rational judgement.

### *Conceptions of Autonomy*

The last point I wish to make regarding the conception of autonomy implied by the Millian Principle is that its initial appeal is due to a different conception of

---

<sup>73</sup> Amdur, 'Scanlon on Freedom of Expression'. 298.

autonomy, namely one that values independent judgment as a good to be promoted<sup>74</sup>. According to the Millian Principle, autonomy is only a constraint on governmental action which protects people from being interfered with in forming beliefs. There is a proviso which applies to a few special cases (such as individuals with severely rationally diminished capacities) and to special circumstances (brief moments of panic in crowded theatres). As discussed in the previous chapter, for autonomy to be achieved we require some criteria to assess the quality of critical judgment and a set of principles to neutralise some of these challenges (such as mental contamination). The problem with the Millian Principle is that it fails to promote the autonomy-related interests citizens have in speech, and therefore it cannot be a justification for its protection.

To sum up the problems found in the notion of personal autonomy assumed by Scanlon in this article, let me highlight to three key difficulties that have been discussed in this section. First, autonomy's appeal is due to a more basic conception of 'self-governance' or 'critical deliberation', which has not been defined by Scanlon. Second, as a result, the notion employed lacks the qualitative criteria necessary for assessing tendencies and processes of belief formation and decision making, so we cannot adequately distinguish between rational, irrational and a-rational processes. Third, if this notion were satisfactory then, as Amdur argues, it would prove too much, insofar as we could argue that every act of expression could be superseded by the agent's judgement, making any restriction on speech unacceptable.

---

<sup>74</sup> Scanlon, (2003) 'Freedom of Expression and Categories of Expression', 97.

## III

## The Millian Principle, Contractualism and Audiences' Interests

In this section I shall discuss the part of Scanlon's argument which relates to citizens' autonomy. In order to do so, I will briefly sketch John Rawls's idea of moral persons choosing principles of justice behind a veil of ignorance, and ask whether it would be reasonable for the parties to choose the Millian Principle.

Scanlon thinks that citizens would not surrender the right to decide for themselves to which information they are willing to be exposed. Although they would allow that in exceptional circumstances, the state could regulate speech (shouting 'fire' in a theatre), it would not be within its normal powers. Scanlon does not provide great detail in his argument, making it difficult to follow his reasoning. Nevertheless, I will sketch why I think his conclusion is not convincing based on a Rawlsian framework.

*Rawls's Conception of the Person*

Rawls thinks that the social role of a political conception of justice is to provide a scheme of mutually acceptable shared institutions, which must be based upon publicly recognised principles. In order to establish the principles of justice governing the basic structure of society Rawls begins from the 'basic model-conceptions'<sup>75</sup>: the concept of well-ordered society, the concept of moral person, and the intermediate model of the original position.

---

<sup>75</sup> John Rawls, 'Kantian Constructivism in Moral Theory' in his *Collected Papers*, 308.

A society is 'well-ordered' if its basic institutions satisfy the requirements of a political conception of justice, there is public knowledge that this is the case, people accept the conception and principles of justice, they believe their society to be just, and are motivated to act according to its rules<sup>76</sup>.

In a well-ordered society citizens are seen as 'free and equal moral persons'<sup>77</sup>. This conception of the person underlines Rawls's justice as fairness, but it is also important in the Millian Principle (if the conception is not exactly the same, the main features are shared),

The members of a well ordered society are moral persons in that once they have reached the age of reason, each has, and views the others as having, an effective sense of justice, as well as an understanding of a conception of their good. Citizens are equal in that they regard one another as having an equal right to determine, and to assess upon due reflection, the first principles of justice [...]. Finally, the members of a well-ordered society are entitled to make claims on the design of their common institutions in the name of their own fundamental aims and highest-order interests<sup>78</sup>.

This conception of the person includes two moral powers. First, persons have the capacity for a sense of justice: 'the capacity to understand, to apply, and to act from (not merely in accordance with) the principles of political justice that specify the fair terms of social cooperation'<sup>79</sup>. This sense must be effective in securing citizens' motivation to act for reasons of justice. The second moral power is 'the capacity to have, to revise, and rationally, to pursue a conception of the good. Such a conception

---

<sup>76</sup> Rawls, *Justice as Fairness*, 5-6, Samuel Freeman, *Rawls* (Oxford: Routledge, 2007), 242.

<sup>77</sup> Notice that the conception of the person described by Rawls here is a political conception of citizens.

<sup>78</sup> Rawls, 'Kantian Constructivism in Moral Theory', 309.

<sup>79</sup> Rawls, *Justice as Fairness*, 18-9.

is an ordered family of final ends and aims which specifies a person's conception of what is of value in human life'<sup>80</sup>. Citizens have two highest-order interests in exercising these moral powers, which are 'supremely regulative as well as effective'<sup>81</sup>.

In addition to these two interests, Rawls assumes that individuals are 'full developed moral persons', entailing that they have, at any given time, 'a determinate scheme of final ends, a particular conception of the good'<sup>82</sup>. Consequently, they have a higher-order interest in pursuing and advancing their own conception of the good. It is important to notice, though, that the latter interest is subordinated to the former set.

Citizens are equal 'in that they regard one another as having an equal right to determine, and to assess upon due reflection the first principles of justice by which the basic structure of their society is to be governed'<sup>83</sup>.

Finally, citizens are free in two respects. First, 'they conceive of themselves and of one another as having the moral power to have a conception of the good'<sup>84</sup>. This implies that they are not tied to particular conceptions, but are able to revise and change their values and final aims. 'As free persons, citizens claim the right to view their persons as independent from and not identified with any particular conception of the good, or scheme of final ends'<sup>85</sup>. The second sense in which citizens are free is that they are regarded as 'being self-authenticating sources of valid claims'<sup>86</sup>, that is, they are entitled to make claims to social institutions in order to pursue their

---

<sup>80</sup> Ibid. 19.

<sup>81</sup> Rawls, 'Kantian Constructivism in Moral Theory', 309.

<sup>82</sup> Ibid. 313.

<sup>83</sup> Ibid. 309.

<sup>84</sup> Rawls, *Justice as Fairness*, 21.

<sup>85</sup> Ibid. 21

<sup>86</sup> Ibid. 23.

conceptions of the good (provided those conceptions are permitted by the political conception of justice)<sup>87</sup>.

Another feature of the political conception of the person is that persons are reasonable<sup>88</sup>. According to Rawls,

reasonable persons are ready to propose, or to acknowledge when proposed by others, the principles needed to specify what can be seen by all as fair terms of cooperation. Reasonable persons also understand that they are to honour these principles, even at the expense of their own interests as circumstances may require, provided others likewise may be expected to honour them<sup>89</sup>.

Being 'reasonable' is a moral idea that includes elements of reciprocity in fair terms of social cooperation according to which 'all who cooperate must benefit, or share in common burdens, in some appropriate fashion as judged by a suitable benchmark of comparison'<sup>90</sup>. These elements situate persons symmetrically in respect to the claims they are expected to make regarding the basic structure of society. It is important to notice that being reasonable places a constraint on the variety of ends and conceptions of the good that citizens are allowed to pursue and endorse within a just society<sup>91</sup>.

In the original position parties choose the principles of justice behind a veil of ignorance. The function of this device is to neutralise the influence of morally arbitrary social and natural circumstances. The parties ignore their place in society, social class, fortune in distribution of natural assets and abilities, intelligence or

---

<sup>87</sup> Ibid. 23.

<sup>88</sup> I will discuss the demands of reasonableness in Chapter 6. Cf. James Boettcher, 'What is Reasonableness', *Philosophy and Social Criticism*, 30 (2004), 597-621.

<sup>89</sup> Rawls, *Justice as Fairness*, 7.

<sup>90</sup> Rawls, 'Kantian Constructivism in Moral Theory', 316.

<sup>91</sup> Cf. Rawls, *A Theory of Justice*, 17.



strength, sex and race. They also ignore the content of the particular conception of the good which they endorse. Moreover they ignore certain facts about their psychological dispositions, such as the level of risk aversion or their (likely) susceptibility to the types of cognitive failures discussed in Chapters 1 and 2. They are, however, aware of some general facts about human society, principles of economic theory and the 'laws of human psychology'<sup>92</sup>.

The parties behind the veil of ignorance choose principles for the distribution of primary goods, which include, first, the basic liberties that form the

background institutions necessary for the development and exercise of the capacity to decide upon and revise, and rationally to pursue, a conception of the good. [Second,] freedom of movement and free choice of occupation against a background of diverse opportunities [...] required for the pursuit of final ends. [Third,] powers and prerogatives of offices and positions of responsibility [...] needed to give scope to various self-governing and social capacities of the self. [Fourth], 'income and wealth, understood broadly as they must be, [as] all purpose means for achieving directly or indirectly almost any of our ends, whatever they happen to be. [Finally,] the social basis of the self respect [which] are those aspects of basic institutions which are normally essential if individuals are to have a lively sense of their own worth as moral persons and to be able to realize their higher-order interests and advance their ends with zest and confidence'<sup>93</sup>.

This theoretical background relates to autonomy in two important respects: According to Rawls, the parties are autonomous insofar as in 'their deliberations

---

<sup>92</sup> Ibid. 118-9.

<sup>93</sup> Rawls, 'Kantian Constructivism in Moral Theory' 314-5, see also John Rawls, 'Social Unity and Primary Goods' in his *Collected Papers*, 362-3.

they are not required to apply [...] any prior antecedent principles of right and justice' (entailing that justice as fairness is a pure procedural theory). Additionally they are said to be moved solely by the highest-order interests in their moral powers and by their concern to advance their determinate but unknown final ends'<sup>94</sup>. The account of the index of primary goods addresses the latter aspect of autonomy.

### *Scanlon's Contractualism and the Millian Principle*

Scanlon's contractualist argument for the Millian Principle is based on a theoretical perspective not vastly different to that of Rawls<sup>95</sup>. According to Scanlon, the parties in the original position would not adopt restrictions on speech except in the very special cases of severely diminished capacities.

According to the model presented above, it would be unreasonable to choose principles that are either based on determinate conceptions of the good<sup>96</sup> or that do not advance our moral powers. It does not follow from this, however, that rejecting the Millian Principle would be an unreasonable choice. To make this claim Scanlon must show either that the Millian Principle is the one that best serves our highest-order interests or, conversely, that not to choose it would diminish the probability of fulfilling those interests.

It is certainly not obvious why the Millian Principle could be the best way to advance our moral interests. In particular, not paying attention to the ecology of speech, that is, to the ways in which speech affects people non-autonomously, seems to inflict a high cost. It is not unreasonable to think that the parties could choose a

---

<sup>94</sup> Rawls, 'Kantian Constructivism in Moral Theory', 315.

<sup>95</sup> Scanlon, 'A Theory of Freedom of Expression', 15-21; Brison, 'The Autonomy Defense of Free Speech', 329; Amdur, 'Scanlon on Freedom of Expression', 290.

<sup>96</sup> For the role of neutrality see Chapter 5.

principle of speech whose intention is to neutralise the possible triggering of mental contamination. This latter strategy could place, for instance, content-neutral restrictions on highly contaminating speech.

However, it is possible that the distinction between the content of speech and the mental processes which produce mental contamination are linked in a more intimate way: it may be that certain contents trigger procedures automatically. One of the suggestions of a domain-specific theory of rationality is that this is most probably the case. If, for instance, racial thinking is linked to a domain-specific module of human kinds, then it is likely that the content of such thought and some of the cognitive processes of this module cannot be disentangled that easily. Although this possibility is contentious, it does suggest that neutralising mental contamination might be linked indirectly to certain content-based regulations<sup>97</sup>. In this respect, and contrary to Scanlon's arguments, it seems plausible 'to suppose that rational autonomous hypothetical contractors would agree to allow governments to protect them from certain harms wrought by people acting autonomously or nonautonomously'<sup>98</sup>.

### *Contractualism and the Ecology of Speech*

According to Scanlon's contractualism, 'an act is wrong if its performance under the circumstances would be disallowed by any systems of rules for the general regulation of behavior which no one could reasonably reject as a basis for informed, unforced general agreement'<sup>99</sup>. It is important to highlight two basic differences

---

<sup>97</sup> Cf. Chapter 5

<sup>98</sup> Brison, 'The Autonomy Defense of Free Speech', 329.

<sup>99</sup> Thomas Scanlon, 'Contractualism and Utilitarianism', in *The Difficulty of Tolerance*, 132.

between Scanlon's and Rawls's model of contractualism: while the rationale for the latter is to devise principles that are reasonably adopted under uncertainty, individuals in the former model have a direct reason for finding non-rejectable principles because the very basis for moral principles is that they must be justifiable to other people. As Scanlon writes, we must be concerned with other people's points of view not because 'we might occupy their positions in some other possible world, but in order to find principles that they, as well as we, have reason to accept'<sup>100</sup>. The second difference is that Rawls's contractualism is political (at least in its later development), while Scanlon's is comprehensive<sup>101</sup>.

It seems that it would be reasonable to reject the Millian Principle insofar as it ignores the ecology of free speech. After all, the 'central audience interest in expression [...] is the interest in having a good environment for the formation of one's beliefs and desires'<sup>102</sup>. In not seeking to neutralise mental contamination, it places audiences in a situation in which they cannot respond to their circumstances in an appropriate manner. For instance, reasonable persons would be motivated to neutralise the effects of aversive racism which is defined as

a modern form of prejudice that characterizes the racial attitudes of many Whites who endorse egalitarian values, who regard themselves as nonprejudiced, but who discriminate in subtle, rationalizable ways. According to the aversive perspective, many Whites who consciously and sincerely support egalitarian principles and believe themselves to be nonprejudiced also unconsciously harbor negative feelings and beliefs about Blacks, which may be

---

<sup>100</sup> Scanlon, *What We Owe to Each Other*, 191.

<sup>101</sup> Another difference is that while Rawlsian principles must be acceptable to all reasonable persons, Scanlon's requirement is weaker; they must be non-rejectable. Cf. *ibid.* 154-5; Scanlon, 'Contractualism and Utilitarianism', 133.

<sup>102</sup> Scanlon, 'Freedom of Expression and Categories of Expression', 91.

based in part on almost unavoidable cognitive, motivational, and sociocultural processes [...]. These unconscious negative feelings and beliefs may be implicit attitudes, whereas the conscious, self-reported egalitarian attitudes of aversive racist may represent explicit attitudes<sup>103</sup>.

In contrast with old-fashion, explicit racists, aversive racists endorse fair treatment of all groups. However, despite their good intentions they still display feelings of uneasiness towards black people and tend to avoid interracial interactions. When they interact, they tend to display anxiety and discomfort<sup>104</sup> and/or to automatically respond in a fashion consistent with the stereotype they consciously reject (for instance, making self-fulfilling prophecies possible).

As audiences, it is in our interests to reduce forms of mental contamination that might threaten the development and exercise of our sense of justice. Aversive racism is one example of the problems people with the right motivation have in responding in the appropriate way. The effectiveness of justice is highly diminished.

The second moral power of audiences might also be compromised by mental contamination. It matters, in particular, not only that we have the capacity for a conception of the good, but also the manner in which this capacity is exercised<sup>105</sup>. As Scanlon himself writes, '[e]xpression is a bad thing if it influences us in ways that are unrelated to relevant reasons, or in ways that bypass our ability to consider these reasons',<sup>106</sup>.

---

<sup>103</sup> John Dovidio, Kerry Kawakami, Craig Johnson, et. al., 'On the Nature of Prejudice: Automatic and Controlled Processes', *Journal of Experimental Social Psychology*, 33 (1997), 512.

<sup>104</sup> John Dovidio and Samuel Gaertner, 'Aversive Racism', *Advances in Experimental Social Psychology*, 36 (2004), 8.

<sup>105</sup> I will distinguish the relation between persons and the way in which they affirm their conceptions of the good regarding liberal institutions and the demands of neutrality in Chapter 6.

<sup>106</sup> Scanlon, 'Freedom of Expression and Categories of Expression', 90.

Reasonable persons could also reject the Millian Principle based on the interests of the victims of speech. In an environment in which racial stereotypes are widely harboured victims suffer from discrimination and unequal opportunities, not because the structural arrangements are explicitly discriminatory, but because people who make important decisions might be (and often are) aversive racists. Moreover, a social environment rife with implicit (and explicit) racial prejudice might affect the way in which the social basis of self-respect is distributed. According to Rawls, among the list of primary goods, perhaps the most important is this one<sup>107</sup>. The basis of the self-respect has two aspects. First 'it includes a person's sense of his own value, his secure conviction that his conception of his good, his plan of life, is worth carrying out'. Second, 'respect implies a confidence in one's abilities, so far as it is within one's power, to fulfil one's intentions'<sup>108</sup>. It seems clear that the importance of this primary good is that without it nothing would seem worth doing; an individual with no self-respect will lack the motivation to pursue her ends and goals. Rawls concludes: 'therefore the parties in the original position would wish to avoid (at almost any cost) the social conditions that undermine self-respect'<sup>109</sup>. The harms brought about by hate speech undermine the self-respect and self-esteem of their victims, providing persons a with *pro tanto* reason to minimise the effects of these harms<sup>110</sup>. Similarly, a principle that allows the protection of some forms of hate

---

<sup>107</sup> Rawls, *A Theory of Justice*, 386.

<sup>108</sup> Ibid. 386.

<sup>109</sup> Ibid. 386.

<sup>110</sup> 'A *pro tanto* reason has a genuine weight, but nonetheless may be outweighed by other considerations. Thus calling a reason a *pro tanto* reason is to be distinguished from calling a *prima facie* reason, which I take to involve an epistemological qualification: a *prima facie* reason *appears* to be a reason, but may actually not be a reason at all, or may not have weight in all cases it appears to. In contrast a *pro tanto* reason is a genuine reason –reason with actual weight– but it may not be a decisive one in various cases'. Shelly Kagan, *The Limits of Morality* (Oxford: Oxford University Press, 1989), 17.

speech 'does not serve the basis for mutual recognition'<sup>111</sup>. According to Scanlon 'a person [...] acts wrongly when she or he simply fails to take notice of considerations that [...] principles [which no one could reasonably reject] hold relevant'<sup>112</sup>. Social situations in which some are treated as inferior, or as having less worth than others, are an objectionable component of the social arrangements of that society. Beliefs in racial superiority, and the stereotypes it produces, present a reason to alter the institutions that preserve them<sup>113</sup>.

### Conclusions

In this chapter, I divided Scanlon's principle into two parts. The first I called the discontinuous responsibility argument. The second is an account of how autonomous agents would regulate speech. I claimed that the Millian Principle relies heavily on the distinction between means and goals and on a particular account of personal responsibility.

I argued that the difference between ends and means is untenable (at least in the terms Scanlon suggests). I also criticised the Millian Principle for not assessing the influences and processes to which people are exposed when making judgments. Some of the difficulties of Scanlon's principle reappear in his discussion about paternalism. I tried to rephrase the discontinuous responsibility argument using Scanlon's later distinction between substantive responsibility and responsibility as

---

<sup>111</sup> Scanlon, *What We Owe to Each Other*, 194.

<sup>112</sup> *Ibid.* 269.

<sup>113</sup> Thomas Scanlon, 'The Diversity of Objections to Inequality', in *The Difficulty of Tolerance*, 204.

attributability. I argued that on the one hand, it is the case that sometimes people should not be held responsible for the acts performed by others based on the former's acts of expression. On the other hand, however, that does not mean that they cannot be held substantively responsible for some wrongdoing or that we cannot impose certain burdens upon them, in this case, restrictions on speech.

The discussion of citizen's autonomy was first contrasted with the Rawlsian original position. I also tried to accommodate the Millian Principle within Scanlon's own version of contractualism. I claimed that people could reasonably reject a principle that does not attempt to neutralise (at least some) cases of mental contamination. It seems that conditions for justifiability demand such evaluations.

From this analysis of the Millian Principle we can deduce two conclusions. First, this version of the autonomy defence of speech is untenable. If speech bypasses autonomy, the latter cannot be employed as a reason to go beyond the harm done through speech. Second, it is in an audience's autonomy-based interests to minimise grave cases of mental contamination, and these interests might be in conflict with free speech. Serious cases of mental contamination (such as aversive racism) and the harm it brings on third parties need to be taken into consideration when establishing principles of free speech. We cannot operate under the assumption 'that having been exposed, an audience is always free to decide how to react: what belief to form or what attitude to adopt'<sup>114</sup>. In this respect I suggest that there is a *pro tanto* reason to control the ecology of speech

However, maybe this reason is not sufficient for restrictions on speech, we must also consider the interests of speakers. This is the task of the next chapter.

---

<sup>114</sup> Scanlon, 'Freedom of Expression and Categories of Expression', 89.



## CHAPTER 4

### RONALD DWORKIN AND THE INTERESTS OF SPEAKERS

In the last chapter I argued that speech is not always positive in terms of audience interests. I suggested that the autonomy defence fails on two accounts. First, if speech bypasses autonomous processes, then it cannot provide the exception to the harm principle originally claimed. In this respect, it is unreasonable that the victims of speech must bear all its costs. Instead, we must look at how the social ecology of speech is built and distribute the costs accordingly. Second, I argued that, contrary to the indications of the Millian Principle, it is reasonable for persons to protect themselves from highly contaminating speech. I suggested that forms of mental contamination which are particularly troublesome for citizens are those which jeopardise the exercise of their two moral powers: the capacity for a sense of justice and the capacity to form, revise, and pursue conception of the good. The latter conclusion is stronger than the former insofar as it imposes a positive duty on government to take action aiming at neutralising the effects of these forms of mental contamination.

However, even if it is in audiences' interests to be protected from the harmful effects of speech, this claim is still not conclusive. We need to look at the interests of

speakers<sup>1</sup> with respect to speech, and when there is conflict we must find a balance. A speaker has two important interests with respect to speech. First, she has an *expressive* interest in ‘being able to call something to the attention of a wide audience’<sup>2</sup> for a number of reasons: she might want to ‘bear witness’ to a viewpoint concerning, say, political justice or to articulate thoughts, attitudes and feelings on matters of personal or broader human concern’<sup>3</sup>. An important dimension of having a sense of justice and being a moral person includes trying to persuade others about the validity of our convictions. Sometimes, speaking our minds is necessary for personal integrity; seeing justice not being done gives a strong reason to express our views and to try to rectify this. Similarly, we want to communicate the reasons we have for valuing our most cherished convictions because we feel that people would be better off if they reflected and also endorsed them. Expressive interests are not necessarily of such serious order; we also want to communicate to others our experiences and thoughts, even the trivial ones, and we want to amuse people and form bonds in which language and expression are important.

Second, a speaker has also *deliberative* interests. Sometimes articulating and publicly expressing a viewpoint can be an important part of subjecting it to sharper assessment and critical analysis. Although these interests have different weight in a theory of free speech, they are both presumptions for expression. My interest in communicating my views about the worthiness of a religious life or about the

---

<sup>1</sup> I refer as ‘speakers’ to all participants who have interests in free speech, even though some of them are not speakers in the usual sense (they might participate by making films, or showing pictures, or painting graffiti, or burning flags). I believe that ‘speakers’ is a better word than participants because audiences and bystanders are also participants in speech situations. The difference, though, is unimportant as it is only a matter of usage.

<sup>2</sup> Scanlon, ‘Freedom of Expression and Categories of Expression’, 86.

<sup>3</sup> Cohen ‘Freedom of Expression’, 224.

unreasonableness of right wing conservatism are generally thought to be far weightier than my interests in telling jokes, or chatting about what I had for lunch<sup>4</sup>.

Ronald Dworkin has produced a number of important arguments defending freedom of speech. Although he has not systematically constructed a theory of free expression, his arguments are important and strong. Their assessment is the focus of this chapter. I will interpret his theory of speech as based mainly on speakers' autonomy. I am not suggesting that he believes that only speakers' are concerned with speech, for he also agrees that it serves audiences' autonomy. Nevertheless, he does suggest that the interests of speakers' alone are sufficient to warrant a free speech principle<sup>5</sup>.

## I

### Free Speech as a Trump over Social Utility

In *Taking Rights Seriously*<sup>6</sup>, Dworkin rejects a utilitarian account of rights in favour of the claim that 'individuals have rights, when, for some reason, a collective goal is not a sufficient justification for denying them what they wish, as individuals, to have or to do, or not a sufficient justification for imposing some loss or injury upon them'<sup>7</sup>. In his earlier defence of free speech, Dworkin uses this framework to reject the view that the social benefits of regulating expression are sufficient justification for them.

---

<sup>4</sup> A way of assigning weights to forms of speech consists in distinguishing between different categories of expression. Cf. Scanlon, 'Freedom of Expression and Categories of Expression'.

<sup>5</sup> Dworkin, *Sovereign Virtue*, 160-1. He seems to agree with the U.S. Supreme Court in that speech might be curtailed only in cases of 'clear and present danger'.

<sup>6</sup> Ronald Dworkin, *Taking Rights Seriously* (London: Duckworth, 1977).

<sup>7</sup> *Ibid.* xi.

In his discussion of obscenity and pornography, he holds that people have a right to moral independence, according to which,

People have the right not to suffer disadvantage in the distribution of social goods and opportunities, including disadvantage in the liberties permitted to them by criminal law, just on the ground that their officials or fellow-citizens think that their opinions about the right way for them to lead their own lives are ignoble or wrong<sup>8</sup>.

Dworkin argues that attempts to regulate or ban pornography or obscenity are based on an assessment of the intrinsic worth of these kinds of expression<sup>9</sup>. The motivation behind regulation is that, if citizens' preferences are to be given due weight, and the majority of citizens prefer to live in a pornography free society, then, there is case for banning such material<sup>10</sup>. A social policy that satisfies most preferences would support a ban on pornography. Dworkin's defence of a right to moral independence provides a constraint on this justification: he argues that the 'moral pain' and the moral offence that pornography and obscenity can produce on people do not provide a reason for their regulation<sup>11</sup>.

Another way in which Dworkin's argument can be understood is to distinguish between two types of preferences that might affect the distribution of burdens and benefits in a utilitarian society. He argues that an individual has a *personal* preference when she desires some goods or opportunities for herself. Clearly a person might want to have access to education and income, to sport and cultural venues, and she wants these opportunities to be open to her. On the other

---

<sup>8</sup> Ronald Dworkin, *A Matter of Principle* (Cambridge, MA: Harvard University Press, 1985), 353.

<sup>9</sup> The presumption against content-based regulation will be addressed in the next chapter.

<sup>10</sup> The same argument can be applied to a variety of issues including homosexuality, divorce, racial segregation, hate speech, etc.

<sup>11</sup> He acknowledges that it could justify its restriction in certain times and places (through zoning, or by selling pornographic magazines only in specialised shops, or by selling them with black covers).

hand, a person has *external* preferences when she prefers some opportunities and goods to be open (or closed) to others<sup>12</sup>. Someone might prefer a swimming pool to be built in his town rather than a theatre, not because he enjoys swimming more than plays, but because he does not want theatregoers to have a good time. Or he might prefer more funding to be allocated to hospitals than schools because he wants his ill friends to be able to get better treatment. An implication of the moral independence idea is that external preferences should be discarded when making political decisions, because a person's rights and opportunities should not depend on how his fellow citizens regard how he leads his life.

Although compelling, this argument does not prove that a strong principle of free speech is to be protected. In particular, it does not provide reasons to protect speech regardless of the harmful consequences it may have. Dworkin's argument shows only that offence and moral pain, *in general*, are not sufficient justification to restrict a person's action. In this sense, Dworkin's argument is a form of minimalism which attempts to show that the alleged harms of speech are, in fact, not harms in the relevant sense. For instance, contrast an effort to regulate the publication of obscene material with an attempt to forbid the practice of homosexuality. Imagine that in both cases the 'bare knowledge' that these phenomena occur in a society produces enough discomfort and moral reprobation in its members to motivate them to forbid such activity. Now, the reason these attempts should fail is the same, namely that discomfort and reprobation are not valid justification for prohibition. In short, Dworkin's argument fails to defend a principle of free speech; it merely shows that external preferences and moral offence cannot be used as a justification to curtail people's activities.

---

<sup>12</sup> Dworkin, *Taking Rights Seriously*, 234.

It is remarkable that Dworkin backs this minimalist strategy in his discussions about the effects of pornography. He argues that as yet, there is no respectable study linking pornography with sexual violence<sup>13</sup> and that, more generally, pornography does not harm women in the sense required for the harm principle to apply<sup>14</sup>. So, Dworkin concludes, given that pornography does not harm women, the only reasons to ban it are based on either external preferences (people who would prefer others not to read pornographic material) or on considerations of offence and moral distress<sup>15</sup>.

## II

### Authenticity and Equality of Resources

This is not the only defence available to Dworkin, for he could concede that pornography indeed harms women but maintain that there are, nevertheless, reasons to protect it. That is, he could change his minimalist for a maximalist strategy. In his later work, he seems to adopt this tactic.

Another defence of free speech is derived directly from Dworkin's defence for equality of resources<sup>16</sup>. Dworkin believes that the fundamental goal of government is to adhere to the abstract egalitarian principle according to which, '[it] must act in a way that makes the lives of those it governs better lives, and it must

---

<sup>13</sup> Ronald Dworkin, *Freedom's Law* (Cambridge, MA: Harvard University Press, 1996), 206. But see Hurley, 'Bypassing Conscious Control' which could provide an explanation for the direct link between exposure to violent pornography and violence against women.

<sup>14</sup> See Dworkin, *Freedom's Law*, 206-7, 217, 219-20, 230.

<sup>15</sup> A critical assessment of Dworkin's views on pornography is Rae Langton, 'Whose Rights? Ronald Dworkin, Women, and Pornographers', *Philosophy and Public Affairs*, 19 (1990), 311-59; see also Brison, 'The Autonomy Defense of Free Speech', 324-5.

<sup>16</sup> Equality of resources is presented and defended in *Sovereign Virtue*, Chapter 2.

show equal concern for the life of each'<sup>17</sup>. Dworkin thinks that a defence of free choice (including freedom of speech) can be directly derived from his account of equality of resources. He argues that equality and liberty are not two different values which may potentially conflict, but, rather, that in a satisfactory account of equality of resources 'liberty becomes a [constitutive] aspect of equality'<sup>18</sup>. Or, in other words, 'liberty is necessary to equality [...] because liberty, whether or not people value it above all else, is essential to any process in which equality is defined and secured'<sup>19</sup>.

Dworkin thinks that an equal distribution of resources must be sensitive to people's preferences<sup>20</sup>. He imagines a situation in which all impersonal resources are to be distributed through an auction. People bid for the resources they desire, until no one 'would prefer someone else's bundle of resources to his own bundle'<sup>21</sup>. This 'envy test' is a necessary condition of an egalitarian distribution<sup>22</sup>.

The auction-based device involves people taking responsibility for the costs of their own ambitions and projects, because the costs of any given resource will fluctuate according to how much others bid for it. How much I bid for the availability of places of worship will be a function of how much I and others value the practice of religion<sup>23</sup>.

---

<sup>17</sup> Ibid. 128.

<sup>18</sup> Ibid. 123.

<sup>19</sup> Ibid. 122.

<sup>20</sup> Notice that while Rawlsian primary goods are derived from citizens' needs (in terms of their two moral powers), Dworkin's resources accommodate individuals' preferences

<sup>21</sup> Ibid. 67.

<sup>22</sup> It is not sufficient, though. An egalitarian distribution still needs to neutralise the effects of 'brute' luck in the distribution of personal resources through a hypothetical insurance device. Cf. *ibid.* 73-82.

<sup>23</sup> People can bid for the availability of religious services but not for the right to practice religion. So, although religion is to be tolerated, the number of churches available will vary according to other people's bids.

Because the envy test might be satisfied in a variety of ways<sup>24</sup>, Dworkin thinks that we need a bridge between it and the egalitarian principle: one of such bridges is the principle of abstraction, which provides an account of true opportunity costs, thus ensuring that the auction is fair. This principle ‘recognizes that the true opportunity cost of any transferable resource is the price others would pay for it in an auction whose resources were offered in as abstract as possible, that is, in the form that permits the greatest possible flexibility in fine-tuning bids to plans and preferences’<sup>25</sup>. This principle allows liberty to figure in the auction because it is necessary for discerning the true opportunity costs of different resources, which is necessary information if government is to treat everyone under its command with equal concern and respect. The principle of abstraction, then, limits free choice under circumstances in which a person’s actions would harm third parties (or put them at significant risk); perfectionist or paternalistic restrictions are in principle disallowed<sup>26</sup>. The principle of abstraction ensures neutrality of justification by describing the resources to be auctioned in as abstract a way as possible and by insisting that ‘the resources people have available [...] be fixed by the costs of their having these to others, rather than by any collective judgment about the comparative worth or projects or personal moralities’<sup>27</sup>.

Another principle that bridges the envy test and the egalitarian principle is the principle of authenticity. This principle gives guidance regarding which liberties are more important than others and, importantly, explains why a speakers’ interests

---

<sup>24</sup> Imagine that, before the auction, all resources are traded for plovers’ eggs and pre-phylloxera claret. After the auction, although someone who does not like plovers’ eggs or pre-phylloxera claret may not want to change her bundle for someone else’s, the government fails to satisfy the abstract egalitarian principle because it does not treat every person with equal concern. Ibid. 67-8.

<sup>25</sup> Ibid. 151.

<sup>26</sup> Some other permissible restrictions aim at correcting the externalities of the auction. Ibid. 156.

<sup>27</sup> Ibid. 154.



based principle of free speech is constitutive of equality of resources. The concern for authenticity derives from the notion that the auction must be sensitive to parties' preferences and life plans. In order to respect the parties as equals, any permissible auction

requires [...] some baseline principles specially protecting the parties' freedom to engage in activities crucial to forming and reviewing the convictions, commitments, associations, projects, and tastes that they bring to the auction, and after the auction, to the various decisions about production and trade that will reform and redistribute their initial holdings<sup>28</sup>.

A general protection of free speech clearly figures in this account of authenticity, for it is advantageous to the parties to express their commitments and beliefs to a wide audience. The principle of authenticity captures the basic idea underlying the autonomy defence of free speech by insisting that speakers have an interest in having 'an opportunity to influence the corresponding opinions of others, on which their success in the auctions in large part depends'<sup>29</sup>. Restricting the opportunity a person has to influence others is an arbitrary way of distorting the opportunity costs of that person's lifestyle; such a distortion constitutes a breach to the egalitarian principle of equal respect. For instance, a person who is not allowed to convince others to bid for sport facilities in her community will find that the costs of exercising are higher than they would have been if she had been free to convince her neighbours to bid for those facilities. If she must invest more resources for securing the sport facilities because she was not allowed to convince others then the auction would not have

---

<sup>28</sup> Ibid. 159.

<sup>29</sup> Ibid. 160.

been fair<sup>30</sup>. It is from this perspective that Dworkin thinks authenticity establishes a free speech principle based on speakers' interests.

However, Dworkin believes that we must also constrain authenticity because it allows too much. As stated, it does not distinguish between different kinds of preferences and might deliver results that fail to treat some members with equal concern. The principle of independence corrects this failure. Just as some lifestyles have higher costs (resulting from wrong interferences, such as restrictions on speech, or policies that *deliberately* increase their costs based on an account of their intrinsic merits<sup>31</sup>), a person's life can be inflicted with distorted opportunity costs if other people's bids reflect contempt or dislike for her. A clear example could be a situation in which members of a racist community bid for areas of land and build segregated zones, leaving members of other races with fewer resources and opportunities than they would have had in the absence of such a prejudice. The principle of independence limits the principle of abstraction by imposing protections to persons who 'are the objects of systematic prejudice from suffering any serious or pervasive disadvantage from that prejudice'<sup>32</sup>.

How should hate speech be accommodated at the intersection of the principles of authenticity and independence? It appears from Dworkin's writing that authenticity takes priority over independence in the sense that the harms done through speech are not sufficient to restrict the authenticity of speakers. This can be defended in two ways.

---

<sup>30</sup> Even though the distribution might be envy-free. She may still prefer to pay the extra resources in order to exercise than being in others' situation.

<sup>31</sup> Of course this does not mean that Dworkin thinks that all ways of life must equally costly. He does not appeal to neutrality of impact. Ibid. 154, 281-3.

<sup>32</sup> Ibid. 161.

First, Dworkin can take a minimalist approach and claim that independence does not apply here because speech does not distort the opportunity costs of victims. On this view, the costs of speech are similar to other costs which are not condemned by equality of resources. For instance, Bob enjoys playing baseball but because he lives in the UK, the cost of doing so is higher than in the US. The higher cost, of course, is a function of others people choices and preferences (in the UK most people bid for football and cricket). Nevertheless, Bob has neither been wronged nor disadvantaged in a justice-based sense.

Now, contrast this case with John, a person who suffers due to other people's racist attitudes. Although John is protected from discrimination in employment, he nevertheless is not protected from the effects of hate speech (or, as I shall discuss in Chapter 6, from the effects of discriminatory associations). Based on a minimalist approach, Dworkin can argue that the effects of speech and association are not detrimental to justice: they are akin to those suffered by Bob. However, this strategy does not seem promising for two reasons. On the one hand, a robust defence of free speech cannot ignore its costs, and, as we have seen, the automatic effects of speech can have grave consequences: someone can fail to obtain a job due to racial stereotypes and implicit associations, people might react aggressively towards her with no good reason, and so on. On the other hand, equality of resources cannot be neutral about those effects, because they flow from racial or sexist prejudice. Equality of resources must condemn 'the attitudes that create disadvantage' based on racial or sexual prejudice<sup>33</sup>.

The second strategy available to Dworkin is a maximalist defence, according to which authenticity takes precedence over independence. This strategy has the

---

<sup>33</sup> Ibid. 162.

advantage of recognising the costs of speech, but, it insists, these are not sufficient to restrict expression.

Notice, though, that, from the perspective of authenticity the interests involved in free speech are not equally situated. As noted in the previous chapter, it makes sense for *audiences* to seek protection from mental contamination, especially from those forms that threaten aspects of our lives to which we attribute great importance, such as the way we treat others. In this sense, the principle of abstraction could be limited so as to accommodate the ecology of speech. It would not be unreasonable to hold that the auction must be sensitive to the environment in which people form and revise their personalities when those limits are established with the intention of realising circumstances in which people are all treated with equal respect and concern.

Although Dworkin could argue that equality of resources *already* places a limit on *public* actions that fail to treat everyone with equal respect and concern by prohibiting state-sponsored racism or legal forms of sexist and racial discrimination, this does not accommodate restrictions on *private* actions. The effects of mental contamination, therefore, undermine this objection, for the harmful impact of aversive racism and similar processes bypass the public/private distinction<sup>34</sup>. Although deliberate governmental distortions of opportunity costs might be worse than those resulting from private actions, there is no reason to suppose that the latter is irrelevant for the purview of justice. Dworkin himself recognises that discrete actions of private actors have serious (intended and unintended) effects on victims and audiences. In his discussion of Mackinnon's claim that pornography harms

---

<sup>34</sup> I will discuss the spill-over effects of mental contamination with relation to private associations in Chapter 6

women he argues that, given that only a minority of men consume pornographic material<sup>35</sup>, other forms of sexism can be more influential. In this respect, he acknowledges that

[c]ommercial advertisement and soap operas create images that are subtle and ubiquitous, and it would not be surprising to learn, through whatever research might establish this, that they indeed do great damage to the way women are understood and allowed to be influential in politics<sup>36</sup>.

Indeed, the idea of mental contamination and the empirical evidence that gives it support highlight the harmful effects of gender and racist stereotypes and implicit associations. Nevertheless, Dworkin thinks that these harms (even if real and important) do not provide sufficient justification to restrict speech.

It is important to bear in mind that the claim defended here is not an example of ‘cultural paternalism’ of the sort Dworkin rejects. Instead it is based on an assessment of how exposure to certain forms of speech might be detrimental to the authenticity of audiences and to the independence of both them and the targets of hate speech. The claim I defend is also not paternalistic in another sense: I am not arguing that it would be *better* for speakers’ authenticity if they were not allowed to express their views because they are immoral or unpopular. The argument is merely that, from the point of view of authenticity, it is not clear why a speakers’ right to hate speech necessarily trumps that of audiences’ and targets’. Moreover, equality of resources cannot be neutral towards the effects of racial and sexist prejudice<sup>37</sup>. In this respect I argue that a fair distribution of resources must be sensitive to the

---

<sup>35</sup> This is itself a controversial claim, many men are exposed to pornography involuntarily.

<sup>36</sup> Dworkin, *Freedom’s Law*, 220.

<sup>37</sup> Dworkin thinks that equality of resources might treat the disadvantages created by prejudice as handicaps in the sense that justice must be sensitive to those differences. Nevertheless, he acknowledges that in one respect, disadvantages created by prejudice are worse than handicaps. Dworkin, *Sovereign Virtue*, 162.

ecology of speech. Dworkin could still give paramount weight to the interests of speakers, but he cannot do so with the strategies addressed here. The next two arguments addressed are both attempts to find an alternative strategy of sustaining this position.

### III

#### Dignity and Integrity

##### *Critical and Volitional Interests*

In contrast to political liberals, who think that political institutions must be justified (or justifiable) independent of a person's ethical convictions, Dworkin thinks that liberalism must be justified by appealing to some non-political values<sup>38</sup>. In order to do so, Dworkin thinks that we should find values at a sufficiently abstract level that most people would endorse. Once such values are found, we must show that certain political institutions are their best interpretation. Political liberals, Dworkin argues, attempt 'to reduce liberal convictions to the public sphere'. For them, 'liberalism seems to be [...] only a theory of right, totally unconnected to the problem of the good'. In contrast, Dworkin 'tr[ies] to connect ethics and politics by constructing a view about the nature and character of the good life that makes liberal political morality seem continuous [...] with appealing philosophical views about the good life'<sup>39</sup>. Or, as he describes one of the goals of his recent book, 'I shall argue [...] that in spite of the popular opinion [that there is unbridgeable divide], we actually can

---

<sup>38</sup> Cf. Rawls, *Political Liberalism*, especially 'Introduction to the Paperback Edition', xxxvii-lxii; Thomas Nagel, *Equality and Partiality* (Oxford: Oxford University Press, 1991).

<sup>39</sup> Dworkin, 'Foundations of Liberal Equality', 191.

find shared principles of sufficient substance to make a national political debate possible and profitable. These are very abstract, indeed philosophical, principles about the value and the central responsibilities of a human life'<sup>40</sup>.

Dworkin distinguishes between two different kinds of interests possessed by people. On the one hand, a person has interests in achieving things she wants to achieve; interests she harbours *only* because she wishes to satisfy them. If she obtains what she desires, her life will go better. For instance, eating fish for supper tonight will improve my well-being only because I desire that meal. Dworkin terms this dimension of well-being 'volitional well-being' and the interests we have in securing it 'volitional interests'<sup>41</sup>. On the other hand, a person has 'critical' interests in achieving certain things, not because she wants them but because their achievement will improve her life. Dworkin cites, as examples, a close relationship with one's children, some success in our professional lives, and moderate grasp about the state of advanced science. The thought here is that these things improve one's critical well-being *independently* of whether we want them or not: well-being in this sense is 'improved by [a person's] having or achieving what he should want, that is, the achievements or experiences that it would make his life a worse one not to want'<sup>42</sup>. Ideally these interests match each other: a person usually wants to obtain what is good for her life in the critical sense and sometimes it is critical for a person to get some degree of success in the goals she wants to achieve. Obviously, however,

---

<sup>40</sup> Dworkin, *Is Democracy Possible Here?*, 7-8. In the next chapters I defend a qualified version of political liberalism in two ways: one argues that a partly comprehensive liberalism is necessary for legitimacy (Chapter 5). The other argues that 'public reason' is complete enough to give legitimate answers to political decisions (Chapter 6).

<sup>41</sup> Dworkin, *Sovereign Virtue*, 242.

<sup>42</sup> Dworkin, 'Foundations of Liberal Equality', 230. Although one can question this distinction (e.g. some hedonistic utilitarians can argue that both kind of interests necessarily collapse together), I will grant the validity of the distinction for the sake of the argument. Notice also that what counts as a critical interest depends, in part, on each person's circumstances. To me, playing football well increases my volitional well-being, whereas it is critical to Wayne Rooney's life.

sometimes they also come apart: myopia, weakness of the will, and addictions all seem to be forms of a person's wants and actions that do not serve her critical well-being. On the other hand, some people simply do not care about how well or badly their lives go<sup>43</sup>. It is not that they are mistaken about their critical interests, but rather that they disregard the quality of their lives: according to Dworkin, 'we think that [...] their lives are defective in a particular and demeaning way: they lack dignity'<sup>44</sup>.

### *Dignity and Political Equality*

The idea of dignity is appealing, according to Dworkin, because it embodies values that most people share regardless of their particular viewpoints on more concrete issues in politics. It is an example of the 'continuity' strategy for justifying liberalism. According to Dworkin, dignity has two aspects. First, human life has an inherent value, and that value is 'axiomatic and fundamental. It is important for no further reason than that [people] have a life to live'<sup>45</sup>. Now, if someone's life has intrinsic value for himself, then he must concede that everyone else's life has the same intrinsic value. The first person's perspective does not make his life more important than others. If this is true, then we are committed to the view 'that it is objectively important that once any human life has begun, that life go well and not be wasted. [We] must also accept that this is equally important for each person

---

<sup>43</sup> Dworkin, *Is Democracy Possible Here?*, 14.

<sup>44</sup> *Ibid.* 14

<sup>45</sup> *Ibid.* 15. Cf. what he calls the 'sacredness' of life. According to him 'human life has an intrinsic, innate value; [...] human life is sacred just in itself'. Ronald Dworkin, *Life's Dominion* (New York: Vintage, 1994), 11.



because [we] have no ground for distinctions of degree any more than for flat exclusions'<sup>46</sup>.

The second dimension of human dignity holds that

Each of us has a personal responsibility for the governance of his own life that includes the responsibility to make and execute ultimate decisions about what life would be a good one to lead. We may not subordinate ourselves to the will of other human beings in making those decisions; we must not accept the right of anyone else to force us to conform to a view of success that but for coercion we would not choose<sup>47</sup>.

This principle of 'special responsibility' grounds the argument that people must be treated not only with concern, but also with respect, and will figure prominently in this version of Dworkin's defence of free speech.

Given that a person has a critical interest in living a good life, Dworkin is able to connect the abstract egalitarian principle with the idea of dignity. Recall that, according to Dworkin, 'government must act in a way that makes the lives of those it governs better lives'<sup>48</sup> and that it must do so in a manner consistent with their dignity. This implies that government attempts to increase its subjects' critical well-being are constrained by two factors: first, in all its actions it must show equal concern to all people (as embodied by the principle of equal value of everyone's life), and second, it must treat them with equal respect (as embodied by the idea of special responsibility). A government that fails to satisfy these conditions loses, to that degree, its legitimacy<sup>49</sup>.

---

<sup>46</sup> Dworkin, *Is Democracy Possible Here?*, 16. See also Dworkin, *Sovereign Virtue*, 5, 240.

<sup>47</sup> Dworkin, *Is Democracy Possible Here?*, 17. In *Sovereign Virtue* Dworkin clarifies that his use of 'responsibility' is similar to Scanlon's 'substantive responsibility'. *Sovereign Virtue*, 489, n. 4.

<sup>48</sup> Dworkin, *Sovereign Virtue*, 128.

<sup>49</sup> On legitimacy, Dworkin writes that citizens 'assume [...] political obligations only if and so long as the community's government respects their human dignity. Only so long [...] as it accepts the equal

There are two avenues for defending free speech based on the notion of human dignity. The first is a direct argument premised on the idea of responsibility, the other emanates from a conception of political equality and democracy. I will briefly explore each here. Dworkin argues for a ‘constitutive’ version of free speech (as opposed to a consequentialist version) on the grounds that free speech is an ‘essential [...] feature of a just political society’<sup>50</sup>. In this respect, free speech is necessary to protect the responsibility each person has to determine their own values, to form convictions, and to communicate them ‘out of concern for others, and out of a compelling desire the truth be known, justice served, and the good secured’<sup>51</sup>. In other words, there is an intimate relation between a person’s special responsibility and her leading an authentic life<sup>52</sup>:

preventing someone from speaking his conscience and conviction to other people is a particularly grave harm. People develop their ethical and moral personalities most effectively in conversation and exchange with others. Speaking out for what one believes —bearing witness and testimony— is in any case for most people an essential part of believing; it is part of the total phenomenon of conviction. Identifying oneself to others as a person of particular beliefs or faiths is part of creating one’s identity, part of the process of self-creation that is at the centre of our personal responsibility<sup>53</sup>.

Free speech is, then, a condition government cannot fail to secure if it is to promote and respect its citizens’ critical well-being.

---

importance of their lives and their personal responsibility for their own lives...’ *Is Democracy Possible Here?*, 97.

<sup>50</sup> Dworkin, *Freedom’s Law*, 199-200.

<sup>51</sup> *Ibid.* 200.

<sup>52</sup> The relation is also between her responsibility and her ethical integrity. I will discuss the latter in the next section.

<sup>53</sup> Dworkin, *Is Democracy Possible Here?*, 153.

The second approach by which free speech can be defended here is based on the idea of political equality, and is linked to the defence of democracy. Dworkin argues that institutions are democratic ‘to the degree that they allow citizens to govern themselves collectively through a partnership in which each is an active and equal partner’<sup>54</sup>. Democracy, understood as such, rejects that majorities *per se* have any moral significance; democracy is a system of collective decision-making that must satisfy certain conditions, conditions mainly aimed at guaranteeing that institutions treat citizens with equal respect and concern, that is, in a manner that is compatible with their dignity<sup>55</sup>.

Certain considerations need to be accommodated in order to understand the relation between democratic equality and freedom of speech. Some of these are captured by the idea of ‘moral membership’. First, in order to be a member in a community in which each individual is treated with respect and concern, an element of reciprocity is needed. Accordingly, Dworkin claims that ‘a person is not a member unless he is treated as a member by others, which means that they treat the consequences of any collective decision for his life as equally significant a reason for or against that decision as are comparable consequences for the life of anyone else’<sup>56</sup>. Second, democracy requires a space of discourse in which collective deliberation takes place. Citizens must be free to discuss and present their views as individuals<sup>57</sup> before any collective decision is taken, ‘and the deliberation must

---

<sup>54</sup> Dworkin, *Sovereign Virtue*, 362. See also ‘democracy means that people govern themselves each as a full partner in a collective enterprise so that a majority’s decision are democratic only when certain further conditions are met that protect the status and interests of each citizens as a full partner in that enterprise’. *Is Democracy Possible Here?*, 131.

<sup>55</sup> For the rejection of the majoritarian premise see ‘Introduction’ in *Freedom’s Law*, and *Sovereign Virtue*, Chapters 4, 10.

<sup>56</sup> Dworkin, *Freedom’s Law*, 25.

<sup>57</sup> As opposed to ‘as citizens’: For Dworkin, individuals come furnished with all their comprehensive views and convictions. Political liberals, on the other hand, claim that political deliberation must be conducted in terms of public reasons, that is, they must put some of their beliefs to one side. Cf.

centre on reasons for and against that collective action, so that citizens who lose on an issue can be satisfied that they had a chance to convince others and failed to do so, not merely that they have been outnumbered'<sup>58</sup>. Democracy also requires citizens to be independent from the collective decisions made by the community: dignity requires that a community does not impose their views on citizens but, on the contrary, it must 'provide circumstances that encourage them to arrive at beliefs on these matters through their own reflective and finally individual conviction'<sup>59</sup>. A government that imposes the views of the majority on its citizens fails to secure that each can live authentic and integrated lives and, to that extent, it compromises its legitimacy. Independence is one of the bridges between our shared convictions regarding the value of human life, equality of resources and our political beliefs.

Another bridge is found in the principle that government must promote three different goals within a democratic society. First, it must secure 'distributive goals' that are fair: for instance, those recommended by equality of resources. Second, it must promote 'symbolic goals' in which the community asserts the equal moral status of its citizens, for instance allocating each person one, and only one, vote is a declaration of the equality of all citizens. Third, it must also secure citizens' 'agency goals', because,

[w]e cannot make our political life a satisfactory extension of our moral life unless we are guaranteed freedom to express our opinions in a manner that, for us, satisfies moral integrity. Opportunity to express commitment to our

---

Rawls, 'The Idea of Public Reason Revisited'; Kent Greenawalt, *Private Consciences and Public Reasons* (New York: Oxford University Press, 1995).

<sup>58</sup> Dworkin, *Sovereign Virtue*, 364-5.

<sup>59</sup> Dworkin, *Freedom's Law*, 26.

convictions is just as important, for that purpose, as the opportunity to communicate those convictions to others<sup>60</sup>.

It is clear why Dworkin believes that free speech is a necessary condition of political equality, and why he gives such an important weight to our expressive interests: a person whose expressive interests are not protected is treated neither with equal respect nor with equal concern. It is important to note that, although Dworkin's defence of free speech appeals to the deliberative interests of both speakers and audiences, he asserts that our expressive interests are sufficient to secure a free speech principle. This perhaps explains why Dworkin seems reluctant to categorise speech. In contrast to other liberals who believe that some forms of speech warrant greater protection because they are better captured by the relation between speech and autonomy<sup>61</sup>, Dworkin supports an extreme form of maximalism, according to which 'the wrong [of censorship] is just as great when government forbids the expression of some social attitude or taste as when it censors explicitly political speech'<sup>62</sup>. In other words, even if speech contributes nothing to processes of deliberation, the fact that a person has an expressive interest in speech suffices for its protection<sup>63</sup>.

This view, however, seems too strong. Autonomy defences of speech are distinctive due to the emphasis they place on deliberative processes and critical reflection. If speech does not serve anyone's deliberative interests then the autonomy-based reasons for its protection are weakened to a considerable degree, for

---

<sup>60</sup> Dworkin, *Sovereign Virtue*, 201-2.

<sup>61</sup> Cf. Scanlon, 'Freedom of Expression and Categories of Expression', 'Content Regulation Reconsidered'.

<sup>62</sup> Dworkin, *Freedom's Law*, 200-1.

<sup>63</sup> Dworkin is not an absolutist about free speech. He thinks that some content-neutral limits can be set on, for instance, political campaigns and donations. Nevertheless he objects to any kind of content-based restrictions. I address this distinction in Chapter 5.

it is clear that self-expression alone is not sufficient to protect speech. There are many ways in which a person can express herself which are best accommodated within a principle of free speech. On the one hand, in the case of harmless actions self-expression is usually best captured by a right to privacy. If I decide to grow a goatee, my right to fashion my facial hair according to my desires is better defended this way than by a right to free expression. On the other hand, self-expression is not sufficient (and perhaps not necessary<sup>64</sup>) in the case of harmful actions: ‘Nero was presumably expressing himself aesthetically when he (perhaps apocryphally) spilled fresh human blood on green grass’<sup>65</sup>.

*The Model of Challenge, Ethics and Integrity*

There is another argument that Dworkin employs to justify free speech, this one based on the idea of ethical integrity. Dworkin provides a controversial defence of liberal equality based on an ethical conception of the good life<sup>66</sup>. Rather than justifying liberalism with political reasons, he thinks that his approach strengthens the case for a liberal political morality<sup>67</sup>. In contrast to this account, the case built on the argument from dignity is less demanding and less controversial. Dworkin does not believe it is necessary to accept this account of ethics to accept liberal equality, but there is, nonetheless, an important connection between the two. His strategy consists in defending his preferred account of political morality as the best interpretation of non-political values.

Recall the ‘abstract egalitarian principle’, according to which government must make the lives of those under its command better, with the constraint that it

---

<sup>64</sup> Not necessary, at least, if a condition of ‘sincerity’ is attached to self-expression.

<sup>65</sup> Hurley, ‘Bypassing Conscious Control’, 318.

<sup>66</sup> Dworkin, *Sovereign Virtue*, 239.

<sup>67</sup> *Ibid.* 241.

treats everyone with equal respect and concern. An important question for this approach is, then, how to measure the value of a human life? What standards can we employ to know whether a person's life goes better or worse? In answer to these questions Dworkin presents his 'model of challenge', which 'supposes that a life is successful insofar as it is an appropriate response to the distinct circumstances in which it is lived'<sup>68</sup>. The model employs the idea of a skilful performance as a paradigm of what it means to live well, and claims that 'living a life is *itself* a performance that demands skill, that it is the most comprehensive and important challenge we face, and that our critical interests consist in the achievement, events, and experiences that mean that we have met the challenge well'<sup>69</sup>.

To assess the success of a performance it is important to take into account the circumstances in which it was executed. We need to know certain things about the challenge's conditions in order to decide whether it was addressed with skill and mastery. Circumstances, however, vary greatly in importance depending on the challenge we face. Some, Dworkin claims, need to be considered as 'limitations' – others are 'parameters'. Limitations are circumstances that do not figure in the description of the challenge people encounter. Examples of limitations are languages, talents, wealth, etc<sup>70</sup>. Essential for a circumstance to be limitation is that it is an aid or a constraint on acting as we have a reason to, but it is not a component of the description of that reason<sup>71</sup>. My life would go worse, for instance, if I did not

---

<sup>68</sup> Ibid. 240.

<sup>69</sup> Ibid. 253. The model, as a conception of the good, has generated a number of controversies. Cf. Richard Arneson, 'Cracked Foundations of Liberal Equality', in Burley, J. (ed.) *Dworkin and his Critics* (Oxford: Blackwell, 2004), 79-98.

<sup>70</sup> As we will see later, wealth is a limitation *only* if it does not contradict what justice allows. Dworkin concedes that the same item might be a limitation for one person, but a parameter to another; this of course depends on the sort of life each person leads.

<sup>71</sup> Cf. Matthew Clayton, 'A Puzzle about Ethics, Justice and the Sacred' in Burley, J. (ed.) *Dworkin and his Critics*, (Oxford: Blackwell), 102.

have some understanding of Borges's poetry (given I think that reading Borges belongs to my critical, rather than to my volitional, well-being), but, when I define the challenge of my life, reading Borges does not enter into the definition.

By contrast, there are other circumstances which necessarily are components in the definition of a challenge. Imagine Roger, a deeply religious person. For him, his faith is not an aid or a limit for what his life means. When he 'decides' what challenge his life presents, religion occupies an essential role. For him, then, religion is a parameter of the good life. He cannot define his life independently from religious practices. Consider what Dworkin writes regarding such instances: '[m]any of our parameters are normative: they define our ethical situation not in terms of our actual situation but of our situation as it should be. Our lives might go badly [...] not just because we are unwilling or unable properly to respond to the circumstances we have, but because we have the wrong circumstances'<sup>72</sup>.

There are two fundamental parameters in relation to the model of challenge: one is the principle of 'ethical integrity'; the other is justice. The former includes the idea that for an achievement to contribute to the ethical value of a person's life, the person himself must value it. When we consider the relation between a person's life, and his attitudes towards it, we can distinguish two different views. On the 'additive' view, the value of a life is independent of the attitudes the agent has towards it. If the person endorses the value he has created, that fact contributes to it having greater value still. If he does not, then, its value remains the same<sup>73</sup>. On the 'constitutive' view, by contrast, it is necessary for a person to identify with the value he has created if it is to contribute to his well-being at all. In Dworkin's words, this view

---

<sup>72</sup> Dworkin, *Sovereign Virtue*, 261-2.

<sup>73</sup> It seems that under the additive view, attitudes towards one's life are not really parameters but limits. If my life can go well without my endorsement, but even better with it, then endorsement does not figure in the definition of living my life well.



rejects that 'some event or achievement can make a person's life better against his opinion that it does not'<sup>74</sup>. This, of course, does not imply that the achievement loses its value all things considered; it might still be valuable to *us* and improve *our* lives. But it will not improve *his* life. Imagine Will, the mathematician genius, who advances maths to an incredible degree, but who also sees no value in this. He thinks that maths is so easy and simple that all his discoveries are just pure mediocrity. On the constitutive view his life is not improved by doing maths, until he reckons maths to be valuable.

Identification, as described in Chapter 2, is clearly insufficient for integrity. Recall Tom, the hard-working student who identifies with his obsessive hard work but became so committed because his mother told him that achievement at school is the only source of self-respect. Identification is necessary, but not sufficient for ethical integrity; Dworkin insists that the latter requires some degree of rational revisability and critical reflection<sup>75</sup>. The principle of ethical integrity combines, then, a subjective and an objective aspect of living a good life: on the one hand, a life goes better only if an individual endorses her activities and plans; on the other, ethical integrity matters only in conjunction with *critical* well-being. Someone who pursues only her volitional interests cannot live a good life no matter how much she identifies with them. In this respect endorsement is necessary for ethical integrity because the latter 'also requires the individual himself to identify, or choose, the central goals, projects, and relationships which make up his life'<sup>76</sup>. For Dworkin, ethical integrity 'make[s] the merger of life and conviction a parameter of ethical

---

<sup>74</sup> Dworkin, *Sovereign Virtue*, 249.

<sup>75</sup> *Ibid.* 271.

<sup>76</sup> Matthew Clayton, 'Liberal Equality and Ethics'. *Ethics*, 113 (2002), 15.

success, and stipulate[s] that a life that never achieves that kind of integrity cannot be critically better for someone to lead than a life that does'<sup>77</sup>.

The second fundamental parameter of the model of challenge is justice. Dworkin argues that 'if living well means responding in the right way to the right challenge, then a life goes worse when the right challenge cannot be faced'<sup>78</sup>. If we accept the model of challenge, then it is difficult to refute that the distribution of resources and liberties must be a normative component of the way in which the challenge is defined. In this respect, people who have less or more than is required by justice lead, *ipso facto*, worse lives than they would have done if they had the proper allocation of resources: justice is not only a moral concept, but an ethical one too<sup>79</sup>.

#### IV

### Hate Speech, Dignity and Integrity

#### *An Internal Conflict*

To what extent, then, is hate speech addressed by the ideas of dignity and ethical integrity? Remember that Dworkin has produced two arguments. First, he thinks that the principle of special responsibility demands that the government cannot curtail the expression of any view (unless there is a severe and imminent danger to third

---

<sup>77</sup> Dworkin, *Sovereign Virtue*, 270.

<sup>78</sup> Dworkin, 'Foundations of Liberal Equality', 260.

<sup>79</sup> Dworkin argues that resources are parameters rather than limitations 'because we cannot describe the challenge of living well without making some assumptions about the resources a good life should have available to it. [...] We must [...] find some suitable account of the way in which resources enter ethics as parameters of the good life, and we have, [...] no alternative but to bring justice into that story by stipulating that a good life is a life suitable to circumstances in which resources are justly distributed'. *Ibid.* 259.

parties' security). Restricting expression is a violation of dignity. Second, he thinks that living a good life consists in encountering the right challenge, and that restrictions on speech create distortions on the challenge a person faces. Free speech is, due to its intimate connection with conviction, a parameter of living well; if someone is not able to speak her mind due to governmental restrictions her life goes, for that reason (and independently of the consequences), worse.

The argument I defend here claims that both dignity and integrity fail to properly accommodate hate speech. In other words, the characterisation of these notions given by Dworkin is compatible with the claim that speakers are not wronged when government restricts the expression of hate speech.

The idea of reciprocity is connected with the idea of dignity in a critical sense. We established that dignity has two dimensions: personal responsibility and the fact that human life has intrinsic value. One of the conclusions of Dworkin's argument regarding the latter is that there is no justification for categorical exclusions; believing that your life has objective value implies that every other person's life has the same intrinsic worth. If this is so, 'then it is impossible to separate self-respect from respect for the importance of the lives of others. You cannot act in a way that denies the intrinsic importance of any human life without an insult to your own dignity'<sup>80</sup>. From this it does not follow that people can legitimately be coerced into not insulting their own dignity; this is a form of paternalism that Dworkin would reject.

It is also the case, however, that dignity cannot be employed as a defence for *hate* speech. If a person loses her dignity by expressing racist views, then she cannot use dignity as a justification for free speech. Given that Dworkin thinks dignity

---

<sup>80</sup> Dworkin, *Is Democracy Possible Here?*, 16.

requires you to recognise that other people's lives have equal intrinsic value, if you deny this, then your own dignity is compromised. So, a person who is prevented from harming his dignity (by committing harmful actions to third parties) cannot offer the dignity defence to reject the regulation because, had he been at liberty to do as he wished, his dignity would have been compromised then too. Dignity provides the basis for a qualified defence of free speech, one that requires free expression of the right view, or, as I will argue in the next chapter, expression restricted by a principle of reasonableness.

It is possible to recast the concerns of dignity and integrity in terms of audiences' interests. Because some forms of speech result in mental contamination, they jeopardise audiences' critical well-being. On the one hand, as noted above, a person might react in a fashion that is inconsistent with her beliefs about justice and morality: aversive racists who think that everybody must be treated with equal respect might display racist behaviour and attitudes due to influences in their environment which they cannot control. This is merely one of the possible ways in which the ecology of speech might compromise the integrity of citizens. On the other hand, it may also be that a government, by its failure to neutralise the effects of hate speech, might show less concern for some of its subjects: citizens who are susceptible to aversive racism might feel that the government could do more to protect them.

Dworkin could argue that the effects of speech, in terms of mental contamination, are similar to other impacts discrete actions have on people. As discussed earlier, according to equality of resources, we should not compensate Joe, a brilliant manufacturer of clepsydras who cannot sell them, because his talents are 'unmarketable'. Similarly, we should not compensate people who suffer from

listening to other people's ideas (not matter how much they dislike them). As we have seen, though, disadvantages created by prejudice and those created by having non-sellable talents are different. Equality of resources is neutral only towards the latter but it cannot be towards the former.

Hate speech is an injustice, then, in three respects. First, it fails to satisfy the minimal threshold of reciprocity required for political equality. Second, it has harmful effects on audiences by, for example, triggering processes of mental contamination which affect their ethical integrity. And third, these processes result from attitudes that liberal equality rejects. As Dworkin recognises, '[c]ommercial advertisement and soap operas create images that are subtle and ubiquitous, and it would not be surprising to learn, through whatever research might establish this, that they indeed do great damage to the way women are understood and allowed to be influential in politics'<sup>81</sup>.

However, it now seems that we have a *prima facie* conflict between the integrity of audiences and the integrity of speakers, for we have seen that a fully integrated life requires expressive freedom. The conflict is apparent: Dworkin thinks that living well requires we face the right challenge. A person whose life is not integrated because of government's coercive action leads a worse life than he otherwise would. But someone who lives in an unjust society<sup>82</sup> also leads a worse life for that reason. Now consider Dworkin's comments about Hitler: 'Of course it would've been better for everyone else if Hitler had died in his cradle. But on the challenge view it makes no sense to say that his life would have been better, as

---

<sup>81</sup> Dworkin, *Freedom's Law*, 220

<sup>82</sup> Notice that the parameter of justice applies to both people who commit injustices and people who live in unjust societies through no fault of their own. Although it is probably worse to actively create injustice than to be part of an unjust society, because justice is a parameter and not a limitation it is not clear how to 'balance' these two situations.

*distinct from no worse*, if that had happened'<sup>83</sup>. This is so because both ethical integrity and justice are necessary conditions of living a good life<sup>84</sup>. One cannot compensate a lack of integrity with more justice, or viceversa: if one fails, well-being also fails. 'Trade-offs do not arise'<sup>85</sup>.

Notice that integrity reflects critical well-being, rather than 'volitional' well-being. Someone who thinks that he needs more resources than he has is not entitled to more (at least for that reason)<sup>86</sup>. Whether or not he is given more depends on an objective account of justice. 'Someone has achieved ethical integrity [...] when he lives out of the conviction that his life, in its central features, is an appropriate one, that no other life he might live be a plainly better response to the parameters of his ethical situation *rightly judged*'<sup>87</sup>. In similar fashion, someone who claims that not being allowed to express his views regarding racial superiority makes his life worse has no claim of justice. The question is not how many resources you believe you should have according to your convictions, but how many you should have according to justice 'rightly judged'<sup>88</sup>. Obviously someone who claims that theft is a constitutive component of his doctrine of the good has no right to steal things, not

---

<sup>83</sup> Dworkin, *Sovereign Virtue*, 268. Emphasis added. See also Clayton, 'A Puzzle about Ethics, Justice and the Sacred', 104.

<sup>84</sup> Dworkin thinks that justice and ethical integrity are both soft parameters: 'Most of us believe that parameters that define the success of a life are all soft' *Sovereign Virtue*, 263. The distinction between soft and hard parameters is that, whereas the latter 'are violated the performance is a total failure, no matter how successful it is in other respects', any violation of the former is a serious, compromising flaw [...] that compromise is not fatal, and can be overcome'. *Sovereign Virtue*, 262. Dworkin is not clear about *how* soft the parameter of justice and ethical integrity are.

<sup>85</sup> Clayton, 'A Puzzle about Ethics, Justice and the Sacred', 104.

<sup>86</sup> Cf. Dworkin's arguments about equality of welfare and defence of equality of resources in *Sovereign Virtue*, Chapters 1 and 2.

<sup>87</sup> Dworkin, *Sovereign Virtue*, 270. Emphasis added.

<sup>88</sup> There seem to be two things going on here. Although Dworkin tries to defend the view I have just elaborated, there are strong criticisms as to what extent he has managed to provide a convincing argument. Clayton, 'Liberal Equality and Ethics'; Cohen, 'On the Currency of Egalitarian Justice', G.A. Cohen, 'Expensive Tastes Ride Again' in Burley J. (ed.) *Dworkin and his Critics* (Oxford: Blackwell, 2004). 3-29, claim that he has not. However, I do not need to go into such details here, my argument claims that Dworkin's defence of free speech fails, even if he is right in asserting that liberal equality follows from his account of ethics.

because abstaining from theft is better for him but because there are reasons of justice that allow governments to forbid such actions. As Matthew Clayton argues,

[b]ecause justice is a parameter of the good life [...], the liberal view that justice should be enforced by the government has a clear justification. Not only is justice an important ideal in its own right that warrants enforcement, its presence is also a requirement of people pursuing successful lives. So, to the extent that government has a duty to act with concern for its citizens, it must enforce justice as a necessary part of enabling people to face appropriate challenges in their lives<sup>89</sup>.

### *Avoiding Cultural Paternalism*

Dworkin argues that his account of ethical integrity blocks paternalism. He thinks that the latter is made self-defeating; you cannot make someone's life better by forcing him to do something of which he does not approve. Regardless of whether or not this argument is tenable, it is irrelevant for my purposes here because I am not claiming that speech regulations are better for racists; just that their lives would not go worse (because justice is a parameter of living well). Dworkin has the following example regarding critical well-being. Imagine Franz, who wants to be a priest. Imagine also that we believe that a religious life is worthless and that we are correct in this. There are three possible solutions: a) we convince Franz to become a cricket player (which is a better life), b) Franz decides to become a priest, or c) Franz becomes a very good cricket player but regrets it bitterly, he would rather have been a priest. Dworkin has no doubt that a) is better than b), and b) is better than c). In this case the only parameter involved is ethical integrity. Now imagine that instead the

---

<sup>89</sup> Clayton, 'A Puzzle about Ethics, Justice and the Sacred', 103.

decision concerned itself with justice (instead of the priesthood he wanted to be a murderer). In this case, option b) would not be worse than c) but it would also be no better (for him). And this is so because no life of injustice is a good life to live, regardless of its level of integration<sup>90</sup>.

Dworkin also thinks that respecting ethical integrity blocks, what he calls, 'cultural paternalism', which is defined as the assumption that 'a political majority has a right to create a culture it wants to live in and raise its children in not for the sake of the minority who might protest but for their own sakes'<sup>91</sup> Is the argument I am presenting culturally paternalistic? First, it is important to recall that the argument is motivated by a concern for the automatic effects of environments on autonomy. Second, and more important for Dworkin's argument, the regulation of speech is motivated by reasons of justice, not by concrete and controversial reasons regarding a conception of the good life. No one can complain about justice being done because justice is required to have a good life and to ensure everyone is treated with equal concern and respect. Regarding the public ecology of speech, it seems fair to make citizens bear some of the costs of maintaining an environment in which people can develop and exercise their sense of justice and autonomy. These costs are especially important when the effects of speech challenge the fundamental assumption of liberal equality both directly (when racists deny equal status to their victims) and indirectly (through processes of mental contamination and aversive racism). This point does not commit us to the view that we should *always* externalise the costs of people's actions<sup>92</sup>.

---

<sup>90</sup> Dworkin, *Sovereign Virtue*, 271-2

<sup>91</sup> Dworkin, R. (2006) *Is Democracy Possible Here?*, 74.

<sup>92</sup> Cf. Andrew Williams, 'Liberty, Equality, and Property', in Dryzek, J., B. Honig, and A. Philips (eds.) *The Oxford Handbook of Political Theory*. (Oxford: Oxford University Press, 2006), 501-3. He distinguishes two forms of sufficientarian egalitarianism. On the one hand, we can internalise the costs



For instance, one way in which the harmful effects of speech can be externalised is by employing ‘more, better speech’ to redress its costs. However, I will argue that this strategy is not attractive because racist speech bypasses rational control. On the other hand, we can internalise the costs of racist speech, by which I mean that non-racists should not bear the costs of racists’ speech. The decision between these two might be ecological in the sense that neutralising mental contamination recommends avoiding exposure to highly contaminating speech. Nevertheless, some might argue that, even though there are reasons to internalise the costs of hate speech, we should compensate racists for not allowing them to express their views. This objection fails insofar as there is no reason to externalise the costs of actions which are unjust. In this respect, racist preferences (and the costs they have on racists) are to be treated as expensive tastes, not as disabilities<sup>93</sup>.

### *Justice and Autonomy*

There are two possible objections to the argument defended in this chapter. First, one can follow Dworkin, who argues that the challenge of living a good life cannot be made more interesting or a better challenge by ‘bowdlerising’ some bad options. However, this claim is not obviously true: a person can think that the absence of certain choices has no detrimental impact on the challenge she encounters. She might deny that ‘our freedoms have value independently of the value we attach to the

---

of actions by limiting liberty thus protecting ‘individuals from bearing the costs arising from others’ decisions. On the other hand, we can externalise those costs making ‘defensible to force some individuals to bear the costs of others’ decisions if doing so is necessary to avoid absolute deprivation without restricting individual liberty’ (501-2).

<sup>93</sup> Racist preferences are similar to expensive tastes in the sense that justice recommends their costs be internalised. They are different in that pursuing expensive activities should not be restricted while racist speech should.

specific things they leave us free to do'<sup>94</sup>. For instance, a person who affirms the specific value of freedom thinks that having the option of playing football is valuable only in as much as playing football is. Being free to play football adds nothing to the activity in question. Similarly, this view seems to entail that by not being free to do something I would not do anyway my life is not worsened. A ban on fox-hunting or bullfighting does not make my life any worse<sup>95</sup>. True, this is a general point, which *only* applies to people who do not want to express racists views and who would not make use of hate speech protection. It clearly does not apply to racists, because a ban would certainly affect them. I do not want to argue here that freedom has only specific value. The aim of these remarks is merely to rebut Dworkin's suggestion that having some options eliminated is *per se* bad. The difference is ultimately dependent on the means by which choice is restricted. It is a bad thing if people's choices are manipulated but it is not necessarily negative if people decide publicly and democratically to eliminate some bad options. In the next chapter, I will defend a robust publicity condition for legitimacy and demonstrate that it is consistent with the regulation of hate speech. If freedom has only specific value, restrictions on hate speech affect only racists and not, as it is sometimes argued, the entire citizenry. Nevertheless, the way in which racists are affected does not make their lives worse, because justice is a parameter of leading a good life and the former, in possession of rights to free expression, cannot hold the latter.

---

<sup>94</sup> The distinction between the 'specific' and 'non-specific' value of freedom is developed in Ian Carter, *The Measure of Freedom* (Oxford: Oxford University Press, 2004), 33. Carter thinks that freedom has non-specific value and interprets Dworkin as a defender of the specific value of freedom. Cf. Charles Taylor, 'What's Wrong with Negative Liberty' in his *Philosophy and the Human Sciences*, 211-29.

<sup>95</sup> More contentiously, it could make my life better, by providing the means to do things I have an independent reason to do.

The second objection builds on the first one, but is more specific. It holds that an important component of dignity consists in having the opportunity to reject bad options. Accordingly, living with dignity requires that a person be exposed to racist ideas, and that she rejects them; that she comes 'clean'. If the objection is successful, restrictions on speech are in conflict with the dignity of non racists. The problem with this objection is that it overlooks the motivation behind the restrictions I support. The basis for such restrictions is not that those views are immoral or wrong, but rather that they affect us in non-autonomous ways: they contaminate our minds. In this respect, non racists have an autonomy-based interest in not being exposed to speech that bypasses their autonomy and threatens the effectiveness of their sense of justice. The regulations defended here are instrumental to autonomy in the sense that they are preferred over other 'remedies' (such as having more, and better speech<sup>96</sup>) because they are more likely to neutralise expression's harmful effects, and do so in a legitimate and just manner.

---

<sup>96</sup> Problems related to this strategy are discussed in Chapter 5.

## CHAPTER 5

### CONTENT REGULATIONS AND LIBERAL NEUTRALITY

In the last chapters I defended the claim that a concern for autonomy supports the regulation of certain kinds of speech, in particular, racist and sexist speech which tends to automatically reinforce social stereotypes. I argued that qua audiences we have an interest in avoiding exposure to highly contaminating speech and in having an appropriate environment in which to develop and exercise critical reflection. Although speakers' integrity might be threatened by regulations on speech, this creates only a *prima facie* problem; if justice is a parameter of having an integrated existence, then the life of racist speakers cannot be made worse by restrictions on hate speech.

In this chapter I address a different concern. I argue that content-based restrictions are compatible with a qualified version of political liberalism. This claim raises (at least) two sets of problems. On the one hand, defenders of free speech argue that content-based regulations are impermissible (in all but few exceptional cases). The strong presumption against this is motivated mainly by a concern for government neutrality. Accordingly, some might suggest that, because I am arguing for content-based restrictions on speech which expresses certain racist views, the

condition of neutrality should also be rejected. I cannot, it might be argued, have both.

The defence I will elaborate here avoids this charge by suggesting that the content-based regulations supported are not based on the intrinsic merits of the views regulated, but rather on the distorting effects that such acts of expression have on individuals' autonomy. Therefore they do not constitute a violation of liberal neutrality.

However, if this defence is successful a further problem arises: it might seem that I am covertly introducing a comprehensive doctrine beneath a smoke screen. The objection would allege that arguing for an autonomy-based liberalism is not neutral in the required sense. My conception of liberalism may seem as partisan as any other political theory based on comprehensive doctrines. People who raise this objection might say that my argument is a form of liberal perfectionism.

These two problems together appear to create a dilemma: either neutrality is rejected for the sake of content-based restrictions on speech or the condition of neutrality is maintained thus making content-based restrictions impermissible. The aim of this chapter is to demonstrate that the justification of content-based regulations suggested here avoids this dilemma.

## I

### The Problem of Content-Based Regulation

*Content-Based and Content-Neutral Regulations*

In American Free Speech doctrine Courts have accepted some forms of *content-neutral* restrictions on speech. Restrictions on the time, place and manner are usually assumed not to threaten fundamental interests in speech, while they minimise the costs that expressive activities have on third parties. Sometimes people have an important interest in not being exposed to speech (regardless of its content), and it seems that this is sufficient to demand its restriction. This interest is satisfied by regulating the time or places of the expressive act. It is acceptable to restrict rock concerts in residential areas at 2AM; or loud conversations in hospitals and libraries; to limit the number of parades and rallies as well as the amount of leafletting included in political campaigns. All of these restrictions pay no attention to the content of speech; they focus solely on the context in which expression takes place.

On the other hand, courts usually have a strong presumption against *content-based* regulation. For instance, banning a U2 rock concert because of the (lack of) quality music or restricting the sales of *The Communist Manifesto* because of the dangerous ideas it defends are clear examples of content regulations. Sometimes a further distinction is made between subject matter and viewpoint discrimination within content-based regulations: a law prohibiting discussion of the *virtues* of abortion is viewpoint discrimination, while a law banning discussion of abortion *tout court* is an issue of subject matter. Although this distinction is somewhat controversial, it is usually assumed that, from the perspective of free speech, viewpoint discrimination is worse than regulation of subject matter. However, the difference between the two is not as clear as we might hope. Content-based regulations might be acceptable in some contexts but not others. Think for example

about a law forbidding anti-abortion protests outside a clinic. Although this law is content-based (and for that matter, viewpoint-based) insofar as it targets demonstrations on the wrongs of abortion, in many cases it would be acceptable<sup>1</sup>. This law does not outlaw football fans cheering for their team, nor Anti-War protesters. Now, an ordinance that restricted every anti-abortionist demonstration would certainly be considered illegitimate.

Although the presumption against content-based regulations is strong, it certainly is not absolute<sup>2</sup>. There are some categories of speech that Courts have been willing to prohibit based on their content, such as private libel<sup>3</sup>, direct incitement<sup>4</sup>, fighting words<sup>5</sup> and obscenity<sup>6</sup>. In order to be acceptable, these restrictions need to satisfy strict scrutiny<sup>7</sup>. It is important to notice that they are motivated less by a desire to prevent the expression of certain ideas and more by a desire to prevent

---

<sup>1</sup> It might be acceptable in terms of security; protesters might jeopardise the safety of the clinic's employees and women who attend it. A further condition might be that there are other options (other places and times) open to those who want to protest against abortion. See also Mill, *On Liberty*, 63.

<sup>2</sup> Cf. James Weinstein, *Hate Speech, Pornography and the Radical Attack on Free Speech Doctrine* (Boulder: Westview Press, 1999), 34-49.

<sup>3</sup> According to Dworkin, prohibition on libel was introduced not in order to restrict what is being said, but in order to allow 'offended citizens to restore their reputations. Dworkin, *Freedom's Law*, 198.

<sup>4</sup> *Brandenburg v. Ohio*, 395 U.S. 444 (1969).

<sup>5</sup> This category 'refers to [expression] that is likely to provoke an immediate violent reaction by the persons to whom the words are addressed and that also conveys little by way of ideas and thus makes only the most limited contribution to public debate'. Fiss, *Liberalism Divided*, 113.

<sup>6</sup> As defined in *Miller v. California*, 413 U.S. 15 (1973). 'A work may be subject to state regulation where that work, taken as a whole, appeals to the prurient interest in sex; portrays, in a patently offensive way, sexual conduct specifically defined by the applicable state law; and, taken as a whole, does not have serious literary, artistic, political, or scientific value'. These three conditions are necessary. See *Miller v. California* 413 U.S. 15 (1972). The test is not uncontested, though, and some people think it is unworkable. Cf. Justice Stewart's remark that 'he couldn't define obscenity but knew when he saw it' (quoted from Dworkin, *Freedom's Law*, 207). Dworkin is also sceptical about obscenity laws.

<sup>7</sup> 'Under the scrutiny test, the government must show that the law is necessary to serve a compelling state interest, and that it is narrowly drawn to achieve that end'. Weinstein, *Hate Speech, Pornography and the Radical Attack on Free Speech Doctrine*, 55.

imminent danger (in the case of incitement), to restore the reputation of offended citizens<sup>8</sup>, and so on.

In the case of fighting words, for instance, the problem is not that certain ideas are expressed, but the manner in which they are delivered. In terms of free speech doctrine, they are only noise. In 1990 Robert Viktora allegedly burned a cross on a black family's lawn. He was charged with violating an ordinance against bias motivated crimes in St. Paul, Minnesota<sup>9</sup>. The ordinance stated that:

Whoever places on public or private property a symbol, object, appellation, characterization or graffiti, including, but not limited to, a burning cross or Nazi swastika, which one knows or has reasonable grounds to know arouses anger, alarm or resentment in others on the basis of race, color, creed, religion or gender commits disorderly conduct and shall be guilty of a misdemeanor<sup>10</sup>.

The Court decided that the ordinance was invalid, because it was viewpoint oriented. Justice Scalia, who delivered the Court's opinion, argued that 'the reason why fighting words are categorically excluded from the protection of the First Amendment is not that their content communicates any particular idea, but that their content embodies a particularly intolerable (and socially unnecessary) mode of expressing whatever idea the speaker wishes to convey'<sup>11</sup>. Scalia objected that the ordinance does not treat each person fairly because it makes a special sub-class of fighting words (namely, fighting words targeting people on the basis of race, colour, creed, religion or gender) as deserving special punishment, while leaving other targets without protection. Someone using abusive language against bigots, for

---

<sup>8</sup> Dworkin, *Freedom's Law*, 198.

<sup>9</sup> *R.A.V. v. City of St. Paul, Minnesota* 505 U.S. 377 (1992).

<sup>10</sup> St. Paul Bias-Motivated Crime Ordinance. St. Paul. Minn., Legis. Code 292.02 (1990)

<sup>11</sup> *R.A.V. v. City of St. Paul, Minnesota* 505 U.S. 377 (1992).



instance, would be not breaking the law<sup>12</sup>. He concluded: ‘Should the government want to criminalize certain fighting words, the Court now requires it to criminalize all fighting words’<sup>13</sup>. Notwithstanding the difficulties of distinguishing between regulations that are content-based and content-neutral, it is usually assumed that the latter are acceptable insofar as accommodating competing interests is relatively simple. In this respect the special protection attached to speech is related primarily to its content rather than to its ‘manner, place and time’.

### *Justifications against Content-Based Regulations*

There are various reasons that motivate such a strong presumption against content regulation. First, there is a widely shared mistrust of government according to which if it has the chance of abusing its power it is likely to do so; if government was given the power to silence criticisms of different views, it would probably use it<sup>14</sup>. Giving the government power to regulate speech risks the possibility of a slippery slope which will eventually result in the government silencing its critics and creating official dogma. There is some truth about this ‘fact of power’: ‘Most people – and

---

<sup>12</sup> Scalia writes: ‘St. Paul has no authority to licence one side to fight free style, while requiring the other to follow Marquis of Queenbury Rules’.

<sup>13</sup> *R.A.V. v. City of St. Paul, Minnesota* 505 U.S. 377 (1992). Similarly, the Indianapolis ordinance on pornography, suggested by C. MacKinnon and others was struck down because it *only* condemned sexually explicit material that presented women as enjoying pain or humiliation or rape (among other things. See Mackinnon’s definition of pornography in her *Only Words* (Cambridge, MA: Harvard University Press, 1993) 121-2 n.32. Non-sexist sexually explicit material would not be outlawed by the ordinance.

<sup>14</sup> An example of this temptation is McCarthyism in the U.S. See also Scanlon, ‘Content Regulation Reconsidered’, 156.

those with power in particular— do not like being criticised or disagreed with and are tempted to use the means at their disposal to avoid such possibilities'<sup>15</sup>.

Second, there is a concern regarding the proper relationship between autonomous citizens and government according to which the former would allow the latter to decide for them what forms of expression they should be able to hear. As the Millian Principle indicates, giving the government the power to make those decisions is incompatible with regarding oneself as a free and autonomous agent<sup>16</sup>. Similarly, as we saw in Chapter 4, Dworkin argues that people are not treated with respect if they are not trusted to decide for themselves about ideas which the government finds offensive or dangerous<sup>17</sup>.

Finally, the most direct rationale for holding content-regulations illegitimate is that governments must remain neutral towards different the conceptions of the good and ideas expressed in society<sup>18</sup>. Government neutrality is derived from a view of what it is for a state to treat everyone with equal respect and concern. According to this view, if the state bases its decisions on the relative merits or intrinsic worthiness of certain views, it will fail to treat people with equal fairness<sup>19</sup>. The standard model of this argument holds that the proper role of government is to guarantee the marketplace of ideas (to use a favourite metaphor of free speech

---

<sup>15</sup> Cohen, 'Freedom of Expression', 233.

<sup>16</sup> Cf. Scanlon, 'Freedom of Expression'; see Chapter 3 for an assessment on this claim.

<sup>17</sup> 'Because we are a liberal society committed to individual moral responsibility [...] any censorship on grounds of content is inconsistent with that commitment' Dworkin, *Freedom's Law*, 205. Emphasis added.

<sup>18</sup> The way in which I believe liberal neutrality should be understood will follow below.

<sup>19</sup> Cf. Ronald Dworkin, 'Liberalism' in Hampshire S. (ed.) *Public and Private Morality* (Cambridge: Cambridge University Press, 1978), 127.

advocates) and not to show bias with any particular view<sup>20</sup>. It is '[t]he people, not the state [who] should choose among competing viewpoints and their choice should not be manipulated by the state by skewing public debate in some special way'<sup>21</sup>. The state, in this view, must refrain from making any judgement about the value, worthiness or truth of different forms of expression<sup>22</sup>. Neutral states, committed to avoiding restrictions on speech's content, however, need not fail to recognise the costs of speech. The usual strategy consists in *correcting* the wrongs speech produces by responding with 'more, better speech'<sup>23</sup>.

### *The Liberal Solution to the Costs of Speech*

It is sometimes suggested that some of the wrongs caused by speech can be corrected by providing public fora in which different views can be expressed. In discussions of hate speech, it is common to advocate a system where the state provides public spaces in which other views can be heard. Because the state recognises that access to media (and to the marketplace of ideas) is unequally distributed, it can subsidise the expression of disadvantaged groups<sup>24</sup>. An active state is the only *legitimate* way of combating the 'silencing' effects of social inequalities<sup>25</sup>. The hope is that, by

---

<sup>20</sup> Cf. 'Freedom of expression requires evaluative neutrality; granting free speech to those with whom the government agrees is not freedom of expression'. Alexander, *Is There a Right to Freedom of Expression?*, 148.

<sup>21</sup> Fiss, *The Irony of Free Speech*, 40. Cf. also *FCC v. Pacifica Foundation*, 438 U.S. 726 (1978): 'For it is a central tenet of the First Amendment must remain neutral in the marketplace of ideals'.

<sup>22</sup> See for instance Alexander's analysis of 'low-value' speech. Larry Alexander, 'Low Value Speech', *Northwestern University Law Review*, 83 (1989), 547-54.

<sup>23</sup> Cf. Justice Brandeis's remark: 'if there be time to expose through discussion the falsehood and fallacies, to avert the evil by the process of education, the remedy to be upheld is more speech, not enforced silence'. *Whitney v. California*, 274 U.S. 357 (1927).

<sup>24</sup> Although there is debate about the extent in which government has a positive duty to guarantee access to public opinion.

<sup>25</sup> For a defence of an active state (although sympathetic to content-based regulations) see Fiss, *The Irony of Free Speech*. See also Scanlon, 'Freedom of Expression and Categories of Expression', 100.

providing fair access to expression disadvantaged voices will be heard, proper debate will take place and people will be convinced by the best reasons.

A variant of this strategy suggests a system of warnings, in which audiences are alerted to the costs of the expression to which they will be exposed before it actually happens. Imagine a person living in a sexist society who, after exposure to pornography, has concluded that, in general, women enjoy rape. In order to correct this view, some liberals recommend debates with feminists be organised, in which the aim is to convince the sexists and pornographers that that have formed the wrong beliefs. After rational consideration, some (hopefully most) people will change their minds and will reject the sexist beliefs. Similarly, the state can impose a warning on pornographic material stating that ‘exposure to this material may persuade you that women enjoy sexual violence’.

How likely is this strategy to succeed in eliminating sexist, racist and homophobic beliefs? In order to succeed, the satisfaction of (at least) two conditions is necessary. First, people must be motivated to change their minds: they must be predisposed to rational scrutiny and critical reflection. Second, they must be capable of altering their beliefs when there are good reasons to do so. Joshua Cohen believes this is the case. He lists a set of background anthropological and psychological facts, among which the ‘Fact of Reasonable Persuasion’ is important:

‘People have the *capacity* to change their minds when they hear reasons presented, and sometimes they *exercise* that capacity’,<sup>26</sup>.

---

<sup>26</sup> Cohen, ‘Freedom of Expression’, 233. Emphasis added.

*The Dynamics of Belief*

Liberals tend to accept the ‘fact of reasonable persuasion’ uncritically, despite research in psychology which has shown we should be less confident of it. Evidence suggests that, although we have the capacity to change our beliefs, we tend not to exercise it as often as we might hope. As Susan Brison writes, many of the free speech doctrines presuppose ‘that one can always choose to accept or reject an ideal, but some ideas are contagious, insinuating themselves into our thoughts without –or even against— our will’<sup>27</sup>.

There are two competing models which attempt to explain how mental systems believe. According to the ‘standard’ Cartesian framework<sup>28</sup>, belief formation consists of two parts. First we understand the content of propositions and later we either accept the content as true or reject it as false. The second process is allegedly independent of the first. According to this model we have a reasonable amount of control over what we believe or disbelieve. Schematically the first model can be represented as:

Exposure to information → Understanding → Rejection or Acceptance.

---

<sup>27</sup> Susan Brison, ‘Review of Alan Haworth’s *Free Speech*’ *Mind*, 113 (2004), 450, 354. For a controversial discussion of the contagiousness on ideas see Richard Dawkins, *The Selfish Gene* (Oxford: Oxford University Press, 1989); Susan Blackmore, *The Meme Machine* (Oxford: Oxford University Press, 1999); Daniel Dennett, *Darwin’s Dangerous Idea* (London: Penguin, 1995); Sperber, *Explaining Culture: A Naturalistic Approach*.

<sup>28</sup> This model has its main inspiration in René Descartes. Cf. Daniel Gilbert, ‘The Assent of Man: Mental Representation and the Control of Belief’ in Wegner, D. and J. Pennebaker (eds.) *Handbook of Mental Control* (Englewood Cliffs: Prentice Hall, 1993), 57-87; Gilbert, ‘How Mental Systems Believe’; Daniel Gilbert, Romin Tafari and Patrick Malone, ‘You Can’t Not Believe Everything You Read’ *Journal of Personality and Social Psychology* 65 (1993), 221-33.

The step between exposure and understanding is passive; we cannot control whether or not we understand a proposition. When we hear something we simply understand it. According to this model it is the second, crucial, step linking the understanding of a proposition to believing it (or not believing it) that is assumed to require an active component of cognition<sup>29</sup>. Without an act of will, we should remain neutral towards the veracity of the proposition.

There is, however, an alternative, ‘Spinozan’ model which suggests understanding and believing propositions are only one process<sup>30</sup>. According to this model, when we understand a proposition we assume it is true. Later, if needed, a different process might be activated in order to reject the proposition. This can be represented as:

Exposure to information → Understanding and Acceptance → Confirmation or Rejection.

The second of these models involves two separate claims

**The Asymmetry Hypothesis.** The first is that acceptance and rejection occur at different times; first comes acceptance and later rejection<sup>31</sup>. The argument is that the denial of propositions is an addendum to information. For instance, the representation of ‘Bob is not an extrovert’ involves both the representation of ‘Bob is

---

<sup>29</sup> Contrast the similarity between the necessity of an act of cognition between understanding and acceptance/rejection on the one hand, and the ‘decoupling’ model which asserts that there is a necessary act of mediation between beliefs/desires and actions. See Chapters 1 and 2.

<sup>30</sup> According to Gilbert, ‘How Mental Systems Believe’ this model was devised by Baruch Spinoza.

<sup>31</sup> Gilbert, ‘How Mental Systems Believe’, 113.

an extrovert' and its denial. This way of understanding denials explains some cognitive biases such as 'denial transparency', the process by which 'on encountering denied information, the person typically has that information available for processing despite the denial', rather than discarding the false information<sup>32</sup>. For instance, in an experiment, some individuals were presented with the sentence 'Bob Talbert is not linked with Mafia'. Other persons were presented with the sentence 'Bob Talbert arrives in City'. Later they had to report the impression they had formed of Bob Talbert. The results show that people in the first group developed a less favourable impression than people in the second group. Notice, that there is no reason to form a negative impression of someone not being linked with Mafia. Wegner and his colleagues suggest the mechanism operating is that individuals first represented 'Bob Talbert is linked with Mafia' and then just added the denial<sup>33</sup>. When they had to assess Bob Talbert the information about him being linked with Mafia (although false) still affected their judgement. A computer would have reacted rather differently: it would have erased the information.

**The Unity Hypothesis.** The second claim (which is more relevant for our purposes) holds that the act of understanding a proposition and that of believing it are in fact the same process. Accordingly, each time we are exposed to a proposition and we understand it, we also believe it. This is an automatic, passive process. Later, we return to the proposition and either confirm or disbelieve it. This second step, in

---

<sup>32</sup> Daniel Wegner, Gary Coulton and Richard Wenzlaff, 'The Transparency of Denial: Briefing in the Debriefing Paradigm', *Journal of Personality and Social Psychology* 49 (1993), 340.

<sup>33</sup> The experiment is reported in Daniel Wegner, et. al., 'Incrimination Through Innuendo: Can Media Questions Become Public Answers?' *Journal of Personality and Social Psychology* 40 (1981), 822-32. The results are also discussed in Wegner, Coulton and Wenzlaff, 'The Transparency of Denial'. Contrast this phenomenon with the 'anchoring bias' in Kahneman Tversky, 'Judgment under Uncertainty'.

contrast, is not automatic but requires a cognitive effort: in order to be triggered, it requires mental resources. The model predicts that under certain circumstances the process of undoing the acceptance of propositions might malfunction. This happens when there is a depletion of the mental resources needed for revising the belief. The prediction is that ‘when some event prevents a person from “undoing” his or her initial acceptance, then he or she should continue to believe the assertion, even when it is patently false’<sup>34</sup>. In contrast, the standard, dualist model predicts that under the same circumstances people will merely understand the proposition, they will neither believe nor disbelieve it.

One of the experiments described in Gilbert et. al. (1993) investigates this possibility. People were asked to read aloud a pair of crime reports which would ‘crawl’ on the screen of a computer<sup>35</sup>. Participants were told that the reports contained both false and true statements; the former appeared in red, the latter in black. In one report, the false statements exacerbated the gravity of the crime; in the other they extenuated it. Some of the subjects were also asked to perform a digit search task. They were told that while they were reading the reports digits would appear in a line under the text. They were asked to press a button every time the number ‘5’ appeared –this group was called the ‘interrupted condition’. Unbeknownst to the participants, digits *only* appeared in conjunction with false statements (the interrupted condition was designed to demonstrate that people might treat explicitly false information as true when they are cognitively loaded). After reading the reports, the participants were 1) asked to recommend prison sentences for the crimes (between 0 and 20 years), and 2) shown 30 statements (4 true, 7 false

---

<sup>34</sup> Gilbert et. al., ‘You Can’t Not Believe Everything You Read’, 222.

<sup>35</sup> As football results ‘crawl’ on BBC One.



and 19 coils), which they had to classify as being 'true', 'false' or 'never shown' in the first report, and the same for the second.

In the case of the prison sentences, people who read false statements extenuating the crime recommended an average of 5.83 years in prison (the control group –those who did not perform the digit search recommended 6.03 years). People who read false statements exacerbating the crime recommended 11.15 years of sentence (the control group recommended only 7.03). In the interrupted condition the difference in years recommended caused by false statements was 5.32 years. In the recognition task the results show that although interrupted and uninterrupted subjects were equally likely to misremember true statements (the proportion was .06 and .03 respectively), those in the interrupted group were more likely to treat false statements as true (0.44 and 0.23 respectively). In fact, people in the interrupted condition identified only 34% of false statements, and misidentified 44% as true<sup>36</sup>.

The results of this experiment<sup>37</sup> suggest that we have a Spinozan belief system. Only if we directly believe what we understand would we display the tendency to treat explicitly false information as true under cognitive pressure<sup>38</sup>. Perhaps the unity hypothesis is too strong. Perhaps there is a difference between understanding and accepting a proposition, and Gilbert draws a conclusion which overestimates his evidence. Nevertheless, a weaker claim seems well supported: the evidence shows that acceptance is significantly easier than rejection and takes fewer cognitive resources. This qualified Spinozan version asserts that the default state of a person's mind is inclined towards acceptance, and that, under cognitive pressure, a

---

<sup>36</sup> Gilbert et. al., 'You Can't Not Believe Everything You Read'. Experiment 1, 223-6.

<sup>37</sup> See also experiments 2 and 3 in the same paper.

<sup>38</sup> Interrupted subjects mistreat false statements as true (.44), but not true as false (.06)

person will tend to believe whatever she is exposed to, even if she knows the information to be false.

It is important to note, in particular, that belief and action are directly linked. People in the interrupted condition not only believed false statements; they also behaved accordingly. For instance, they recommended tougher prison sentences based on false ideas than people in the control condition. Notice, however, that the differences in the two groups are independent of the subjects' prior beliefs and their intellectual capacities. They are due to environmental features which thwarted their critical capacities. It was because they were occupied with searching for digits that they stopped disbelieving the information they knew was false. Even in cases in which people are not busy, they do not use the opportunity to abandon false information. Instead they store it, and sometimes it permeates their judgements. In general, people are gullible and credulous; only on rare occasions they are critical. This dynamic of belief can, in certain circumstances, lead to cases of mental contamination.

#### *'More, better speech' and Mental Contamination*

Another important cognitive bias which challenges Cohen's 'fact of reasonable persuasion' is 'confirmation bias'. In general, people have the tendency to assess empirical evidence in a biased manner. When participants in an experiment were asked to evaluate reports on the deterrent impact of the death penalty on crime they made their assessment in accordance with their prior beliefs on the effectiveness of

capital punishment<sup>39</sup>. Participants were shown two different reports on the death penalty. Those were fairly general and inconclusive. One presented evidence favouring the deterrence-effect; the other offered evidence suggesting that capital punishment has no impact on crime. Those who held a prior belief that the death penalty has deterrent effects interpreted the first report as confirming their views. Moreover, they claimed to have confidence that the study upon which the report was based was scientifically strong and had merit. On the other hand, they were not impressed by the second report; they doubted the methodology, questioned the objectivity of the study and so on. Not surprisingly, people who originally thought that the death penalty is not effective in preventing crime reacted in the opposite fashion. Moreover, people in the experiment also showed a ‘polarisation bias’: those who supported the death penalty before the experiment became more convinced of its deterrent effects. And those who opposed it, objected to it more strongly. The experiment demonstrates that people tend to select only those pieces of information which are consistent with their prior views and tend to disregard information that contradicts them. They do not critically evaluate the information with which they are presented.

These experiments pose a serious problem for the ‘more, better speech’ strategy. It suggests that people can exercise large amount of control over what they believe, and that they evaluate evidence and information impartially. But people cannot (to the required degree) select their convictions in such a neutral manner.

---

<sup>39</sup> Charles Lord, Lee Ross and Mark Lepper, ‘Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence’ *Journal of Personality and Social Psychology* 37, (1979), 2098-2109.

Environmental conditions have a strong influence, and people arrive at beliefs by means they believe to be inadequate.

We can distinguish two levels at which the fact of reasonable persuasion operates. One is the rational, conscious tier which is the target of strategies such as ‘more, better speech’. The other is an automatic unconscious level: as we have seen, people store false information that is later used in judgments, people can mistake false information for truth and use this as a reason for action (recommending prison sentences, for instance), and they also tend to interpret information in a selective, biased manner. Yet, they are not aware of this happening: these processes are not ‘felt’. So, while we can have ‘more, better speech’ at one level, there may be environmental features which trigger the above processes and jeopardise the benefits that rational debate is supposed to deliver. Moreover, ‘more, better speech’ is also limited because people often do not accept that they are susceptible to these cognitive biases. They are sure they can control their own minds, what things they believe, and what causes their actions.

Open debate is a beneficial phenomenon, but might not always be effective, especially in cases where processes bypass rational control, such as automatic processes, implicit cognition and social stereotypes. ‘More, better speech’ can help ‘if the additional speech provides people with scientific evidence that their behaviour is actually subject to such direct, automatic [...] influences [...] so that they can take steps to guard such influences then perhaps the additional speech will have some tendency to mitigate the harmful effects’<sup>40</sup>. As we have seen, however, leaving people to counteract the effects of mental contamination on their own presents

---

<sup>40</sup> Hurley, ‘Bypassing Conscious Control’, 322-3.

serious difficulties. People tend to underestimate their susceptibility to mental contamination and to overestimate their corrective powers. Even if they recognise their limitations they might be constrained in other sense, perhaps they lack the motivation to neutralise those influences<sup>41</sup>. Instead, I suggest, avoiding exposure to the sources of mental contamination would be a more effective strategy.

Raising awareness of the biases we sometimes display is, as I have argued, a condition for a just society. It fails, however, in trying to undo the damage of certain forms of speech. It can have *preventive* value—warning people about the processes that might challenge their autonomy—but not a great amount of *corrective* value. Despite this, the arguments developed here surely do not reject the fact of reasonable persuasion; they merely qualify it. Essentially, although people have the capacity to change their minds when they are presented with reasons, and sometimes they exercise that capacity, they also have the tendency to interpret new evidence in a biased way, under cognitive pressure they regard information they should know to be false as being true, and they get anchored with false information, which is later used to make judgements. We need to be more careful in our application of the ‘more, better speech’ strategy and be aware of its ecological limitations.

### *Autonomy and Content-Based Regulations*

The argument I defend holds that because autonomy is inconsistent with mental contamination, there are valid reasons to restrict freedom of speech in sensitive

---

<sup>41</sup> Just as some people choose to smoke even when they know (and accept) that smoking harms third parties as well as them, some people might agree that they would be influenced in harmful ways and still do nothing about it. Cf. Ibid. 323.

areas, such as racist and sexist issues. These restrictions are not motivated by an assessment of the internal merits of the views proposed but rather by the effects they have on autonomy. In this respect I make two arguments. On the one hand, I note that in a strong understanding of free speech autonomy is often employed to defend speech, despite the harm the former can cause. But, in fact, because autonomy is bypassed it cannot play the exempting role. On the other hand, I argue that in order to achieve autonomy certain external conditions must be satisfied, amongst which is a social environment in which critical deliberation is facilitated. Protection of hate speech can impair the ecological conditions of autonomy<sup>42</sup>. In this sense, the restrictions suggested are autonomy-enabling rather than autonomy-disabling<sup>43</sup>. This second argument seems (*prima facie*) much more perfectionist than the first, insofar as it bestows autonomy with a considerable weight within a consequentialist structure. The rejection of this claim will occupy the next sections.

Although restricting racist speech is a necessary condition of accomplishing the autonomy-enabling environment I argue for, clearly it is not sufficient; as many liberals argue, governments must still take active steps in order to distribute access to media and public fora fairly. These steps might involve establishing expenditure limits to political campaigns, providing public funding to political parties, subsidising the expression of minority or unpopular views, and so on<sup>44</sup>.

There is also another worry I wish to address. I have argued that racist speech must be regulated not on an assessment of its content's merits, but due to its

---

<sup>42</sup> Remember Scanlon's remark that the most important interest that audiences have in free speech relates to having a good environment for the formation of beliefs. 'Freedom of Expression and Categories of Expression', 91.

<sup>43</sup> For what I have said so far it is clear that restrictions on hate speech must be broader than the usual categories of 'fighting words' or 'face-to-face vilification'.

<sup>44</sup> Arguments for this can be articulated from a worry about the quality of public debate (Fiss, *The Irony of Free Speech*) or the fair value of liberties (Rawls, *Political Liberalism*, 289-371).

results<sup>45</sup>. In this sense, it may be thought that the restrictions are not content-based. I think this captures half the truth. On the one hand, according to this interpretation, the argument I am presenting justifies only context-based restrictions, that is, it only applies to contexts in which those automatic processes are prone to be activated. But there may other contexts in which racist speech does not activate those processes. If this is so, then the restrictions are only context-dependent, rather than content-based. On the other hand, however, it seems that racist speech has the capacity to trigger those processes in virtue of its content. If the content was different, then the effects would not arise. Both are necessary conditions: there are certain contents expressed in certain contexts which threaten autonomy. Complicating matters further, we still know little about the relation between context and content; hard questions must be answered regarding the level of domain specificity and flexibility of our mental processes in order to devise a definitive answer.

The upshot of the argument is, then, that speech expressing a particular viewpoint (namely, racist and sexist speech) ought to be regulated not because its content is mistaken (although it is), but because it is likely to trigger mental contamination. The argument holds that certain features in the environment trigger automatic responses that bypass rational autonomy. Racist and sexist speech constitute important instances of mental contamination because we have a special propensity to be contaminated by those ideas<sup>46</sup>. Moreover, racist and sexist ideas have a significant role in the public ecology of our societies, affecting the manner in which people act. Although the remarks discussed in this section regarding the

---

<sup>45</sup> Not because the merits of the view it presents, but because the way in which it threatens autonomy.

<sup>46</sup> Cf. Chapter 1. At the end of that chapter I survey Hirschfeld cognitive theory of race, according to which we have a specific-domain cognitive module for 'human kinds', which is susceptible to racial categories. Cf. Hirschfeld, *Race in the Making*.

‘more, better speech’ strategy apply to propositions in general, when combined with instances of hate speech they pose a serious threat to citizens’ moral powers, in particular, as I have suggested, to their capacity for a sense of justice.

Nevertheless, there are good reasons to believe that some spill-over effects will take place<sup>47</sup>. For instance, imagine the discussion of racist science in academic fora. Academic debate seems to be a good environment in which to form beliefs: there are not many distractions, people can focus on only one thing, discussion is supposed to be clear, rational, direct, free from manipulation, and so on. But it is important to know whether the effects of those debates can ‘spillover’ to other spheres of life that do not satisfy those environmental conditions<sup>48</sup>.

## II

### Legitimacy, Autonomy, and Anti-Perfectionism

So, is the argument defended perfectionist? Some might argue that although I have rejected a version of perfectionism (one which argues speech might be regulated on an assessment of its merits), I have covertly instilled another: a version of perfectionism which attaches overriding value to personal autonomy. The claim defended, some might argue, violates neutrality insofar as the value of an autonomous life is a partisan idea, with which reasonable people can disagree. So,

---

<sup>47</sup> In Chapter 6, problems of spillover effects are discussed in relation to Nancy Rosenblum’s ‘Moral Uses of Pluralism’. Cf. Nancy Rosenblum, *Membership and Morals: The Personal Uses of Pluralism in America* (Princeton: Princeton University Press, 1998).

<sup>48</sup> On the other hand, maybe we might be willing to compromise part of the truth (e.g. in case there were any racial differences) for moral reasons. It may well be the case that morality overrides some truths.



while I have successfully avoided justifying restrictions on hate speech on the basis that racism is morally wrong, the argument is perfectionist insofar as it rests on an account of the value of personal autonomy.

The debate regarding liberal neutrality and perfectionism is an extensive one, and the response I develop will be necessarily sketchy. The term ‘neutrality’ itself has ambiguities and misunderstandings, and it is therefore not a very useful one. Nevertheless, I will abide with its use to follow the conventions of related literature. Another difficulty is the number of misunderstandings and ‘straw-men’ invoked in typical discussions, which obscure the merits of the positions debated. In what follows I will comment briefly on what liberal neutrality is *not*. Later I will sketch a view about what perfectionism (properly understood) *is* concerned with. Finally, I will defend a restricted version of neutrality which grants a moderate concern for autonomy<sup>49</sup>.

### *What Neutrality Is and Is Not*

There are two significant interpretations of neutrality. One claims that liberal policies should affect everyone equally<sup>50</sup>. ‘Impact neutrality’ is the view that, because government must treat all its citizens as equals, its policies must also affect everyone’s life prospects equally. According to this version of neutrality, ‘[o]ne of the main goals of governmental authority, which is lexically prior to any other, is to ensure for all persons an equal ability to pursue in their lives and promote in their

---

<sup>49</sup> This position is heavily indebted to Raz, *The Morality of Freedom*; Joseph Raz, ‘Facing Up: A Reply’, *Southern California Law Review*, 62 (1989), 1153-235; Joseph Raz, *Practical Reason and Norms* (Oxford: Oxford University Press, 1990), and Clayton, *Justice and Legitimacy in Upbringing*.

<sup>50</sup> Cf. Raz, *The Morality of Freedom*, 115.

societies any ideal of the good of their choosing'<sup>51</sup>. Liberals, such as Rawls, Dworkin and Nagel, reject neutrality in this sense; they acknowledge that policies will affect people differently, and they think this is not a cause for concern. A scheme of basic liberty will make some lifestyles more difficult to lead than others. Consider again Bob who likes playing baseball but now that he lives in the UK, he finds that it is easier to get participants for a game of football than baseball. The 'cultural' market has a different impact on his life than it has in Wayne's life (Wayne is an avid football player). Nevertheless, Bob cannot claim unfair treatment by the government just because his lifestyle is not as easy to pursue as Wayne's. Neither Rawls's 'primary goods' nor Dworkin's 'equality of resources' support compensating people for these kinds of differences<sup>52</sup>.

Another sense in which neutrality can be understood is as a means of justification: governments cannot ground policies on the intrinsic merits of controversial conceptions of the good. This is the sense in which most liberals use the term neutrality<sup>53</sup>. Governmental policies must be independent from comprehensive doctrines<sup>54</sup>. For instance, governments can fail to be neutral by subsidising religious schools on the belief that religious lives are better than non-

---

<sup>51</sup> Will Kymlicka, 'Liberal Individualism and Liberal Neutrality', *Ethics*, 99 (1989), 883.

<sup>52</sup> They do support compensation for other inequalities, which does not concern us here. It seems that one needs to be welfare egalitarian in order to support impact neutrality. Cf. Dworkin, *Sovereign Virtue* Chapter 1.

<sup>53</sup> Cf. Dworkin, 'Liberalism'; Dworkin, *Sovereign Virtue*, 153-5, 281-2; Peter de Marneffe, 'Liberalism, Liberty, and Neutrality', *Philosophy and Public Affairs*, 19 (1990), 253-74; Nagel, *Equality and Partiality*, Chapter 14; Rawls, *Political Liberalism*, 190-1; Rawls, *A Theory of Justice*, 80-1.

<sup>54</sup> Rawls uses the term 'doctrine' to refer to 'comprehensive views of all kinds' Rawls, *Political Liberalism*, xxxvii-xxxviii n. 2. 'Comprehensiveness' refers to doctrines or conceptions which 'include [...] what is of value in human life, and ideals of personal character, as well as ideals of friendship and of familial and associational relationships, and much else that is to inform our conduct, and in the limit to our life as a whole' Ibid. 13.

religious ones. I will explain how neutrality in this sense is possible later; the point of import here is to distinguish between its impact and justificatory senses.

A common charge against neutrality is that it is logically inconsistent: it is impossible because it is self-defeating. In essence, neutrality must be neutral regarding itself, that is, liberalism must be neutral towards neutrality and non-neutrality, and this is self-contradictory. Liberalism, it is argued, cannot claim both neutrality and non-neutrality at the same time. This charge, however, misconstrues the nature of liberal neutrality. Liberalism does not need to be neutral on every issue, and certainly not towards neutrality itself. Nor it does need be neutral towards its own conception of justice and associated values. Liberals believe these are important, and that institutions must be organised in order to pursue them. These values include the idea of fair social cooperation, equality, liberty, and so on. Neutrality is not a nihilist position. Liberals are not neutral towards the interests of citizens as specified in their theory of justice. Rawls believes that citizens have two fundamental interests (to develop and exercise an effective sense of justice and to be able to form, revise and rationally to pursue a conception of the good<sup>55</sup>), and that institutions must promote those interests. To be sure, he does not wish to be neutral between arrangements that promote them and those that do not. So, liberals are not neutral towards neutrality, conceptions of justice, and the values associated with them, or the interests that it assigns to citizens<sup>56</sup>.

---

<sup>55</sup> Rawls, 'Kantian Constructivism in Moral Theory', 312.

<sup>56</sup> Rawls distinguishes between procedural neutrality and neutrality of aim. The former only appeals to 'a procedure that can be legitimated, or justified, without appealing to any moral values at all'. *Political Liberalism*, 191.

In essence, liberals believe that governments must be neutral *only* towards a subset of ideas: the set of (reasonable) conceptions of the good<sup>57</sup>. These include among others the value of religious life, sexuality, the worth of family and friends, the intrinsic merits of art, sport, and different philosophical theories. Liberal neutrality claims that governments must not base policies on an assessment of the value of these activities. In short, neutrality of justification ‘is the principle that the principles of justice that regulate basic social and political institutions must be justifiable in terms of moral and political values that any reasonable person would accept as a basis of moral claims regardless of his or her particular conception of the good’<sup>58</sup>.

### *Rawls, Liberalism and Legitimacy*

Rawls provides a good example of how liberalism can be neutral between competing conceptions of the good. He believes that a conception of justice is *political* if it satisfies three conditions<sup>59</sup>. First, it must be worked out for a specific domain, namely, for political, social and economic institutions. It should cover guidelines and principles of distribution, entitlements, burdens, benefits and conduct that apply only to those institutions, not to other spheres of life. So, a political conception regulates

---

<sup>57</sup> For the time being I will leave the discussion about reasonableness aside. Suffice to say that reasonableness as applied to a person involves a ‘disposition to promote and honour fair terms of cooperation and willingness to recognise the burdens of judgment and accept their consequences’. Rawls, *Political Liberalism*, 49 n. 1. A reasonable doctrine affirms an intelligible view of the world, contains norms about how to single out values and how to balance them in case of conflict, and belongs to an evolving tradition. Ibid. 59.

<sup>58</sup> de Marneffe, ‘Liberalism, Liberty, and Neutrality’, 253.

<sup>59</sup> For what follows see Rawls, *Political Liberalism*, 11-15.

taxation, but does not provide directives about how to choose sexual partners<sup>60</sup>. Second, a political conception is justified in a particular manner: it is presented as a 'freestanding' view. Its justification is neither derived from nor dependent on comprehensive doctrines. In contrast with other comprehensive conceptions, such as Kant's, Mill's, Aquinas's or Aristotle's, a political conception does not appeal to matters of personal virtues, non-political values, religious beliefs or metaethical theories. This does not mean that it cannot be supported by comprehensive doctrines; rather that this is not necessary<sup>61</sup>. Third, a political conception derives its content from implicit ideas present in the public political culture of democratic societies. Rawls takes the commonly shared ideas of freedom and equality to their logical conclusion<sup>62</sup>, ideas which have been developed during recent centuries and form the core of democratic societies. A political conception of justice is neutral towards comprehensive doctrines, then, in the sense that it is presented and can be justified independently of any of those doctrines. Similarly, Rawls's theory maintains neutrality between different conceptions of the good by specifying a list of primary goods<sup>63</sup>, which can be shared by citizens independently of the particular conceptions of the good they affirm. '[I]t is enough that citizens view themselves as moved by the two highest-order interests of moral personality [...], and that their particular conceptions of the good, however distinct their final ends and loyalties, require for their advancement roughly the same primary goods'<sup>64</sup>. An adequate list of primary

---

<sup>60</sup> The subject matter of a political conception constitutes what Rawls calls the 'Basic Structure'. Cf. Chapter 1.

<sup>61</sup> Cf. '[A]...political conception can be seen as part of a comprehensive doctrine but it is not a consequence of that doctrine's nonpolitical values' *ibid.* 155.

<sup>62</sup> Thomas Nagel, 'Rawls on Liberalism' in Freeman, S. (ed.) *The Cambridge Companion to Rawls* (Cambridge: Cambridge University Press, 2003), 63.

<sup>63</sup> Cf. Rawls, 'Kantian Constructivism in Moral Theory', 362.

<sup>64</sup> *Ibid.* 361.

goods, it is claimed<sup>65</sup>, treats every person as a free and equal moral agent (this conception of the person is a political conception, not a comprehensive one).

The explanation for why a political conception of justice is the appropriate fashion by which to establish fair institutions is based on aspects of democratic societies: the ‘fact of reasonable pluralism’ and the ‘principle of liberal legitimacy’. First, people affirm a variety of reasonable conceptions of the good. They disagree about religion, virtue, philosophy, the meaning of life, our place in the universe, and other comprehensive subjects. This plurality of views is neither a historical accident nor a thing to be regretted; it is the natural outcome of practical reason used freely<sup>66</sup>. These disagreements could only be eliminated through oppressive power<sup>67</sup>. If the liberal conception of justice depended on a comprehensive view—one on which people disagree—it would not generate the consensus necessary for satisfying the principle of liberal legitimacy:

Our exercise of power is fully proper only when it is exercised in accordance with a constitution the essentials of which citizens as free and equal may reasonably be expected to endorse in the light of principles and ideals acceptable to their common human reason<sup>68</sup>.

---

<sup>65</sup> I say it is ‘claimed’ because there are some difficulties concerning whether it actually treats every one as free and equal. This debate, though, is not relevant to the point about neutrality made here. Cf. Thomas Nagel, ‘Rawls on Justice’, *The Philosophical Review*, 82 (1973), 220-34.

<sup>66</sup> Rawls, *Political Liberalism*, 36.

<sup>67</sup> The explanation of how reasonable disagreement comes about is provided by the ‘burdens of judgment’. Ibid. 54-58.

<sup>68</sup> Ibid. 137

According to Rawls then, the only means of generating an overlapping consensus (a consensus for the right reasons<sup>69</sup>) is to articulate a political conception of justice, which is neutral between competing conceptions of the good<sup>70</sup>. By an overlapping consensus it is meant that people in a well-ordered society normally act in conformity with reasonably just laws and endorse a liberal conception of justice, for many different reasons each stemming from their conceptions of the good, including their comprehensive moral views<sup>71</sup>. If people could not agree with the principles ruling their society, it would fail to treat every citizen as a free and equal moral agent.

### *Perfectionism and Anti-Perfectionism*

Now contrast the view described above with a perfectionist position: perfectionism is the view that ‘there is no fundamental principled inhibition on governments acting for any valid moral reason’<sup>72</sup>. According to Raz, this is a natural view. Once a person identifies a valuable aspect of the good life, the fact that it is valuable is a reason for bringing it about. The government has a duty to make citizens comply with reasons that apply to them independently if by following when government’s directives citizens are more likely to comply with those reasons than when following their

---

<sup>69</sup> Contrast this view with a consensus generated as a ‘modus vivendi’. Imagine a regime in which Buddhists are not persecuted, not because the Christian majority think that enforcing Christianity is wrong, but because they lack the power to convert or annihilate Buddhists.

<sup>70</sup> Some people are sceptical regarding the possibility of reaching agreement on the conceptions of justice. The argument holds that only certain disagreements are to be taken into account, that is, only disagreements among reasonable conceptions of justice. This condition makes disagreements about justice less pervasive, by keeping only a family of liberal conceptions. For a full articulation of this argument see Clayton, *Justice and Legitimacy in Upbringing*, 19-24; and Rawls, *Political Liberalism*, Lecture IV.

<sup>71</sup> Freeman, *Rawls*, 366.

<sup>72</sup> Raz, ‘Facing Up’, 1230.

independent judgment<sup>73</sup>. If there is a valuable course of action, then it is the duty of government to bring it about. This is the core of the naturalness of perfectionism. As Joseph Chan argues, ‘people care about the quality of their lives and have an interest in leading a good life. If the state’s existence is to help citizens to pursue their interests, it seems natural that the state should assist citizens by promoting [them]<sup>74</sup>’. This position does not exclude any valid moral reason: all moral claims are ‘fair play’ for governments.

Perfectionism does not need to be committed to the enforcement of the good life in all cases. It acknowledges that there are strategic reasons for not using power for promoting of the good. In some cases, the value of an activity is conditional upon it being done freely. For instance, it is likely that a person’s prayers have religious value only if she believes that praying is a worthwhile activity. If this is so, then the government should not make praying compulsory. The reason is strategic: even if there was nothing morally wrong with enforced prayers, making this compulsory would be self-defeating.

Another consideration that perfectionists might have is that the enforcement of some valuable action might diminish the value of autonomy, given that the latter requires endorsement. Imagine a frustrated painter who thinks that all his pictures lack quality. Imagine also that, contrary to his opinion, we think his pictures are very good and artistically valuable. Should the government force him to keep painting for the sake of art, or should he be allowed to drop the pursuit? From a perfectionist

---

<sup>73</sup> The normal justification thesis claims that ‘the normal way to establish that a person has authority over another person involves showing that the alleged subject is likely better to comply with reasons which apply to him (other than the alleged authoritative directives) if he accepts the directives of the alleged authority as authoritatively binding and tries to follow them, rather than by trying to follow the reasons which apply to him directly’. Raz, *The Morality of Freedom*, 53.

<sup>74</sup> Joseph Chan, ‘Legitimacy, Unanimity, and Perfectionism’, *Philosophy and Public Affairs*, 29 (2000), 5-6.



perspective the answer depends on the balance between the value of art and the value of autonomy. Perhaps the value autonomy is outweighed by his masterful creations; perhaps it is not. Some might make the stronger claim that autonomy is a *condition* of the good. For instance, Will Kymlicka asserts that ‘no life goes better by being led from the outside according to values the person does not endorse. My life goes better if I am leading it from the inside according my beliefs about value’<sup>75</sup>. Imagine that listening to Sibelius is better than listening to Stravinsky, but pursuing one’s autonomous preference for Stravinsky might be better than listening to Sibelius under coercion.

Although some of these considerations present important obstacles to perfectionism, it might be responded that it still has the resources to overcome them. For instance, Hurka argues that, instead ‘of forcing people into a single best activity’, perfectionists might go the milder option ‘of forbidding a single worst activity’. Or even for the still weaker possibility of giving subsidies to good activities or heavily taxing worse ones<sup>76</sup>. Which of these options is chosen depends on strategic reasons and calculations.

With the practical issues circumvented perfectionists must then fully commit to the promotion of valuable ways of life. Citing principled reasons for not bringing about value is not an option open to them. In this respect the idea of ‘modest perfectionism’ is puzzling. For instance, Chan argues that modest perfectionism is

---

<sup>75</sup> Will Kymlicka, *Contemporary Political Philosophy: An Introduction* (Oxford: Clarendon Press, 1990), 203-4. Quoted by Thomas Hurka, ‘Indirect Perfectionism: Kymlicka on Neutrality’, *The Journal of Political Philosophy*, 3 (1995), 39-40. Notice however that it is not clear whether Kymlicka’s arguments support perfectionism. Kymlicka thinks that his view supports liberal neutrality, while Hurka argues that they can lead to an indirect perfectionism.

<sup>76</sup> Hurka, ‘Indirect Perfectionism’, 44, 47, 49 respectively.

‘local, noncoercive, mixed and multicentered’<sup>77</sup>. This version of perfectionism, however must provide an explanation of why, if one accepts the perfectionist claim at all, we should not opt for the more extreme version or some other middle-ground. It could be that modest perfectionism is appealing, but its appeal might be generated by anti-perfectionist reasons.

Liberal neutrality, on the other hand, cites principled reasons to refrain from acting on the validity of comprehensive views. As such, neutrality has a peculiar structure. The idea that supports neutrality provides an exclusionary reason to refrain from acting on valid reasons dependent on controversial conceptions of the good. Raz distinguishes between first- and second-order reasons. The former are reasons that appeal directly to certain states of affairs. The latter are ‘any reason[s] to act for a reason or to refrain from acting for a reason’<sup>78</sup>. A particular type of second-order reasons is exclusionary reasons, which demand we refrain from acting for a reason. Raz argues that in cases in which first-order reasons conflict, this should be resolved by assessing the relative strength of both reasons, and opting for what we have most reason to do. So, if I have a reason to go to the cinema tonight but also a reason to meet some friends at the pub, what I should do depends on which of those two reasons is stronger. On the other hand, when conflict arises between a first-order reasons and an exclusionary reason, ‘such conflicts are resolved not by the strength of the competing reasons, but by a general principle [...] which determines that exclusionary reasons always prevail’<sup>79</sup>. In this respect exclusionary reasons are reasons for not acting on the balance of reasons. Raz gives the example of Ann who

---

<sup>77</sup> Chan, ‘Legitimacy, Unanimity, and Perfectionism’, 16.

<sup>78</sup> Raz, *Practical Reason and Norms*, 39.

<sup>79</sup> *Ibid.* 40.

is trying to decide how to invest her money. One night, a friend phones offering her the chance of a good investment. The decision, however, must be taken immediately. Ann had a horribly tiring day at work and does not feel in the mood to make important decisions, so she declines the offer. Her state of mind and her tiredness are reasons not to make a decision, rather than a reason to decline the offer *per se*. She explains that ‘she is [refusing] the offer not because she thinks that the reasons against it override those in its favour but because she cannot trust her own judgment at the moment’<sup>80</sup>. In this respect, she has a reason not to act on the merits of other reasons. Notice though that exclusionary reasons do not imply that the excluded reason has no force; they do not nullify or cancel it. It is merely that the latter cannot outweigh the former.

Construed as an exclusionary reason, anti-perfectionism means only that the government cannot be motivated by certain considerations derived from comprehensive doctrines. Anti-perfectionism is not committed to the view that comprehensive doctrines have no merit, or are false. This should dismiss the claim that anti-perfectionists are sceptical about the good. For example, Hurka thinks that ‘Rawls’s defence of liberal neutrality combines the defeatist assumption that political philosophy cannot aim at true political principles’<sup>81</sup>. Given what has been said here, this criticism does not stand. Political principles do not aim at the good because there is a reason not to act from the validity of comprehensive doctrines of the good, not because political liberals think the good cannot be deciphered or attained<sup>82</sup>. As Raz writes, exclusionary reasons are reasons for not being motivated

---

<sup>80</sup> Ibid. 37.

<sup>81</sup> Hurka, ‘Indirect Perfectionism’, 56.

<sup>82</sup> Cf. Rawls’s comments on *modus vivendi* in *Political Liberalism*, xxxix.

in one's action by certain (valid) considerations'<sup>83</sup>: they do not deny the truth of the reasons excluded.

### *Legitimacy, Publicity, and Anti-Perfectionism*

So, what motivates the exclusionary reasons that support anti-perfectionism? In particular, they are supported by the principle of liberal legitimacy discussed above, which holds that the use of power is appropriate only if it is used in terms which citizens (as free and equal moral agents) may be reasonably expected to agree<sup>84</sup>. Elaborating further on the principle of liberal legitimacy, Rawls argues that a political conception of justice needs to satisfy what he calls the 'publicity condition'. This condition has three levels. First, it is satisfied when 'society is effectively regulated according to public principles of justice: 'citizens accept and know that others likewise accept those principles, and this knowledge in turn is publicly recognized'<sup>85</sup>. The second concerns the sort of beliefs by which principles of justice can be justified, 'that is the general beliefs about human nature and the way political and social institutions generally work, and indeed all such beliefs relevant to political justice'<sup>86</sup>. These beliefs are modelled by the original position. It is important for the sake of the argument defended in this thesis to acknowledge that beliefs about human psychology are recognised as public justifications, and, as such, can be legitimately used for modelling a political conception of justice. The third condition stipulates that the full justification of the conception of justice is publicly known, or

---

<sup>83</sup> Raz, *Practical Reason and Norms*, 185.

<sup>84</sup> Rawls, *Political Liberalism*, 137.

<sup>85</sup> *Ibid.* 66.

<sup>86</sup> *Ibid.* 66.

at least publicly available<sup>87</sup>. The publicity condition, then, restricts the sort of justifications that can guide governmental policies; it restricts the ‘pool’ of available justificatory reasons to those that every citizen can reasonably accept. Publicity secures the base for an overlapping consensus that guarantees the legitimate exercise of political power. These are the goods that motivate the exclusionary reasons provided by liberal neutrality against perfectionism<sup>88</sup>.

The idea of autonomy is constitutive of both the legitimacy and the publicity conditions. It seems that only by endorsing freely the principles that rule coercive institutions can autonomy be respected. As Clayton argues, ‘[f]ull autonomy is a condition achieved only if citizens freely identify with the constraints that they face, in the sense that they understand the content and justificatory bases of the constraints, and freely accept them’<sup>89</sup>. It is important, for our purposes, that full autonomy is a political concept; it is not derived from a comprehensive doctrine. It is specified by the political conception of the person as free and equal and as having the two moral powers: an effective sense of justice and the capacity to form, revise and pursue a conception of the good. According to Rawls, the only circumstances in which full autonomy can be realised is when citizens ‘act from principles of justice that specify the fair terms of cooperation they would give themselves when fairly represented as free and equal persons’<sup>90</sup>. The publicity condition can also be defended from an anti-paternalistic principle, according to which the government

---

<sup>87</sup> Ibid. 67.

<sup>88</sup> Cf. ‘Liberalism asks that citizens accept a certain restraint on the power of the state to enforce some of their most deeply held convictions against others who do not accept them, and holds that the legitimate exercise of political power must be justified on more restricted grounds which belong in some sense to a common or public domain’. Nagel, *Equality and Partiality*, 158.

<sup>89</sup> Clayton, *Justice and Legitimacy in Upbringing*, 15.

<sup>90</sup> Rawls, *Political Liberalism*, 85.

fails to respect citizens' autonomy when the latter are ruled by principles they cannot be reasonably expected to accept, even if the rules are grounded on true premises<sup>91</sup>.

Contrast this condition of publicity with that endorsed by Thomas Christiano, who argues that legitimacy requires weak publicity. According to him, publicity demands that 'the principles of justice be ones that that people can in principle see to be in effect'<sup>92</sup>. Christiano defends this weak notion of publicity on the grounds that the 'requirements of consensus would impose an impossible burden on principles of justice. Theories cannot get off the ground if they require agreement on principles themselves as a condition of justice'<sup>93</sup>. Given the impossibility of full publicity, Christiano suggests, we must adopt a weaker version by 'publicly embody[ing] justice in a way that is compatible with a range of disagreements about what justice requires'<sup>94</sup>.

One way of answering Christiano's challenge is by highlighting that not all possible disagreements about justice are fatal for an overlapping consensus. Rawls insists that the consensus for which we should aim is restricted only to *reasonable* citizens. Reasonableness is associated 'with the willingness to propose and honor fair terms of cooperation' and 'with the willingness to recognize the burdens of judgments and to accept their consequences'<sup>95</sup>. If a person who fails to satisfy the threshold of reasonableness disagrees with the principles of justice, the consequences of such disagreement are not relevant for legitimacy. Indeed Rawls argues that political liberalism aims at a consensus among a family of liberal conceptions

---

<sup>91</sup> Clayton, *Justice and Legitimacy in Upbringing*, 15.

<sup>92</sup> Thomas Christiano, 'The Authority of Democracy', *Journal of Political Philosophy*, 12 (2004), 270.

<sup>93</sup> *Ibid.* 274.

<sup>94</sup> *Ibid.* 274.

<sup>95</sup> Rawls, *Political Liberalism*, 49 n. 1.

(among which he thinks ‘justice as fairness’ is the most reasonable). He argues that liberal conceptions have three conditions. First they specify certain ‘rights, liberties, and opportunities’. Second, they give special priority to these freedoms. And third, they specify ‘measures assuring all citizens, whatever their social position, adequate all-purpose means to make intelligent and effective use of their liberties and opportunities’<sup>96</sup>. Weak publicity seems to have two problems. On the one hand the argument that *reasonable* agreement is impossible is not tenable. On the other, weak publicity is not appealing insofar as it does not sufficiently respect the political conception of autonomy<sup>97</sup>. It seems clear then, that full publicity is a necessary condition for the satisfaction of citizens’ first moral power. ‘Only if the full explanation and justification [of the political conception of justice] is publicly available can citizens come to understand its principles in accordance with the idea of society as a fair system of social cooperation’<sup>98</sup>.

### *Autonomy and Anti-Perfectionism*

I have developed a concept of autonomy far more demanding than the political account defended by Rawls<sup>99</sup>. Next I want to consider briefly whether it is possible to maintain the appeal of a more comprehensive version of autonomy within an anti-

---

<sup>96</sup> Ibid. xlvi. See also Clayton, *Justice and Legitimacy in Upbringing*, 19-24; Burton Dreben, ‘On Rawls and Political Liberalism’, in Freeman, S. (ed.) *The Cambridge Companion to Rawls* (Cambridge: Cambridge University Press, 2003), 328-9.

<sup>97</sup> Christiano argues that ‘this is where [...] democracy comes in’. ‘The Authority of Democracy’, 274.

<sup>98</sup> Rawls, *Political Liberalism*, 78.

<sup>99</sup> Cf. Chapter 2.

perfectionist framework<sup>100</sup>. Rawls distinguishes between fully and partially comprehensive doctrines; the former ‘covers all recognized values and virtues within one rather articulated system’, the latter ‘comprises a number, but by no means all nonpolitical values and virtues and is rather loosely articulated’<sup>101</sup>. Is it possible to argue for a non-perfectionist liberalism which gives weight to autonomy?

Consider the following distinction. We can have an *inclusive* idea of neutrality, according to which the government must be neutral towards different conceptions of the good, but also towards different conceptions of the relation between a person and her conception of the good<sup>102</sup>. On the other hand, we can also have a *restricted* view of neutrality, which applies only to conceptions of the good, and not to the way in which a person forms, revises and pursues them. While the former demands neutrality towards the idea of autonomy, the latter does not; it allows liberals to attach moral significance to how a person forms her attachments and goals. Restricted neutrality appeals to the value of autonomy in a manner consistent with the second moral powers of citizens. It acknowledges that forming, developing and pursuing a conception of the good requires the satisfaction of certain conditions.

This is a controversial claim, for someone might reject the assertion that autonomy forms part of her well-being. Despite accepting that autonomy is controversial, a partially comprehensive liberalism insists that autonomy is valuable. Nevertheless, it rejects perfectionism by citing *principled reasons* for not promoting

---

<sup>100</sup> The argument that follows is highly indebted to Clayton, *Justice and Legitimacy in Upbringing*, 26-7.

<sup>101</sup> Rawls, *Political Liberalism*, 13.

<sup>102</sup> Matthew Clayton, ‘White on Autonomy, Neutrality and Well-Being’, *Journal of Philosophy of Education*, 27 (1993), 104-6.



certain (valuable) aspects of a person's life<sup>103</sup>: we can invoke some part of the truth about non-political values for justification, but not the whole truth. We can appeal to part of the controversial value of autonomy, while still maintaining principled limits regarding what the government can appeal to.

Contrast this view with Rawls's position. Rawls distinguishes between two forms of the identity of persons. On the one hand, qua citizens we have an institutional, public identity, which is not tied to any particular conception of the good. In relation to this identity we have, as citizens, an interest in being able to revise and detach ourselves from the particular conception we endorse. On the other hand, we have a noninstitutional identity, which specifies our deeper attachments and goals<sup>104</sup>. In some cases, these two identities can conflict. Rawls gives the example of the person who changes his faith. As a follower of a different religion he might think he has become a different person, but for public matters his institutional identity has not changed: he is still bound by the duties and obligations he was prior to changing his faith. In this sense, Rawls thinks that, by defending a purely political conception of the person, a political conception of justice will generate the necessary agreement for an overlapping consensus.

This is relevant to the distinction between inclusive and restricted neutrality in the following way. Imagine Paul, a person who thinks that he is better off by not subjecting his faith to critical analysis. He thinks that there are many temptations in the world, and that it is likely that he will be misguided into believing falsehoods. As a believer, he thinks that it is in his interest not to be able to revise his conception of the good, or in other words, he rejects that he has an interest in developing his second

---

<sup>103</sup> As we have seen, this option is not available to perfectionists.

<sup>104</sup> Rawls, *Political Liberalism*, 30.

moral power. Why should he affirm, qua citizen, that he has an interest which, qua person, he rejects? As Clayton argues, 'it is unclear why those who reject the comprehensive claim that they have an interest in being able rationally to reflect on, and revise, their conception of the good should accept that they have such an interest as citizens'<sup>105</sup>. Think about Betty, a fellow believer of Paul's who, after subjecting her faith to critical scrutiny, remains strongly convinced of it. From an inclusive sense of neutrality we must remain indifferent to Paul's and Betty's cases. If, in contrast, we affirm a restricted sense of neutrality, we can perceive differences between them. Under this interpretation, we affirm that Betty takes the right course of action, by establishing a direct link between her moral autonomy and her political autonomy. To be sure, this view is controversial (in virtue of being partly comprehensive), but it is, nevertheless, more plausible than the alternative. A restricted sense of neutrality unites our intuitions regarding the value of autonomy with an anti-perfectionist political morality.

To conclude I will briefly summarise what has been said in this chapter. Most liberal discussions of free expression regard restrictions on hate speech as illegitimate when they are content based. The presumption against content regulations emanates from a popular account of political neutrality. A common strategy to neutralise the harmful effect of racist speech is to have 'more, better speech'. Against Cohen's 'fact of reasonable persuasion', I argued that, despite people being technically able to change their minds, they tend not to. People usually do not assess new information neutrally, they form impressions on irrelevant

---

<sup>105</sup> Clayton, *Justice and Legitimacy in Upbringing*, 25.

information, get anchored, and are subject to the distortions emanating from information framing. I also suggested that the manner in which we form beliefs is biased towards acceptance and that assessing and rejecting propositions we understand consumes considerable cognitive resources, which, in stimuli-rich environments, might not be always available. These claims support the argument I made in previous chapters, which asserted that the best way to defend ourselves from mental contamination is by controlling the environment in ways favourable to our deliberative capacities. Content-based restrictions on speech are justified, not by an assessment of the intrinsic merits of racist speech, but by the effect this has on autonomy.

I also reject the claim that the position delineated above is perfectionist. I discuss some misunderstandings about the meaning of liberal neutrality and interpret neutrality as an exclusionary reason, justified by the goods of publicity and legitimacy. A condition of full publicity is defended on an account of citizen's autonomy. Finally, I defended a restricted sense of neutrality, which is consistent with an anti-perfectionist autonomy-based liberalism, and my argument in favour of content-based regulations of speech.

## CHAPTER 6

### FREE ASSOCIATION, PUBLIC ECOLOGY, AND FREE SPEECH

Most liberals agree that freedom of association is a vital component of life in a free society and therefore must be protected from governmental intervention. However, there is still much disagreement concerning the limits to which government can legitimately impose upon the activities and the membership criteria such groups are allowed to employ. In this chapter, I draw from the arguments developed in those previous to argue that there are legitimate reasons for governments restricting what private associations are allowed to express and the admissions criteria they are allowed to employ. In doing so, I shall focus attention specifically on racist speech and racist associations.

There are striking similarities between the defences of freedom of association and freedom of speech. First, both are usually defended on grounds of autonomy. Being free to associate in one's own terms is, for some people, a condition of personal autonomy. And, in similar fashion to the defence of speech, this idea can be used to defend free association as a constraint on governmental actions. On this view, the state fails to respect a person's autonomy if it interferes with her joining an association. Autonomy can, additionally, figure in a consequentialist argument,

which claims that a government's promotion of the value of autonomy requires an environment in which people are free to associate on their own terms<sup>1</sup>.

Defences of freedom of association can employ a minimalist or a maximalist approach<sup>2</sup>. Regarding the former, it can be argued that the impacts of associations are not harmful to third parties, and, consequently, associational freedoms are an instantiation of a more general principle of minimal liberty. On the other hand, it can be argued that the harmful effects they have on third parties are trumped by the benefits of associations for members or for society in general.<sup>3</sup>

Free association resembles freedom of expression in the sense that there is a strong presumption against regulating the 'content' of associations. Although it is permissible to restrict associations' activities on the basis of 'manner, time and place'<sup>4</sup>, governments are usually barred from restricting them in virtue of an assessment of the merits of their membership policies and/or pursuits. For instance, preventing an association of environmentalists from gathering close to a nuclear plant might be acceptable when the restriction is justified on 'contextual' grounds, but not when it is made only on an assessment of environmentalism. In general, interference can be justified only by showing that an association presents a clear and present danger to society; not merely that it has some harmful consequences on others.

Lastly, it is sometimes suggested that the costs of associations can be neutralised by forming other groups that will provide opportunities to the excluded

---

<sup>1</sup> Free associations can also be defended on grounds of democracy by highlighting the contribution that associations make to democratic pluralism. Cf. Robert Dahl, *Dilemmas of Pluralist Democracy: Autonomy vs. Control* (New Haven: Yale University Press, 1982).

<sup>2</sup> I discuss the difference between minimalist and maximalist defences of speech in Chapter 1. Cf. also Cohen, 'Freedom of Expression'.

<sup>3</sup> The arguments I will discuss in this chapter mainly address this approach.

<sup>4</sup> These restrictions usually are addressed by liberty of assembly.

people. This strategy is similar to the ‘more, better speech’ remedy discussed in Chapter 5. The idea is that the legitimate way to neutralise the harmful effects of associations (or speech) involves setting up new opportunities, either for association or for speech, which will redress the setbacks caused by others’ associational or expressive activities. This alternative is preferred to interfering directly with citizens’ activities.

The claim I defend in this chapter holds that an association’s activities are to be constrained by the condition of reasonableness, and that their exclusionary practices must be justifiable by public reasons. These conditions are better suited to protecting the goods of associational freedom. The argument defended might face the following objection: although at the core of liberal regimes is the idea that governments have a duty to treat everyone with equal respect and concern, private associations and individuals (in their private lives) are not bound by this requirement. So, a liberal regime rules out racism or sexism only at state level; in the private sphere citizens are allowed to associate on racist grounds, or express racist views if they so wish.

I will proceed as follows. I begin by describing the fundamental goods it is claimed voluntary private associations protect and/or promote, and the constraints under which they are usually assumed to operate. I will illustrate how associations can relate to those goods through three examples of racist associations. In the second section, I will assess a prominent theory of associations: Nancy Rosenblum’s ‘Moral Uses of Pluralism’<sup>5</sup>. Finally, I will develop an argument for legitimate restrictions on associations constrained by an account of reasonableness and public reason<sup>6</sup>.

---

<sup>5</sup> Nancy Rosenblum, ‘Compelled Association: Public Standing, Self-respect and the Dynamics of

## I

## Associational Freedom

Within liberal societies people associate for many reasons and with many goals. In order to understand why governments should not interfere with and in many cases protect these associations it is useful to identify the goods that associations are supposed to provide. While doing so, it is important to bear in mind that freedom of association necessarily (analytically) involves a right of discrimination. My freedom to associate with a person X for an activity A is meaningless unless I also have the possibility of not being joined by person Y. Although this point is somewhat trivial, it is important to notice that it says nothing regarding the reasons by which people can legitimately deny membership to others.

*Associational Goods*

**Authenticity.** It is usually argued that associational freedom is a fundamental component of living a life on one's own terms. As George Kateb writes: 'To be free, to live as one likes, includes associating on one's own terms, which means engaging in relationships of all sorts, finding or trying to find pleasure in them, and also finding in them opportunities for many kinds of experience'<sup>7</sup>. In order to lead a life that I can truly call 'mine', I must be able to engage in relationships with the people I choose, and I must be able to do so in the terms with which we voluntarily agree.

---

Exclusion', in Gutmann, A. (ed.) *Freedom of Association* (Princeton: Princeton University Press, 1998), 75-108; Rosenblum, *Membership and Morals*.

<sup>6</sup> Stuart White, 'Freedom of Association and the Right to Exclude', *The Journal of Political Philosophy*, 5 (1997), 373-91.

<sup>7</sup> Kateb, 'The Value of Association', 37-8.

The possibility of joining people with whom I share interests, values and commitments is a fundamental part of leading a life with which I can identify myself. Similarly, many of a person's important and dearest projects are collective enterprises. In order to be successful these projects require several persons acting together. If carried out in solitude these projects will fail. It is important to note also that the capacity of associations to protect authenticity would (arguably) be self-defeating if their terms were imposed by government. Essentially, this is because people do not merely want to associate with others; they also wish to do so in terms they find satisfactory and mutually acceptable.

**Intimacy.** This good refers to the capacity for companionship, friendship and camaraderie. Although, in some sense, the goods of intimacy resemble authenticity, they seem to be 'less serious'<sup>8</sup>. This is not to argue that intimacy is not an important good; only that it is less 'articulated' than those covered by authenticity. People usually establish meaningful 'spontaneous' relationships with others. These relationships arise in daily life and usually lack any specific aim (apart from companionship itself). Groups of friends who meet in the pub or eat lunch together are perfect examples. These relationships, however, can flourish and become much more important. When a person chooses a partner or to have a close relation with her family she also enjoys the good of intimacy. Intimacy, however, is fragile in the sense that it is mainly based on trust and 'good-faith'; governmental intervention can very easily spoil the meaning, worth and enjoyment of these sorts of associations.

---

<sup>8</sup> Cf. Sam Fleischacker. 'Insignificant Communities' in Gutmann, A. (ed.) *Freedom of Association* (Princeton: Princeton University Press, 1998), 273-313.



**Personal Development.** Another important set of goods delivered by associations concerns the development of both moral and psychological capacities of members. In virtue of being voluntary, it is usually argued, associations provide an environment in which members acquire moral dispositions for cooperation, respect for public rules, and self-esteem<sup>9</sup>. The dynamics of associational life generate virtues such as respect, reciprocity, trust, solidarity and mutual support. For instance, Rawls argues that the ‘morality of association’ requires a complex set of abilities which include the capacity to recognise that other individuals might have different points of view, different perspectives, different ends and wants. Moreover, it is not only recognition that associations can foster, they can also enlighten our understanding of others’ reasons as reasons that deserve respect and endorsement. Finally, associations encourage control and modification of our behaviour in respect of other people’s points of view<sup>10</sup>. In short, the idea is that, by fostering cooperation, associations trigger the moral development of their members<sup>11</sup>.

**Expressive Goods.** Sometimes people associate in order to put forward a ‘voice’. In this sense, a central aim of associations is to communicate a point of view (or a subject matter) to the public agenda. The relation between private associations and free speech is usually assumed to grant a right of association<sup>12</sup>. Given their

---

<sup>9</sup> Rosenblum, ‘Compelled Association’.

<sup>10</sup> Rawls, *A Theory of Justice*, 409-410. Rawls goes on to say that ‘we may suppose that there is a morality of association in which the members of society view one another as equals, as friends and associates, joined together in a system of cooperation known to be for the advantage of all and governed by a common conception of justice’. Ibid. 413.

<sup>11</sup> This holds for both, children and adults.

<sup>12</sup> For instance in the U.S. Constitution, there is not a right of association as such, but associations have usually been protected under the First Amendment doctrine (free speech and rights to assembly) and under the freedom of religion. Cf. Kent Greenawalt, ‘Religious Association’ and Amy Gutmann,

expressive goals, groups are protected from governmental intervention, and subject to the same limits, for the same reasons that individual expression is protected. For instance, Justice Brennan links association and free speech when he writes: ‘According protection to collective effort on behalf of shared goals is especially important in preserving political and cultural diversity and in shielding *dissident expression* from suppression by the majority’<sup>13</sup>.

There seem to be two tiers to these expressive goods. On the one hand, associating with others is an effective way of conveying a message to the public arena. Because it is much easier for this to get across and to be heard if it is proffered by many people (rather than just one), freedom of association increases the potential of the communication<sup>14</sup>. This aspect of expressive associations is usually referred to as its ‘voice’. On the other hand, associations also have internal expression capacities. Associations themselves are a forum for expressing and discussing ideas among members. By joining people with similar interests a person can discuss and express her views with greater ease than she could in the public arena. The connection between this aspect of expression and the good of ‘authenticity’ is obvious: part of what living a life in my own terms presumably requires is to be listened to and heard by people with interests similar to my own<sup>15</sup>. Because what is

---

‘Freedom of Association: An Introductory Essay’ both in Gutmann, A. (ed.) *Freedom of Association* (Princeton: Princeton University Press, 1998).

<sup>13</sup> Quoted by Greenawalt, ‘Religious Association’ 114. Emphasis added.

<sup>14</sup> Other things being equal: it is probably true that a rich person has a greater capacity at reaching audiences by himself than an association of poor people.

<sup>15</sup> Gutmann writes: ‘When the primary purpose of an association is expression of a point of view [...] then its freedom to select its members consistently with its expressive purposes is essential to its members’ exercise of free speech through the association [...] Any meaningful right to free speech must protect associations whose primary purpose is expressive from political interference in their membership policies insofar as that interference is directly related to its expressive purposes’. Gutmann, ‘Freedom of Association’, 11-12.

said is kept indoors I believe that this dimension of expression is much better covered under authenticity and intimacy<sup>16</sup>.

**Effective Pursuit of Goals.** The last good I wish to discuss is a fundamentally instrumental role held by associations. Many associations are joined as a means of achieving other goals. By coordinating actions, lobbying governments, providing contacts or forming networks, associations tend to be more effective at achieving goals than individuals. The degree of member involvement varies greatly in associations that provide this sort of good<sup>17</sup>. Although many involve regular face-to-face contact between members, others do not require much cooperation at all: sometimes the only requirement is a joining fee, and the involvement of members is minimal (think, for instance of associations such as Oxfam, the Wild Life Trust, or English Heritage or the friends of CBSO).

### *Conditions for Associations*

Although associational life is a fundamental component of liberal democracies, there are limits placed upon associations' activities and goals. In the literature concerning associational freedom there seems to be a consensus on four conditions that must be satisfied.

---

<sup>16</sup> Arguments of free speech are especially relevant when the expressive activities are targeted to a public audience.

<sup>17</sup> Gutmann distinguishes between primary, secondary and tertiary associations based on the degree of members' involvement. Cf. Gutmann, 'Freedom of Association', 10.

**Basic Rights.** Associations cannot engage in activities that violate the *basic rights* of citizens (there cannot be a British Association of Artful Back-Stabbers<sup>18</sup>). Nor can associations have violent, illegal goals (the Association of Tax-Evaders cannot exist). For instance, there are some important controversies regarding the extent to which associations can pursue illegal goals, or employ violent means to achieve their aims<sup>19</sup>. It is not clear what limits should acts of civil disobedience abide. For instance, think about parades of Animal Right's Activists, or the Hunting Association<sup>20</sup>.

Similarly, associations are usually barred from violating the *basic rights* of their members. Ritual torture, for instance is not a protected practice, even if inflicted on voluntarily participants. This point is controversial: some people might argue that insofar as a person gives consent, as the maxim *volenti non fit injuria*<sup>21</sup> states, the (harmful) practices with which she engages are self-regarding, and interferences are illegitimate examples of paternalism<sup>22</sup>. In contrast, it might be argued that the government has a duty to defend a basic set of rights even against the will of the person involved. Fortunately this dispute has no direct implications for the argument defended here, so I will set it aside.

**Right of Exit.** One of the essential conditions that liberal governments impose upon associations is that they must respect their members' right of exit. Associations cannot force members into life membership; the exit door must be

---

<sup>18</sup> White, 'Freedom of Association and the Right to Exclude', 379.

<sup>19</sup> A thorough discussion of these examples is far beyond the reasonable limits of this chapter.

<sup>20</sup> For a theory of civil disobedience see Will Smith, 'Democracy, Deliberation, Disobedience'. *Res Publica*, 10 (2004), 353-77. For a discussion on, and qualified defence of paramilitary associations see Rosenblum, *Membership and Morals*. Chapters 7-8.

<sup>21</sup> Thanks to Andrew Walton for bringing this maxim to my memory.

<sup>22</sup> Cf. The Spanner Case. [www.spannertrust.org](http://www.spannertrust.org) [accessed 28<sup>th</sup> August 2007]/

always open. This condition can be understood in two senses. On the one hand government can forbid voluntary ‘slavery’. Associations cannot force their members to stay for life: there must always exist the legal possibility of leaving the association. For instance liberal governments ban the practice of associations taking away members’ passports and other official documentation. However this condition is too weak: associations can still make exit unreasonably hard by denying members the actual conditions for leaving. In this respect, it seems that associations need to guarantee that its members are able to leave. Although it is not obvious to what extent associations have a positive duty to ensure exit, they are barred from taking certain actions that preclude exit. For example, consider the *Susan and Toni Alamo Foundation v Secretary of Labor* case<sup>23</sup>. The Alamo Foundation is a non-profit religious organisation which

derived its income by operating service stations, retail clothing and grocery outlets [...] staffed by Foundation “associates” [...] In the 1985 case involving the Alamo Foundation, the secretary of labor filed action under the Fair Labor Standards Act, charging the group with violating minimum wage and overtime provisions<sup>24</sup>.

Although private associations may be exempt from labour and economic regulations, the court was troubled by the fact that members-employees ‘were entirely dependent on the Foundation for food, shelter, and clothing’<sup>25</sup>. Rosenblum agrees that the Court was right in forcing the association to satisfy the Fair Labor Standards Act’s requirement because it was precluding members’ possibilities of exit.

---

<sup>23</sup> *Tony & Susan Alamo Foundation v. Secretary of Labor*, 471 U.S. 290 (1985). See Rosenblum, *Membership and Morals*, 99-100.

<sup>24</sup> *Ibid.*

<sup>25</sup> *Ibid.* 100.

**Voluntary Entrance.** Private associations cannot force entrance. This is the one of the defining characteristics of private ‘voluntary’ associations. There are a few controversial cases in which associations can force individuals to become members; for example, to become part of a trade union can be a condition for employment<sup>26</sup>. There are also other ‘associations’ to which we seem to belong without having voluntarily agreed to such as religion, culture, family, social class, but it is dubious whether they qualify as private associations in the sense required<sup>27</sup>. On the whole, however, this is typically an enforced condition.

**No Enforcement of Criminal Law.** The last duty of private associations is that they cannot enforce criminal law<sup>28</sup>. Governments must remain neutral towards the particular conceptions of the good which different private associations support. Enforcing the view of one association as criminal law would be an unfair treatment to individuals who do not belong to it and/or share its particular conception of the good. Imagine, for instance that the Catholic Church were able to criminalise homosexuality because it conflicts with traditional dogma. Homosexuals (and in fact, people who merely believe the opposite) would be treated unfairly.

Similarly, associations do not have equivalent powers to the government in the capacity of enforcing criminal law. A legitimate state has a range of sanctions, ranging from fines to imprisonment, available in order to extort citizens’ compliance

---

<sup>26</sup> I will ignore these examples for the purpose of this chapter; it seems that forcing entrance to trade unions might be acceptable when free riding (by non-unionised workers) might be an important threat to the bargaining power of the union. Cf. Stuart White, ‘Trade Unionism in a Liberal State’ in Gutmann, A. (ed.) *Freedom of Association* (Princeton: Princeton University Press, 1998), 330-56.

<sup>27</sup> Cf. Michael Walzer, ‘On Involuntary Association’, in Gutmann, A. (ed.) *Freedom of Association* (Princeton: Princeton University Press, 1998), 64-74.

<sup>28</sup> Cf. Rawls, ‘The Idea of Public Reason Revisited’, 158-9.

with the law. In contrast, a private association's right to enforce its rules is more limited: they may use some means to ensure that members comply with rules, but the right to enforce them is usually conditional on individuals remaining members. Disputes within associations are settled by members, and government tends to intervene only in cases in which an association violates the rights of citizens. Think for instance of the controversy in Christianity regarding ordaining homosexuals. A government will refuse to intervene, unless the rights of some citizens are under threat. In many instances important internal disputes lead to schisms in the association rather than an involvement of government to enforce one 'interpretation' of the 'true aims' of the association<sup>29</sup>.

### *Examples of Associations*

The following table illustrates how three racist associations could justify their activities and admission policies based on the associational goods described above.

	Racist political party	All White Church	Racist private club
Authenticity	X	X	X
Intimacy	—	X	X
Personal Development	X	X	X
Expression	X	X	?
Effective means to other goals	X	?	X

**The Racist (White) Political Party.** This association might defend its right to exist, its practices and its membership criteria by virtue of four of the associational goods discussed above. First, they can claim that excluding non-white people is a

<sup>29</sup> Cf. Rosenblum, *Membership and Morals*, Chapter 3.

necessary condition of being a racist party: associations sometimes demand would-be members pass a 'screening' process. If they accept people who do not satisfy a basic members' profile the 'viability' of the association might be jeopardised<sup>30</sup>. In this example, if a racist party accepts non-white people it would cease to be a racist white party. Second, they can argue that the association's goal is to promote and advance their members' conception of the good, such that living their lives according to their own principles is best achieved by being part of such an association. Third, they can claim that their personal moral development is triggered by the cooperation and solidarity of belonging to an association whose goals and principles they share. Fourth, they can assert that the party protects important expressive goods. On the one hand, it propagates their shared conception of the good in the public sphere; on the other, it serves as an exploration of their views and beliefs by, for instance, organising conferences, holding debates, inviting speakers and so on. Finally, the members of this party can also argue that it is an important and effective medium of achieving their political goals and for establishing their views in the electoral agenda<sup>31</sup>.

**The all-White Racist Church/Cult.** What reasons (if any) can a church provide in order to defend its right to exclude people on the basis of race, and to legitimise its practices? Similar to the political party, the church can appeal to authenticity goods. They can argue that their religious beliefs impose a duty on them

---

<sup>30</sup> For a critical discussion of the 'viability' test (regarding special exemptions in education to the Amish in the Pennsylvania) see Rosenblum, *Membership and Morals*, 94-8.

<sup>31</sup> It is not clear whether a political party can appeal to intimacy. Usually they are large associations with annual meetings and not too intense interaction among ordinary members.



to act according to their racist principles<sup>32</sup>. They can assert that there are intimate goods that their association protects (usually members of churches have lots of face-to-face interaction: they organise socials, masses, and charity programs. They also usually befriend each other and offer support, friendship and company). Although the intensity of interaction varies from church to church (ranging from ‘totalistic’ religions ‘which immerse members in the organisation and take up every moment of their lives’<sup>33</sup> to a more relaxed ones), religious life almost always comprises of levels of intimacy and fellowship. Third, members of this church can appeal to their personal development: churches promote cooperation and solidarity among coreligionists. Another important ground on which churches can claim protection is that of the expressive goods that churches are said to protect. As Stuart White writes: ‘where an association’s primary purposes concern the exploration and/or propagation of distinctive religious and/or philosophical beliefs, the right of association members to exclude specifically to protect association’s distinctive beliefs has an especially strong presumption of legitimacy’<sup>34</sup>. They are the keepers of the religious doctrine they represent, and an important function is that of disseminating this across society.<sup>35</sup>

**The all-White Racist Private Club.** This is the paradigmatic example of an association promoting intimacy. Just as I have a right to invite to my birthday party

---

<sup>32</sup> Some people argue that religious associations and religious practices deserve special protection (compared to non-religious ones) because religion imposes special duties on people. Cf. Greenawalt, ‘Religious Association’; Paul Bou-Habib, ‘A Theory of Religious Accommodation’, *Journal of Applied Philosophy*, 23 (2006), 109-26.

<sup>33</sup> Rosenblum, *Membership and Morals*, 98.

<sup>34</sup> White, ‘Freedom of Association and the Right to Exclude’, 385.

<sup>35</sup> I do not think religious associations are merely instrumental to other goals, unless one considers them to be means for salvation. In general, religious activities are described as ends in themselves.

whoever I want (and for my own reasons), so private clubs appeal to a right of intimacy that makes it ‘permissible for people to practice *whatever pattern of exclusion they like*’<sup>36</sup>. In *Roberts v Jaycees*<sup>37</sup>, Justice Brennan holds that intimate associations are highly selective; they require seclusion from others in critical aspects of the relationship’ and ‘congeniality is essential to their purposes’<sup>38</sup>. Stuart White nuances the idea of intimacy by defining two requirements that intimate associations must satisfy. First, he argues that ‘there must be a strong and mutual familiarity, ordinarily grounded in regular, intensive, “face-to-face” interaction, between more or less all the organisation members’. Second, ‘the pursuit and enjoyment of intimacy-related goods, such as friendship and love, must be the organisation’s primary associative purpose’<sup>39</sup>. Because goods of intimacy are related to authenticity the private club can also appeal to the latter in order to exclude others. In a similar fashion, personal development can also be fostered by the solidarity and cooperation of fellowship and camaraderie.

Whether the private club can appeal to expressive goods is more controversial. It is obvious that there are likely to be expressive activities within the

---

<sup>36</sup> White, ‘Freedom of Association and the Right to Exclude’, 386.

<sup>37</sup> *Roberts v Jaycees*, 468 U.S. 409, (1984). ‘Appellee United States Jaycees is a nonprofit national membership corporation whose objective, as stated in its bylaws, is to pursue such educational and charitable purposes as will promote and foster the growth and development of young men’s civic organizations... Two local chapters in Minnesota have been violating the bylaws for several years by admitting women as regular members, and, as a result, have had a number of sanctions imposed on them by appellee, including denying their members eligibility for state or national office. When these chapters were notified by appellee that revocation of their charters was to be considered, members of both chapters filed discrimination charges with the Minnesota Department of Human Rights, alleging that the exclusion of women from full membership violated the Minnesota Human Rights Act’.

<sup>38</sup> Gutmann, ‘Freedom of Association’, 9; Greenawalt, ‘Religious Association’, 113. Justice Brennan agreed with the court in their ruling that the Minnesota Jaycees accepting women as full members was not a violation of Jaycees freedom of association. Justice O’Connor ruled out the Jaycees discriminatory policy by defining them as ‘public accommodation’, that is, she denied that the Jaycees protected intimate goods. See, *Roberts v Jaycees*, see also Rosenblum, ‘Compelled Association’.

<sup>39</sup> White, ‘Freedom of Association and the Right to Exclude’, 390.

club, they probably make racist jokes and informally discuss racist science. However, they do not have expressive purposes targeted at the public agenda: they are not a 'voice' in the public arena. In this case there might be a conflict between intimacy and expression: insofar as they define themselves primarily as an intimate association their expressive claims are weakened, if they define the organisation as a 'voice' their intimate character becomes less strong. Similarly, their intimate character makes more difficult for them to become a means to another goal.

## II

### A Theory of Free Association

#### *The 'Moral Uses of Pluralism'*

In this section I will examine Nancy Rosenblum's theory of freedom of association, or as she calls it the 'moral uses of pluralism'<sup>40</sup>. In her book, she develops an emphatic defence of private associations based on the effects that 'associational life has on the moral dispositions of members personally and individually'<sup>41</sup>. Associational life, she argues, is of fundamental importance because it promotes and encourages cooperation, autonomy, and self-respect. Albeit different, these goods are related to the associational ones identified in the first part of this chapter. The capacity for cooperation is an important component of personal development and, presumably contributes towards living an authentic life. Autonomy, understood as

---

<sup>40</sup> The theory is developed at length in Rosenblum, *Membership and Morals*. A shorter and sketchier statement is presented in her 'Compelled Association'.

<sup>41</sup> Rosenblum, *Membership and Morals*, 3-4.

self-rule, overlaps, in several respects, with authenticity. Similarly, it seems that in many cases the development of autonomy requires a person to be able to form intimate bonds and relationships. The capacity for self-respect must figure in the value of authenticity, intimacy and personal development in the sense that in many instances, the value of intimate relations are conditional to fostering (or at least insofar as they do not threaten) self-respect.

Based on Rawls's 'morality of association'<sup>42</sup>, Rosenblum argues that moral development arises when

members come to appreciate that the group's system of cooperation requires a variety of actions and points of view, and they learn to take on the perspectives of others. Ties of friendly feeling and trust are generated as they see that others intend to do their share. As individuals become attached to these arrangements and develop mutual confidence, they are motivated to comply with the obligations of membership and to live up to the ideals of their station. Qua members, they learn to habitually overcome the vices that impede cooperation<sup>43</sup>.

The link between moral development and cooperation is further expanded by the effects of association participation on individuals' self-respect. According to the 'moral uses of pluralism', self-respect is 'inseparable from actual participation in an association's ends, an affirmation of the worth of an individual to a particular group will depend on internal assessment of his or her usefulness'<sup>44</sup>. These three goods, she insists, make associational life a cornerstone of a liberal society, and provide stringent reasons for protecting associations' activities.

---

<sup>42</sup> Rawls, *A Theory of Justice*, § 71, 409-13.

<sup>43</sup> Rosenblum, *Membership and Morals*, 50-1.

<sup>44</sup> *Ibid.* 62.

According to the ‘moral uses of pluralism’, liberal theory is typically overly optimistic about the effects of freedom of association on liberal democracy: there is a leading proposition in political theory and public policy, Rosenblum argues, which holds that ‘the relation between civil society (that is, our associational life) and liberal democracy *is, or must be,* reciprocally supportive; that liberal democratic character and commitment must find their origin and vital complement in an array of independent groups’<sup>45</sup>. This idea of liberal congruence is rejected by the ‘moral uses of pluralism’ for two reasons.

First, there is no reason to suppose that associations will enforce the values of liberal democracy. In a plural society, Rosenblum argues, associations praise many different conceptions of the good that may be incompatible with those of liberalism. Associations may encourage undemocratic and illiberal practices<sup>46</sup>: racism, sexism, xenophobia, aristocracy, hierarchy, and machismo among many others. There is no reason to suppose that within a liberal society, the so-called civic sphere will necessarily reproduce the values liberals praise<sup>47</sup>. However, the fact that associations are illiberal does not preclude them from contributing to the moral development of their members. So, rather than focusing on the effects that associations have on democracy (and on society in general), the ‘moral uses of pluralism’ insists that we should value to the effects they have on the development of members ‘personally and individually’.

---

<sup>45</sup> Ibid. 10. Emphasis added.

<sup>46</sup> Provided they satisfy the constraints discussed above, that is, they do not violate the basic rights of non-members (and probably members), their right to exit is effective, entrance is voluntary and they do not enforce criminal law.

<sup>47</sup> Cf. Yael Tamir, ‘Revisiting the Public Sphere’, in Gutmann, A. (ed.) *Freedom of Association* (Princeton: Princeton University Press, 1998), 214-38.

It seems that if the government were to impose liberal practices all the way down its attempts would fail on two grounds. First, it would 'fail the fundamental test of treating its citizens as free and reasonable beings who have the right and responsibility to decide what kind of associational life is most conducive to their own well-being and well-being of their society'<sup>48</sup>. Second, it is likely that governmental intervention renders the 'moral uses of pluralism' self-defeating: in order to derive the benefits of cooperation and self-esteem members need to feel that the association is in an important sense 'theirs', that is, that they are participating on their own terms. It seems difficult to deliver the benefits of association (in terms of moral development) by governmental engineering<sup>49</sup>.

Rosenblum's defence of free association is essentially maximalist; she suggests that the benefits of associations outweigh the harmful effects. This is clear in the subject of exclusion, in which the 'moral uses of pluralism' suggests associations can legitimately exclude provided that certain conditions are met. In addition to the general constraints outlined above, Rosenblum suggests that many forms of exclusion are acceptable insofar as they do not entail second-class citizenship. In her discussions on the Jaycees case, Rosenblum argues that, by forcing the Jaycees to accept women as full members, the court was endorsing the view (which she rejects) that 'second-class *membership* in this sort of voluntary

---

<sup>48</sup> Gutmann, 'Freedom of Association', 22-3.

<sup>49</sup> This is an empirical claim: it could be that moral development is less likely to occur if enforced from 'above'. It could also be that moral development is essentially a by-product of other activities, and therefore impossible to orchestrate deliberately. Examples of activities that are essential by-products are discussed in Elster, *Sour Grapes*, Chapter 2. This claim seems too strong. There is no reason to assume that government attempts to promote moral development through associations is completely doomed to failure (although, it is possible that members will develop to a lesser extent than if left to associate in their own terms). A third possibility is that the value of development is conditional upon associations being independent from government. If this is the case, Rosenblum has not argued for this stronger claim. Unfortunately, it is not clear which of these three claims 'the moral uses of pluralism' defends.

association implies second-class *citizenship* tout court'<sup>50</sup>. In essence her claim is that public standing and membership are independent; that arbitrary discrimination does not compromise equal public standing. Rosenblum goes so far as to claim that the 'moral uses of pluralism' offers forms of self-respect that are incommensurable with the same good when distributed by public institutions<sup>51</sup>.

Moreover, Rosenblum argues that other forms of membership can compensate for the costs of exclusion. The strategy is similar to the 'more, better speech' approach discussed in Chapter 5. Whatever the damages of exclusion, the legitimate remedy for government is to establish new forms of associations, rather than force exclusionary ones to modify their admissions criteria. She seems to suggest that forcing associations to accept otherwise excluded members will not benefit anyone. On the one hand, existing members need some sort of 'gatekeeping' in order to retain control over their own affairs<sup>52</sup>. On the other, would-be members will not benefit from association because compelled association might 'impose[] heroic obligations on victims'<sup>53</sup>. To illustrate, imagine that you are not wanted in a private school due to your religious beliefs, but that the government has forced the school to accept members of your faith. Attending the school knowing that everyone has been forced to accept you might be psychologically demanding. According to 'the moral uses of pluralism', a better solution for the excluded group is to form an alternative association; women excluded from the Jaycees could form a women-only Jaycees type of association. Liberal governments are responsible for insuring the

---

<sup>50</sup> Rosenblum, 'Compelled Association', 87. Emphasis added.

<sup>51</sup> Rosenblum, *Membership and Morals*, 62.

<sup>52</sup> *Ibid.* 64.

<sup>53</sup> Rosenblum, 'Compelled Association', 92.

background conditions that ‘facilitate the ceaseless formation of new associations’<sup>54</sup>.

As she sums up:

There is little reason to suppose that [...] public policy should aim at mandating inclusiveness and eliminating rejections, dismissiveness, secretiveness, even discrimination despite the pain these cause those who are excluded [...] The principal thing is that alternative associations are available for those alien or unwanted. And that associations do not keep their members captive or permanently cut off; that members have a real choice of discontinuing affiliation<sup>55</sup>.

The second important argument against liberal congruence claims that the illiberal *private* practices and attitudes maintained within some associations will not spread to other spheres of society, and, in particular, not to the *public* domain in which individuals interact as equal citizens. So, according to Rosenblum, the main problem with the ideal of liberal congruence is that it assumes (without evidence) that ‘dispositions and practices shaped in one association spillover to other contexts’<sup>56</sup>. Instead, the ‘moral uses of pluralism’ believes that even if the racist private club promotes attitudes against black people among its members (by making jokes, discussing racist ‘science’, etc), qua citizens they can still detach themselves from those beliefs and respect black people (qua citizens) in the public sphere. Liberals, she claims, adopt a simplistic ‘transmission belt’, according to which the ‘formative effects’ of private associations will be transmitted to the public sphere<sup>57</sup>.

---

<sup>54</sup> Rosenblum, *Membership and Morals*, 60.

<sup>55</sup> *Ibid.* 64.

<sup>56</sup> *Ibid.* 38.

<sup>57</sup> *Ibid.* 48. This transmission belt works for both ‘good’ and ‘bad’ effects.



Rosenblum's argument against spillover effects holds that individuals 'exercise [their] capacities for discrimination and moral adaptation all the time even among seemingly close situations. This includes a refined capacity to *resist* spillover'<sup>58</sup>. Moreover, Rosenblum asserts that the experience of pluralism enhances people's ability to make distinctions between spheres and to adjust their behaviour according to different expectations. She writes, 'even if we are subject to (or inflict) prejudice, arbitrariness, or deference in one domain, we may be able to exhibit an iota of tolerance in public arenas of fairness in hiring. If spillover and reiteration were the only or principal dynamic, the confinement of attitudes, behaviour, and moral dispositions would be inconceivable'<sup>59</sup>.

This argument, however, seems too strong. For we can object that spillover and reiteration effects need not be the only or principal dynamic processes in order to be significant for political morality. There well may be other process occurring, and, if at least some spillover takes place, we need to assess the degree at which it jeopardises citizens' ability to exercise their sense of justice.

To sum up, the 'morality of association' holds that associational life encourages cooperation, which triggers the moral development of members. Associational life also operate as 'compensation to disappointed social and political expectations'<sup>60</sup> by providing an environment in which groups of excluded individuals can form their own associations, and where they too can then get the benefits of the 'moral uses of pluralism'. In this sense it counterbalances exclusions suffered in other spheres. Additionally, freedom of association enhances the capacity

---

<sup>58</sup> Ibid. 49. Italics in original.

<sup>59</sup> Ibid. 50.

<sup>60</sup> Ibid. 64.

of moral agents to discriminate among contexts and behave accordingly. For instance, one can make racist comments at home or at the club, but not in the street or at school.

Regarding the costs of associations and exclusions the ‘moral uses of pluralism’ claims that spillover effects do not occur<sup>61</sup>, that government should not enforce liberal practices without compromising the very goods that associations protect, and that second-class membership is independent from second-class citizenship. The only conditions required for this to be upheld is that governments guarantee a pluralist environment favourable to the formation of new associations, that existing associations respect the right (and the possibility) of exit of members and that no basic rights are violated.

### *The ‘Moral Uses of Pluralism’: A Critique*

Rosenblum’s dismissal of the spillover effects of private attitudes in the public sphere, along with the important value she allocates to associational goods, leads her to underestimate the costs of associational life on society and to somewhat overestimate the benefits on members’ lives. The impact of associations affects not only on their members’ lives, but society in general. We need a more refined understanding of social conditions that enhance citizens’ moral development (both members and non-members).

Rosenblum does not link democratic private associations and liberal democracy because she believes there is no guarantee private associations that are

---

<sup>61</sup> Or at least that we should not expect them to happen.

conducted by democratic practices will democratise the public sphere. For this reason, she argues that freedom of association must only be concerned with the experience of pluralism, taking as a 'metric' of success how much associations contribute to their member's abilities to cooperate. However, this line of reasoning seems too short-sighted. The ability of a racist club to enhance cooperation among its members is a short-term effect, which needs to be balanced against other long-term, enduring changes in society. For example, society's attitudes about racism may change if racist clubs are protected and/or people discriminated against on racist grounds might lose (even more) power, becoming (even more) disadvantaged. A problem with the 'moral uses of pluralism' is that it is myopic to all these long-term effects.

Rosenblum is worried about the social ecology of private associations. She argues that government must keep interference to a minimum, and that it must promote a wide range of plural associations. Nevertheless, it seems arbitrary to focus only on the social ecology of associations and not on their public ecology. This problem is also reflected in Rosenblum's solution to the problem of exclusions. Although she might be right that building a net of associations for disadvantaged groups could counterbalance injustices suffered through exclusion, she does not show why this is the best solution. Her argument again does not extend further than the claim that cooperation is fundamental to the moral development of members and that costs are somewhat negligible. It may well be the case that a more efficient way to neutralise the bad effects of certain associations could be to restrict their freedom to engage in certain activities (namely, those that are too costly to non-members). In order to neutralise the costs of an all-White Church (such as the change of attitudes

about non white persons, discrimination, and so on) for instance, it is possible, and perhaps probable, that forming an alternative church or a support group for excluded people excluded will not sufficiently counterbalance those costs. It is likely that restricting some of the all-White Church practices would to be more effective.

Restrictions on admissions policies seem to be especially costly to citizens who are prevented from joining racist associations. It could be objected that the costs they face can be reduced insofar as there are alternative (non racist) associations which they could form. According to some, this response must acknowledge that there is a loss; that in one aspect the lives of racists go worse by the prohibition of racist associations. Although internalising the costs might be justified (given the impact these groups have in society), perhaps racists should be compensated for their being worse off<sup>62</sup>.

This objection is not convincing. First, one could conditionalise the interests a person has in joining an association, stating that associational goods are conditional on associations being just. According to this response, the moral development achieved within a racist association has no weight. In this respect, the 'deficit' in moral development suffered by racists is only *prima facie*.

A second, more modest, response holds that the moral development that associations facilitate is not 'tied' to any particular association being available. Following Raz's idea that autonomy requires an adequate range of options, the absence of a particular association may not be detrimental to a person achieving the goods that associations offer if there are others available. It is difficult to see why the goods delivered by fox-hunting associations cannot be also delivered by drag-

---

<sup>62</sup> Cf. Chapter 4 in relation to speakers' interest in having hate-speech protected.

hunting associations. This is another aspect in which Rosenblum overestimates the goods of associations. Associational freedom might be significantly beneficial to members, but it does not follow that the goods of having the option to become a member of any particular association is similarly significant<sup>63</sup>.

This recognises that there is a loss, but one which can be redressed by alternative options. Another response would be to conditionalise the interest a person has in joining an association, stating that associational goods are conditional on such groups not being unjust.

However, it could still be argued that society should bear the long-term costs of associations on the basis that public policy must be concerned only with public attitudes and not with private beliefs and practices. This idea takes us to the core of Rosenblum's argument: the rejection of spillover effects. As noted above, Rosenblum holds that we have a refined capacity to discriminate between contexts. Thus, privately held beliefs and prejudices will not spillover into public life. Although this is an attractive idea, there is an important collection of evidence that suggest Rosenblum is far too optimistic about our mental control.

Rosenblum writes that 'the logic of [liberal] congruence does not come automatically equipped with a social or psychological dynamics to explain why dispositions cultivated in one association can be expected to be stable and transmitted to other spheres'<sup>64</sup>. It is true that liberalism itself does not provide a psychological theory that explains why spillover sometimes happens. However, research over the past 20 years has shown that much of our behaviour is caused by automatic unconscious responses rather than acts of will. Many of these findings

---

<sup>63</sup> Thanks to Matthew Clayton for this suggestion.

<sup>64</sup> Rosenblum, *Membership and Morals*, 38.

suggest that attitudes cultivated in one sphere are likely to spillover into other spheres, bypassing the agent's control.

As we have discussed in past chapters, there is a significant amount of evidence to support this claim. These responses occur regularly, and are directly caused by the environment. Automatic behaviours include not only bodily movements, but also the pursuit of goals. In general, it does not matter whether individuals are aware of the stimuli that cause their behaviour. Although automatic responses are not unavoidable, in most cases they are uncontrollable. Neutralising automatic behaviour requires significantly more mental resources than people usually have at their disposal.

A significant portion of automatic behaviour is related to 'implicit associations'. As Greenwald and Banaji write 'the signature of implicit cognition is that traces of past experience affect some performance, even though the influential earlier experience is not remembered in the usual sense—that is, it is unavailable to self-report or introspection'<sup>65</sup>. For instance, some individuals were presented with the task of generating complete words in response to an incomplete series of letters and word fragments. The answers given were most commonly words from a list to which these persons had been exposed to previously (with or without their awareness)<sup>66</sup>. In another experiment, some individuals were shown two series of letters. They were asked to press a button when both series formed actual words (e.g. CAR/HOUSE), and a different button when at least one did not (e.g. APPLE/KUPOD). Interestingly individuals took longer to recognise words that were

---

<sup>65</sup> Anthony Greenwald and Mahzarin Banaji, 'Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes', *Psychological Review*, 102 (1995), 4.

<sup>66</sup> *Ibid.* 4.

unrelated (e. g. BREAD/NURSE, BUTTER/DOCTOR) than words that were related (e.g. BREAD/BUTTER, NURSE/DOCTOR)<sup>67</sup>.

Another component of implicit social cognition, as we have reviewed, is social stereotypes: 'a cognitive structure that contains the perceiver's knowledge, beliefs, and expectancies about some human group'<sup>68</sup>. As noted above, in Western societies black people are associated with laziness, aggression, rhythm, low intelligence and poverty among others traits. White people, in contrast, are associated with being intelligent, clean and successful. Gender stereotypes relate women with dependency and underachievement, while males are related with aggressiveness and success<sup>69</sup>. Not surprisingly, Greenwald and Banaji report that readers of academic essays assume that the best essays have been written by men rather than women<sup>70</sup>.

There is an important body of evidence which suggest that even people who describe themselves as being low-prejudiced<sup>71</sup> behave according to the stereotype under certain circumstances. This is caused by the automatic and implicit elements of stereotype activation, just as other representation stereotypes become automatically activated when features of the social group are perceived. In an

---

<sup>67</sup> John Dovidio, Nancy Evans and Richard Tyler, 'Racial Stereotypes: The Contents of Their Cognitive Representations', *Journal of Experimental Social Psychology*, 22 (1986), 23. Another example of implicit associations is the 'halo effect' discussed in Chapter 2.

<sup>68</sup> Hamilton, D. and Troler, T., 'Stereotypes and Stereotyping: An Overview of the Cognitive Approach', in Gaertner, S. and J. Dovidio, (eds.) *Prejudice, Discrimination, and Racism* (New York: Academic Press, 1986), 133. Quoted by Greenwald and Banaji, 'Implicit Social Cognition', 15.

<sup>69</sup> Cf. Devine, 'Stereotypes and Prejudice'; Greenwald and Banaji, 'Implicit Social Cognition'.

<sup>70</sup> Greenwald and Banaji, 'Implicit Social Cognition', 15.

<sup>71</sup> In this context being low prejudiced means to *explicitly* reject the content of the stereotype. There are several exercises that measure people's levels of prejudice. For instance, see John McConahay, Bettee Hardee, and Valerie Batts, 'Has Racism Declined? It Depends upon Who's Asking and What is Asked', *Journal of Conflict Resolution*, 25 (1981), 563-579. For discussion see: Devine, 'Stereotypes and Prejudice'; Dovidio, Evans and Tyler, 'Racial Stereotypes'; Dovidio, Kawakami, et. al., 'On the Nature of Prejudice'; Dovidio and Gaertner, 'Aversive Racism'. Irwin Katz and Glen Hass, 'Racial Ambivalence and American Conflict Value: Correlational and Priming Studies of Dual Cognitive Structures', *Journal of Personality and Social Psychology*, 55 (1988), 893-905.

experiment Dovidio and his colleagues<sup>72</sup> primed subjects with schematic faces of Black and White people. The faces were flashed for a fraction of a second so that the subjects were not aware of it. Later they had to pair a string of letters ('PPPPPP'— a supposed cue for persons —or 'HHHHH'— a supposed cue for houses) that were used as distractors with positive ('good', 'kind' and 'trustworthy') or negative traits ('bad', 'cruel', 'untrustworthy'). The results show that people primed with faces of African-Americans reacted faster to negative traits than to positive ones, and that when they were primed with White faces they reacted in the opposite fashion. This is because negative traits are associated with Blacks, and the positive traits with Whites. The races are stereotyped. It is noticeable that the results obtain independently of whether people were low- or high-prejudiced toward Blacks. As Devine writes 'priming will automatically activate the cultural stereotype for both those high and low in prejudice'<sup>73</sup>. It is not only racial or gender stereotypes that elicit automatic responses: people primed with words related to the elderly walked slower than neutral controls<sup>74</sup>, and people asked to think about 'intelligent' stereotypes (such as university professors) did better at general knowledge tasks than people who thought about 'stupid' stereotypes (such as football hooligans)<sup>75</sup>. As noted above these findings have generated the idea of 'aversive racism' —the notion that people who endorse egalitarian, non-prejudiced beliefs suffer from mental

---

<sup>72</sup> Dovidio, Kawakami, et. al., 'On the Nature of Prejudice', Exp. 1. See also Devine, 'Stereotypes and Prejudice', Exp 2.

<sup>73</sup> Devine, 'Stereotypes and Prejudice', 5. Dijksterhuis and Bargh explain the dynamics of stereotype activation as follows: 'the effects of stereotype activation in changes in overt behaviour can be explained by a series of steps. First, stereotypes activate associated traits. These traits, in turn, activate more concrete behaviour representations. Finally, these behaviour representations activate the motor programs responsible for actual behaviour'. Dijksterhuis and Bargh, 'The Perception-Behavior Expressway', 24.

<sup>74</sup> Bargh, Chen, and Burrows, 'Automaticity of Social Behavior'; see also Chapter 2.

<sup>75</sup> John Bargh, 'The Cognitive Monster: The Case against the Controllability of Automatic Stereotype Effects', in Chaiken, S. and Y. Trope (eds.) *Dual-Process Theories in Social Psychology* (New York: Guildford Press, 1999), 374.



contamination and display negative feelings and beliefs towards people of other races.

Another important collection of evidence suggests that a similar, automatic mechanism operates regarding goal-oriented behaviour. They are also mental representations (just as stereotypes and concepts) capable of become automatically activated<sup>76</sup>. For instance, a group of students were presented with two 10X10 matrices of letters with some words hidden in them. Matrix 1 contained some words related to high performance such as 'win', 'compete', 'succeed', 'master', as well as some neutral words, such as 'lamp', 'building', and so on. Matrix 2 had the same neutral words but those related to high performance were substituted for other neutral words ('carpet', 'river', 'robin'). Not surprisingly, students assigned to Matrix 1 performed significantly better than those assigned to Matrix 2. The idea of this experiment was to activate the goal to perform well without students' awareness<sup>77</sup>. It seems to have succeeded.

A person regularly making the same choice in a particular situation develops a tendency to employ the same mental processes, which, in turn, leads to a removal of the conscious role in the process<sup>78</sup>. Think about the process of driving a car. It is by repeating the same action in particular situations many times (pressing the clutch pedal – changing the gear – releasing the clutch pedal slowly – pressing the

---

<sup>76</sup> John Bargh, Peter Gollwitzer, et. al., 'The Automated Will: Nonconscious Activation and Pursuit of Behavioral Goals', *Journal of Personality and Social Psychology*, 81 (2001), 1015.

<sup>77</sup> Ibid. Exp. 1, 1016-7.

<sup>78</sup> See diagram in Tanya Chartrand and John Bargh 'Automatic Activation of Impression Formation and Memorization Goals: Nonconscious Goal Priming Reproduces Effects of Explicit Task Instructions', *Journal of Personality and Social Psychology*, 71 (1996), 469.

accelerator pedal) that the entire process becomes automatic; experienced drivers do not need to focus their attention on this process<sup>79</sup>.

An important feature of automatic goal activation is that it produces the *same* behaviour as goals performed at will. In this sense, automatic behaviour bypasses about the source of activation; once the mechanism is activated it will produce the same conduct<sup>80</sup>. For instance, in another experiment, Bargh and his colleagues<sup>81</sup> activated the goal of cooperation in some students by using the Scrambled Sentence Test (the priming words used were 'dependable', 'helpful', 'support', 'reasonable', 'honest', 'cooperative', 'fair', 'friendly', 'tolerant' and 'share'). Later, they had to undertake a resource-dilemma task against a presumed other player<sup>82</sup>. They were to catch fish (that had to be replenished periodically) from a common pool. Every turn they had the option of keeping all the fish they had caught or returning some to the lake. In the game's instructions, some students were told that they had to cooperate so that the common pool was not depleted. Results show that people who were primed to cooperate, returned almost as many fish (31.1) as the students who were explicitly told to do so (32.1). People who were both primed and told to cooperate returned more than all the others (35.1). In contrast, students who were neither primed nor told to cooperate returned the least (24.9). After the experiment, when students were asked to report about their willingness to cooperate those primed did not mention that they had such a goal. These

---

<sup>79</sup> This mechanism is not 'habit' as conceptualised by classical behaviourist theory, but 'instead is behaviour that is flexibly responding to environmental events as they unfold in the ongoing situation'. Bargh, Gollwitzer, et. al., 'The Automated Will: Nonconscious Activation and Pursuit of Behavioral Goals', 1025.

<sup>80</sup> Chartrand and Bargh, 'The Chameleon Effect', 476.

<sup>81</sup> Bargh, Gollwitzer, et. al., 'The Automated Will: Nonconscious Activation and Pursuit of Behavioral Goals'. Exp. 2.

<sup>82</sup> The game resembles a prisoner's dilemma: the best outcome for both is to cooperate, but the rational individual action is not to cooperate.

findings suggest that nonconscious-goal effects on behaviour do not require the individual to become consciously aware of the fact that this goal is being pursued in the situation. They also support the notion that participants who are unaware of the activation of nonconscious goals will remain unaware of their subsequent operation to guide behaviour<sup>83</sup>.

How does all this evidence relate to the ‘moral uses of pluralism’? In her criticisms of the idea of liberal congruence, Nancy Rosenblum rejects the view which holds that ‘the relation between civil society (that is, our associational life) and liberal democracy *is, or must be, reciprocally supportive*’<sup>84</sup>. This claim has two different parts. One is a normative assertion, and will be addressed in the next section. The other is an empirical one. In contrast to it, the evidence surveyed above provides reason to suspect that spillover effects *are likely to happen*<sup>85</sup>. Even if the evidence here is not definitive, we have a *prima facie* reason to doubt Rosenblum’s optimism about the ability of individuals to compartmentalise and contain their racist attitudes in private. It seems we do not have the capacity to control our behaviour to this extent, and we cannot make the successful context-sensitive discriminations that she assumes possible. For there are situational features in the environment that trigger responses, of which we are not aware, and usually do not control. As discussed in previous chapters, neutralising mental contamination<sup>86</sup> requires a significant amount of mental resources. And even with this there is no guarantee of

---

<sup>83</sup> Bargh, Gollwitzer, et. al., ‘The Automated Will: Nonconscious Activation and Pursuit of Behavioral Goals’, 1024.

<sup>84</sup> Rosenblum, *Membership and Morals*. 10.

<sup>85</sup> In the next section I will defend the view that there are normative reasons to support a form of congruence.

<sup>86</sup> Mental contamination is defined subjectively as people reacting in ways they would not approve. Cf. Wilson and Brekke, ‘Mental Contamination and Mental Mediation’; see also Chapter 2.

success. A preferred alternative might be to control exposure to contaminating factors.

Features in the environment activate automatic behaviour, and we do not control this. We are as likely to respond in a way 'X' to stimulus 'A' whether it happens in public or in private. People who associate Blacks with negative features and display negative attitudes towards them will do so in both spheres, because we do not come equipped with a public/private discrimination device (as Rosenblum implies we do). These findings suggest that we must be more vigilant about the ecology of speech and association, and about socialisation processes not only in children (who are very vulnerable<sup>87</sup>) but in adults too. Given the problems posed by mental contamination, it seems that we have an interest in not being exposed to certain stimuli that at some other point of our lives will cause behaviour we would regret. These problems highlight the harmful costs of spillover effects, not only to direct targets of racial stereotypes and racist- and gender-based exclusions, but to other members of society too. Although mental contamination (presumably) will not affect a person who is more likely to join a racist club, church or political party, it affects a large number of citizens by setting the environmental conditions that trigger, for instance, aversive racism. In this sense the creation and maintenance of racist institutions that produce, reproduce and pass on racial stereotypes and attitudes poses a direct threat to the moral development of an important section of the population.

---

<sup>87</sup> Young children are cognitively vulnerable in that, for instance, they have not yet fully developed the structures that allow rejecting propositions likely (Cf. Gilbert, 'How Mental Systems Believe'). Also, they have yet to reach a minimal threshold of autonomy. Cf. Clayton, *Justice and Legitimacy in Upbringing*.

The argument I develop here highlights another shortcoming of the ‘moral uses of pluralism’. Rosenblum believes that the primary element of a person’s moral development is the ability and willingness to cooperate with other members of one’s association. Although this surely is an important element, it is not a complete. As noted above, citizens have fundamental interests in developing and exercising a sense of justice and in devising, reviewing, and pursuing a reasonable conception of the good. These interests motivate a normative claim regarding the public ecology of associational freedoms, and support the claim that a modest version of liberal congruence is appealing. In the next section, I will develop this view.

### III

#### Public Reason, and Reasonable Association

In this section I will discuss an idea of freedom of association constrained by the notion of reasonableness. In this sense, I address Rosenblum’s claim that liberal congruence is normatively inappropriate. The main claim I will defend is that restricting associations on grounds of reasonableness better protects the associational goods described in the first section. Like Rosenblum, the view I will defend begins by acknowledging the fact of pluralism, and the effects it has on the moral development of citizens. However, in sharp contrast, my argument for reasonable association takes a wider view by recognising the ecological effects of pluralism on society, and not only those on an association’s members.

*The Condition of Reasonableness*

The notion of reasonableness emanates from acknowledging that we live in a society deeply divided by different comprehensive and incompatible doctrines of the good. People hold different ideas about the value of life, the sources of morality, our place in the universe, the nature of religion, and so on. A characteristic feature of this pluralism is that different doctrines are incompatible. There is no (more comprehensive) doctrine that accommodates them all. Pluralism is a fact that can be neither ignored nor fought. As Rawls states, it must be assumed that it 'is the normal result of the exercise of human reason within the framework of the free institutions of a constitutional democratic regime'<sup>88</sup>.

Rawls argues that there is a constraint on the plurality of comprehensive doctrines to which public justification is required: pluralism must be reasonable. Reasonable persons have two basic features. First, reasonableness involves a dimension of *reciprocity*, according to which persons are willing to engage in practices the rules of which no one could reasonably reject<sup>89</sup>. According to Rawls, '[p]ersons are reasonable in one aspect when, among equals say, they are ready to propose principles and standards as fair terms of cooperation and to abide by them willingly, given the assurance that others will likewise do so'<sup>90</sup>. Reasonable persons not only are moved by self-interest, they also seek for its own sake to cooperate with others in terms all can accept<sup>91</sup>. So, being reasonable imposes a demand to recognise others as free and equal. Citizens are free in the sense that a) they can exercise their

---

<sup>88</sup> Rawls, *Political Liberalism*, xviii.

<sup>89</sup> Cf. Scanlon, *What We Owe to Each Other*, 191-7.

<sup>90</sup> Rawls, *Political Liberalism*, 49

<sup>91</sup> *Ibid.* 50.

moral powers (see below), b) they are self-authenticated source of valid claims<sup>92</sup> and, c) they are responsible for their ends, and capable of adjusting these ends in light of demands of justice<sup>93</sup>.

Reasonable persons, must also, be willing to accept the '*burdens of judgement*'<sup>94</sup>. The burdens of judgement explain (together with the free exercise of reason) the sources of the plurality of doctrines in our societies. When we contemplate, say, the existence of God, the demands of morality, the meaning of life, the virtues of nuclear power stations, or the case for animal rights, there are many different considerations that must be taken into account. The burdens of judgment include, among other things, the complexity of evidence and the hardness of assessing it (both practically and theoretically), the disagreement about the relative weight of different considerations, the problem of conceptual vagueness and the existence of difficult, borderline cases, the manner and degree in which our personal experience influence our attitudes towards certain values rather than others, and so on<sup>95</sup>. The burdens of judgement explain why different people reach different reasonable doctrines when considering these issues. It is important to note that reasonable persons tolerate diverse views because they believe that such plurality is the outcome of the free exercise of reason, and also because they desire to engage with others on fair terms of cooperation (according to which people must be able to exercise their reason freely), not because they lack the power to 'convert' others to

---

<sup>92</sup> Ibid. 32.

<sup>93</sup> Ibid. 33-4.

<sup>94</sup> Ibid. 54-8.

<sup>95</sup> Ibid. 56-7. Rawls also writes: 'Citizens are reasonable when, viewing one another as free and equal in a system of social cooperation over generations, they are prepared to offer one another fair terms of cooperation according to what they consider the most reasonable conception of political justice; and when they agree to act on those terms, even at the cost of their own interests in particular situations, provided that other citizens also accept those terms'. Rawls, 'The Idea of Public Reason Revisited', 136.

their own views. Similarly, reasonable persons do not regret the fact of reasonable pluralism. While the first aspect of reasonableness involves a moral element (namely, the idea of reciprocity) the second aspect can be characterised as having an epistemic one. The idea of the reasonable is nicely displayed by Rawls:

reasonable persons see that the burdens of judgment set limits on what can be reasonably justified to others, and so they endorse some form of liberty of conscience and freedom of thought. It is unreasonable for us to use political power, should we possess it, or share it with others, to repress comprehensive views that are not unreasonable<sup>96</sup>.

The demands of reasonableness, then, can be understood as the disposition to understand society as a fair system of cooperation, and to propose and honour fair terms of cooperation, between free and equal citizens<sup>97</sup>. These demands are derived from a view of citizens as free and equal and correspond to the fundamental moral powers that citizens are described to possess<sup>98</sup>. As noted above, citizens have two highest order interests (each corresponding to their 'moral powers'). First, a person is interested in developing and exercising the capacity to understand, to apply, and to act from a conception of justice<sup>99</sup>. Second, a person has also a fundamental interest in forming, revising and rationally pursuing a conception of the good. Individuals are also assumed to have, at any given time, a particular conception of the good, which raises another higher order interest (albeit one less fundamental than those mentioned above) in pursuing and advancing their own particular conception of the

---

<sup>96</sup> Rawls, *Political Liberalism*, 61.

<sup>97</sup> Cf. '...in attributing freedom equality and the basic moral powers to others a reasonable citizen thereby takes a series of commitments. She commits herself to beliefs and actions compatible with a respect for the other's basic moral powers' James Boettcher, 'What is Reasonableness', *Philosophy and Social Criticism*, 30 (2004), 606.

<sup>98</sup> Reasonableness is a disposition and not a moral power because it embodies a willingness to exercise those moral powers. Cf. *Ibid.* 604.

<sup>99</sup> Rawls, 'Kantian Constructivism in Moral Theory', 312.



good<sup>100</sup>. These powers entail the development and exercise of critical and deliberative capacities that are necessary in order to endorse a sense of justice and to develop, affirm and revise a conception of the good. Here is one of the important points at which the ‘moral uses of pluralism’ and the approach I defend overlap: at the core of both there is a concern for the moral development of individuals.

### *Public Reason*

Given the fact of pluralism, on what grounds can reasonable people reach an agreement (acceptable to them all) about the basic institutions of society? According to Rawls, in order to be acceptable citizens’ political claims must be made in accordance with public reason<sup>101</sup>. ‘The idea of public reason’, writes Rawls, ‘specifies at the deepest level the basic moral and political values that are to determine a constitutional democratic government’s relation to its citizens and their relation to one another’<sup>102</sup>. The fundamental features of public reason are that it must present a political conception of justice not dependent on comprehensive doctrines and that it must satisfy the criterion of reasonableness discussed above<sup>103</sup>. As noted in previous chapters, a *political* conception of justice is necessary given the fact of reasonable pluralism. Essentially, because individuals differ deeply in the doctrines they endorse, the only means of reaching an ‘overlapping consensus’<sup>104</sup> is by presenting conceptions of justice that are independent (‘freestanding’, as Rawls

---

<sup>100</sup> Ibid. 313.

<sup>101</sup> Rawls, *Political Liberalism*, 212-54; ‘The Idea of Public Reason Revisited’, 131-80.

<sup>102</sup> Rawls, ‘The Idea of Public Reason Revisited’, 132.

<sup>103</sup> It is important not to confuse public reason with reasonableness; the latter is necessary, but not sufficient for the former.

<sup>104</sup> Rawls, *Political Liberalism*. Lecture IV.

writes) from the doctrines people endorse<sup>105</sup>. Political conceptions have three fundamental features: ‘they apply to the basic structure of society’, ‘they can be presented independently from comprehensive doctrines’, and ‘they can be worked out from fundamental ideas seen as implicit in the public political culture of a constitutional regime’<sup>106</sup>. Citizens’ political conceptions of justice must be reasonable in the sense stated above: they must satisfy the criterion of reciprocity and must accept the burdens of judgement<sup>107</sup>. To sum up, by satisfying these criteria public reason can form the basis for an overlapping consensus among free and equal citizens who hold different and incompatible reasonable comprehensive views.

There are two important considerations regarding public reasons. The first concerns the subjects to whom it applies; the second concerns its scope. Public reasons, Rawls argues, binds only the ‘discourse of judges on their decisions [...]; the discourse of government officials, especially chief executives and legislators; and finally, the discourse of candidates and their campaign managers’<sup>108</sup>. Although other agents are free from the demands of public reason, Rawls still insists that ‘ideal citizens are to think of themselves as if they were legislators’<sup>109</sup>. In this respect, citizens have a moral (not legal) duty to justify their decisions in terms of public reasons.

---

<sup>105</sup> Independence from comprehensive doctrines does not mean that political conceptions cannot appeal to any values. It implies that values directly dependent on the soundness (or truth) of comprehensive doctrines are eschewed. Nevertheless, political conceptions can (and surely must) appeal to *political* values.

<sup>106</sup> Rawls, ‘The Idea of Public Reason Revisited’, 143.

<sup>107</sup> Cf. Rawls’s ‘principle of liberal legitimacy’. Rawls writes that ‘our political power is fully proper only when it is exercised in accordance with a constitution the essentials of which all citizens as free and equal may reasonably be expected to endorse in the light of principles and ideals acceptable to their common human reason’ Rawls, *Political Liberalism*, 137.

<sup>108</sup> Rawls, ‘The Idea of Public Reason Revisited’, 133.

<sup>109</sup> *Ibid.* 135.

Second public reason is to be applied only to certain decisions. There is some ambiguity in Rawls's description of the scope of public reason<sup>110</sup>. On the one hand, the *narrow* view suggests that political liberalism

strives[s] for publicly based justifications for questions regarding the constitutional essentials and basic questions of distributive justice but not in general for all the questions to be settled by the legislature within a constitutional framework. We should distinguish, then, between these two cases. The first attainable [...] and desirable, the second neither attainable nor desirable<sup>111</sup>.

It is unclear, however, why we should restrict public reason only to constitutional essentials and matters of basic justice. It seems that insofar as binding decisions are taken, those decisions must be justifiable to all reasonable citizens. We can concede that there is degree of urgency regarding constitutional essentials that the funding of art galleries might not have, but this is not a reason to conclude that non-essentials should not satisfy public reason. In this respect it seems a natural interpretation of Rawls's idea of legitimacy suggests a different, *broad*, view, according to which 'the ideal of public reason ought to be applied, whenever possible, to all political decisions where citizens exercise coercive power over one another'<sup>112</sup>.

A possible objection to this broad view might be to claim that public reasons concern *only* decisions whose subject is the basic structure of society<sup>113</sup>. As noted in the first chapter, the nature of the basic structure might be interpreted in two

---

<sup>110</sup> Cf. Jonathan Quong, 'The Scope of Public Reason', *Political Studies*, 52 (2004), 233-50.

<sup>111</sup> Rawls, *Justice as Fairness*, 91 n.13. In contrast, in *Political Liberalism*, Rawls grants that the focus on constitutional essentials is the strongest case for public reasons. He suggests that once essentials are justified, then 'it is usually desirable to settle political questions by invoking the value of public reason. Yet, this may not always be so'. Rawls, *Political Liberalism*, 215.

<sup>112</sup> Quong, 'The Scope of Public Reason', 234.

<sup>113</sup> Here I return to some issues discussed in Chapter 1

different ways. The ‘coercive-based’ interpretation claims that only coercive institutions are to be included. The ‘profound effects’ interpretation, in contrast, suggests it is sufficient for institutions to have deep effects on citizens’ lives in order to be the subject of political justice<sup>114</sup>. The latter interpretation is appealing because it acknowledges that society’s public ecology affects the kind of persons we are: it influences our ambitions, development of talents, the conceptions of the good we pursue, and the capacity for a sense of justice. In addition, institutions belonging to the public structure must provide public guidelines to citizens acting in good faith<sup>115</sup>.

One of the reasons liberalism values freedom of association is its capacity to deliver certain goods. These associational goods are valuable because they affect the moral development of citizens as well as their autonomy. Indeed, as I have suggested, this is the main reason we think that associations warrant special protection despite the harmful effects they might have on third parties. If we accept the profound effects reading of the basic structure, then we can include associations within it. Moreover, their belonging to the basic structure could explain why associational freedom is so important for liberalism.

This is not to argue that principles of justice should *directly* rule the internal life of associations<sup>116</sup>. There is a difference between the kind of reasons by which we justify the organisation of institutions and the principles by which they are governed<sup>117</sup>. Associations can produce admissions policies that are not themselves

---

<sup>114</sup> Cohen, ‘Where the Action Is’ 18.

<sup>115</sup> Cf. Chapter 1; Williams, ‘Incentives, Inequalities, Publicity’; Clayton, *Justice, and Legitimacy in Upbringing*, 35-40.

<sup>116</sup> Rawls, ‘The Idea of Public Reason Revisited’, 158-9.

<sup>117</sup> Rawls does not think that justice as fairness is the only political conception of justice, nor that public reason gives a definite answer to all political problems. Regarding some issues it might be underdetermined or incomplete. This is no a reason, though, to reject the broad view. For an elegant

consistent with principles of justice but which have, nevertheless, been sanctioned by public reasons. Imagine, for instance, that we are forming a Symphonic Orchestra and we need to decide who can participate. After deliberation, we agree on a membership policy based exclusively on natural talents. Everyone wishing to be included must pass a ‘musical talents test’<sup>118</sup>. Although the membership policy is, in some sense, inconsistent with the principles of justice, if the procedures for reaching that decision and its justification are based on public reasons, then it seems legitimate.

Note, however, this does not validate racist decisions. The main problem is that it is difficult to imagine procedures consistent with public reason whose outcomes are racist policies. Moreover, it seems that *any* procedure which can produce such an unreasonable outcome can be questioned, and presumably rejected from the standpoint of public reason.

### *Reasonable Association*

How does the argument developed in this section relate with the associational goods discussed at the beginning of the chapter? The important question is this: are there legitimate public reasons to which associations can appeal in order to restrict membership on racist grounds? To respond in the positive associations must show that their discriminatory practices are justified by a reasonable freestanding political conception of justice. In this respect, equality, rather than discrimination should be

---

defence of public reason see Andrew Williams, ‘The Alleged Incompleteness of Public Reason’, *Res Publica*, 6 (2000), 199-211.

<sup>118</sup> Assume for the sake of the argument that there is a machine which grades musical talent so that everyone who scores over eight can enter the Orchestra.

the rule. And any deviation from it must be publicly justifiable. The burden of the proof is on excluder, not on the excluded.

**Authenticity Reassessed.** As noted above, citizens have a major interest in forming, revising and pursuing a conception of the good, and in living in accordance with their own convictions and ideals. Obviously, this interest does not grant an absolute right: authenticity is always constrained by the rights of others. Charles does not have a right to stab Manchester United fans merely because he wants to live according to his convictions. There are two ways in which we can understand this conflict. First, we can think that we must balance the authenticity of Charles's interests in stabbing Manchester United fans against those of his victims. In our conclusion, we would need to acknowledge that there is a loss in authenticity from Charles's point of view, but that this is justified by the protection of the interests of his victims.

Another, more attractive, interpretation takes a different stand: it qualifies our rights by claiming that we have no right to be unreasonable. A person is unreasonable if she rejects any of the following claims: 1) 'that political society should be a fair system of social cooperation for mutual benefit', 2) 'that citizens are free and equal', and 3) 'the fact of reasonable pluralism'<sup>119</sup>. Under this interpretation, the conflict between Charles's authenticity and the interests of his victims is only *prima facie*: because Charles view is unreasonable there is no conflict to be balanced. Charles suffers no harm in being prevented from living according with his view.

---

<sup>119</sup> Jonathan Quong, 'The Rights of Unreasonable Citizens', *The Journal of Political Philosophy*, 12 (2004), 315.

If my argument is valid then racist claims based on authenticity grounds are excluded because they are unreasonable and costly to others (both to victims and to bystanders). The costs can be divided on costs to direct victims. Racist (as well as other exclusive) associations deprive victims of economic, social and political goods<sup>120</sup>: they usually deny access to employment<sup>121</sup>, pressure groups<sup>122</sup>, education<sup>123</sup>, civic activities<sup>124</sup>, and so on. Racist associations are also costly to non-direct victims: as noted above, they create, enforce and transmit stereotypes that promote and sustain aversive racism. We, thus, have a prima facie clash of authenticity claims. On the one hand, there are racists who want to live according to their own views, and on the other, there are citizens who do not want to be racists,

---

<sup>120</sup> White discusses the goods that ‘categorical’ exclusion has on victims. He, nevertheless, reaches a very different conclusion. He argues that expressive and intimate associations have ‘strong presumption of legitimacy’ and can only be overridden when there is a risk of ‘profound material deprivation, e.g. starvation, or exclusion from a decent minimum of education’. White, ‘Freedom of Association and the Right to Exclude’, 379 n. 11.

<sup>121</sup> For instance, anti-discrimination laws in the US forbid denying employment on grounds of race, gender, and so on. However, associations might be exempt if they discriminate on religious grounds or for expressive reasons. Cf. Rosenblum, *Membership and Morals*; Greenawalt, ‘Religious Association’. Also, exemptions can be granted to small businesses on the grounds that they are family businesses (and thus are protected by intimacy). Rosenblum, ‘Compelled Association’, 89. But see *Boy Scouts of America et. al. v Dale* (99-699) 530 U.S. 640 (2000) in which the Court ruled out the attempt of the Boy Scouts to fire Dale because he was a homosexual. The court ruled that anti-discrimination law had precedent because Dale’s homosexuality neither threatened the First Amendment rights of the Boy Scouts, nor significantly affected members’ ability to carry out their purposes.

<sup>122</sup> The Jaycees case is a good example of how exclusion deprives people from networks and influential groups.

<sup>123</sup> Until recently, there were still segregated schools. Nowadays, racial segregation in schools has been banned. Cf. *Runyon v. McCrary*, 427 U.S. 160 (1976). For instance, Bob Jones University did not allow Black students to enrol until 1976, and until 2000 interracial dating was forbidden. <http://archives.cnn.com/2000/US/03/04/bob.jones/> [accessed 18/09/07]. But cf. *Parents Involved in Community Schools v. Seattle School District No. 1 et. al.* 551 U.S. \_\_\_\_ 2007.

<sup>124</sup> Cf. *N.Y. City Ancient Order of Hibernian v. Dinkins*, 814 F. Supp 358 (1993) in which New York city’s Irish Lesbian and Gay Organization were denied the right to participate in the St. Patrick’s Day Parade. Also see. *Welsh v. Boy Scouts*, 993 F.2d 1267 (1993). M. Welsh, a 7 years old boy, was excluded from the Boy Scouts for refusing to express a belief in god. The court found that the Boy Scouts had a right to exclude him because believing in God was a fundamental part of the Boy Scout’s value system.

but who behave as such without awareness. According to my argument, racists' claims have no force in virtue of being unreasonable<sup>125</sup>.

**Intimacy Reassessed.** Liberalism authorises private spaces in which no intervention seems to be justified. These 'brackets of privacy'<sup>126</sup> involve the spheres of sex, marriage, close friends, family, and so on. It seems that here people can legitimately exclude for whatever reasons they so wish (or even for no reason at all). I can refuse to invite someone to my birthday party based on a whim, and without offering explanations to anyone.

Is racist behaviour to be tolerated in intimate associations (for instance, think of someone whose choice of partner is driven by racist preferences)? The question is difficult to answer. On the one hand, it seems that, just as in the birthday party example, a person is free to choose intimate companionship for whatever reasons she wants. But, on the other hand, it also seems that if a person is to have a (reasonable) sense of justice she should act in accordance with it. If this is the case, then a person should have no right to choose sexual partners based on racist preferences. Nevertheless, there is additional problem because many times we cannot know whether we are making decisions based on racist views or not<sup>127</sup>. Consider the aversive racist: he sincerely believes that his behaviour is not racist, although it is. It seems that certain decisions (such as who to marry) are so complex and involve so many attitudes operating at different levels of awareness that it is difficult to provide

---

<sup>125</sup> It seems that even if we accept that citizens have a right to be unreasonable, the balance between racists and aversive racists would be inclined towards citizens who want to neutralise the effects of racism.

<sup>126</sup> I owe this phrase to Fabienne Peter.

<sup>127</sup> I do not mean to claim that we never know. It is clear that someone can disregard people from a certain race as a possible sexual partner. However, in most cases the dynamics of sexual attraction are too complex, and the ultimate reasons for our preferences might not be transparent to us.



a definitive answer. In this sense, we could say that, in certain aspects, racists are protected by the opaqueness of the mind<sup>128</sup>.

There is an important caveat to this problem: families within the basic structure are obliged to educate children in a manner that facilitates the development of their moral powers<sup>129</sup>. This puts parents under a duty not to educate their children in the context of a racist environment. This duty covers not only explicit opinions, but also implicit attitudes that children could learn. So, even though a person could choose her partner on racist grounds, once the pair have children they have a duty to control their racist (implicit and explicit) attitudes in order that their children's moral development not be distorted. Parenting rights are, in a sense, conditional. Parents must raise their children in a manner consistent with a reasonable conception of justice. Given that a child learns and imitates implicit attitudes of her parents, actual and would-be parents must behave in a way that does not reflect racism<sup>130</sup>. A person expecting a child has an important reason to avoid racist attitudes before he becomes a parent. Remember that controlling implicit attitudes is difficult, so a person who before having children displays racist attitudes will find it difficult to eradicate them once her children are born, thus making it less likely that she will use her parental rights properly.

---

<sup>128</sup> By opaqueness of the mind I mean that our minds are not transparent to us, the fact that some times we cannot know the reasons that cause our behaviour. Cf. Wilson, *Strangers to Ourselves*. I would tentatively say that I believe we have no right to act on racist preferences in these intimate spheres. But given that in many circumstances people just do not know whether they are acting on such preferences and attitudes (and given that it is difficult for external individuals to see whether someone is acting on racist grounds), people could not be penalised for choosing a partner on racist grounds.

<sup>129</sup> Cf. Chapter 1; Okin, *Justice, Gender, and the Family*. Chapters 2, 5. Matthew Clayton argues for the stronger claim that parents are bound to raise their children in accordance with public reason. Clayton, *Justice and Legitimacy in Upbringing*, 87-128

<sup>130</sup> Cf. Okin's view that families' internal life must be just in order children to develop a sense of justice. Okin, *Justice, Gender and the Family*, 97-101.

There seems to be a problem, however for private clubs here, which try to enforce racist membership policies, and justify those policies by appealing to intimacy. First, they cannot appeal to our minds' opacity because their membership policy is explicitly racist. They cannot say 'I'm sorry, I just didn't know I was acting as a racist'. Second, it is doubtful that the brackets of privacy extend to social clubs. After all, they are institutions that have a public presence: they own buildings, they are formally integrated (in contrast with the birthday party), they have explicit codes of conduct and constitutions, they make claims on public funds through tax exemptions or subsidies, and so on. These features weaken these clubs' claims to intimacy and suggest they belong to the public sphere. If this is the case, then they are more constrained by public reasons and conditions of reasonableness than marriages, friends, etc.

**Moral Development Reassessed.** This associational good is Nancy Rosenblum's main focus. I have criticised her for offering an overly narrow conception of moral development, according to which the most salient feature seems to be the ability to cooperate with members within one's voluntary associations. In other words, Rosenblum fails to accommodate sufficiently for the public ecology of associations. The argument I developed here shares the concern for the development of moral and critical capacities, but it avoids the flaws of Rosenblum's approach by focusing on citizens' moral development rather than that of members. Even though racist associates may genuinely foster capacities for cooperation and the self-esteem of members, my claim holds that moral development is conditional on being reasonable. In this respect I follow Dworkin, who, as discussed in Chapter 4, argues

that justice is a parameter of living well; a life of injustice is not a good one no matter how autonomous it is.<sup>131</sup>

I also hold that citizens have an important interest in being protected from the deleterious effects that racism has on their development. I claim that racist churches, clubs and political parties have detrimental effects on the social environment, which forms the background in which social attitudes develop<sup>132</sup>. Phenomena such as aversive racism pose an important concern for citizens who wish to develop their moral and deliberative capacities according to their own conceptions of the good. As I argued in the last chapters, an attractive strategy for protecting ourselves against these distorting environmental features is to minimise exposure to them.

**Expressive Rights Reassessed.** In order to make claims based on expressive rights, associations must present themselves as articulating a ‘voice’. Obviously, the attitude we have to this claim depends on our views about free speech. If the argument I have defended in this thesis is tenable, then freedom of speech might be legitimately regulated in order to respect and/or promote autonomy.

Now, consider the differences between the ‘voice’ of the racist Church and that of the racist political party. In the case of the latter, it could be argued that political speech deserves greater protection because it serves a vital interest of citizens<sup>133</sup>. Nevertheless, it seems that, because it matters more than other categories of speech, we must also be more concerned about its ecology. It is because political speech and political decisions are of such great importance that Rawls argues they

---

<sup>131</sup> Cf. Clayton, *Justice and Legitimacy in Upbringing*, 143.

<sup>132</sup> Cf. Lessig, L. (1995) ‘The Regulation of Social Meaning’.

<sup>133</sup> More important than the interest we have in commercial or artistic speech.

must be constrained by the requirements of public reason while non-political speech must not<sup>134</sup>. As citizens we have an important interest in the public ecology of speech; because we want to participate as free and equal members, we have a special concern regarding how political speech and deliberation are executed and also in their content. Ultimately, these concerns direct us to design policies that enhance our deliberative capacities by, for instance, neutralising mental contamination.

On the other hand, the ‘voice’ of a racist Church is not political in this respect: it does not ‘have to do with the electoral process and the activities of government’<sup>135</sup>. It does have, however, political implications insofar as it is ‘intended and received as a contribution to public deliberation about some issue’<sup>136</sup>. Nevertheless, the ‘voice’ it promotes is unreasonable, and, as such, these doctrines have detrimental effects on the ecology of people’s political beliefs through processes of mental contamination.

**Effective Pursuit of Goals Reassessed.** The argument I develop here is less permissive in the goals that associations can legitimately pursue. Associations are not bound merely by a theory of ‘clear and present danger’; their goals must also be reasonable<sup>137</sup>. Racist associations, such as racist political parties fail in appealing to this justification, because racist goals cannot be justified by public reasons.

---

<sup>134</sup> I owe this insight to Susan Hurley and Fabienne Peter.

<sup>135</sup> Scanlon, ‘Freedom of Expression and Categories of Expression’, 101.

<sup>136</sup> Cass Sunstein, *Democracy and the Problem of Free Speech* (New York: Free Press, 1995), 130.

<sup>137</sup> This condition allows for political disobedience, because it is not limited to goals that are legal. Cf. Smith, ‘Democracy, Deliberation and Disobedience’.

To conclude, freedom of association is a keystone to liberal legitimacy, but we cannot give a *carte blanche* to associations. Private associations are gatekeepers of certain goods but we must carefully assess how such groups actually protect and distribute these benefits. I have argued that it is important to balance the goods that associations offer their members with the costs that they impose on society. Nevertheless, I argue that the benefits offered by racist associations provide no reason for their protection. Limiting associational freedom on the basis of reasonableness provides, I believe, an appealing solution insofar as it distributes the goods (especially the goods of moral development and personal autonomy) of associations in a satisfactory fashion. This not only allows members to benefit, but also enhances pluralism's usefulness to liberal and democratic institutions.

## CONCLUSIONS

In this thesis I defended the claim that liberalism is compatible with certain restrictions on freedom of expression. The aim has been to rebut popular defences of free speech and association, according to which these freedoms are necessary to protect, respect, or promote individuals' autonomy.

The thesis articulates and defends two claims. One, modest in its scope, asserts that, if autonomy is bypassed by automatic, nonconscious processing, it cannot play the exemptionary role it is often allocated. The argument in defence of autonomy-based free speech claims that speech causes harm, but that this is justified because the benefits speech has on personal autonomy outweigh this. If speech bypasses autonomy, however, the justification becomes untenable. The second, stronger claim holds that, if speech is detrimental to autonomy, and we think the latter to be of value, then there is a strong reason to neutralise those effects. Moreover, given the nature of those harms, the most plausible strategy to do so is to impose some restrictions on expression.

In Chapter 1 a background of political morality was detailed against which the ensuing argument was developed. Here I argued that the most plausible interpretation of freedom of expression construes it as an independent principle; that is, as a principle which grants protection to speech beyond a general principle of minimal liberty. Under this interpretation, there is a relationship between autonomy

and speech, in virtue of which freedom of expression should be protected despite of the harmful effects it has on third parties.

I located the autonomy defence of free speech within the framework public/private distinction. I suggested this dichotomy is relevant in understanding discussions of free speech for two reasons. First, liberals tend to treat autonomy as a constraint on governmental interference. A primacy of private responsibility is invoked, according to which agents should bear the costs arising directly from their actions. In this regard, government should not infringe upon the liberties of a person who is not directly responsible for a harmful act. Second, the public/private distinction is relevant in the sense that liberals believe political principles concern only certain forms of conduct. In particular, it is usually argued that they do not apply to private actions of citizens. Against the idea of discontinuous responsibility, I noted that there is sufficient evidence suggesting features in the environment directly cause much of our behaviour, and this fact cast doubts on the plausibility of such a conception. I also argued that principles of justice must apply to institutions which have profound-effects on people's lives. Essentially, because there are many features in the environment which bypass rational control, those features, which affect importantly the opportunities of citizens, must be included within the remit of justice.

In order to defend this argument, in Chapter 2 I revised a familiar set of conditions that autonomy needs to satisfy. I linked those conditions with different challenges. In particular, I surveyed academic literature on recent developments in Social and Cognitive Psychology regarding automatic behaviour and unconscious processing. Ideomotor theories suggest that perception and action are linked in a

direct fashion which dispenses of consciousness (or 'acts of will'). The idea is that 'perception is for doing'. For instance, it has been argued that the mere perception of behaviour tends to increase the likelihood of that same behaviour. Similarly, experiments show how individuals tend to mimic and imitate others. Another related instance of ideomotor theories suggests that the perception of an action automatically triggers related concepts. In particular, I focused on the existence of social and racial stereotypes that automatically increase the tendency of people to behave consistently with such stereotypes. I highlight that these automatic tendencies are the normal state of our minds; they operate continuously. However, I do not want to suggest that tokens of automatic processes are necessary or unavoidable; only that their control is both costly and difficult. It is costly insofar as it demands a large amount of mental resources; it is difficult because, in most cases, even with introspection people ignore which tendencies are operating, their magnitude, and their direction.

From these ideas I developed the category of 'mental contamination', which is defined as 'the process by which a person ends up with an unwanted judgement, emotion or behavior because of mental processing that is unconscious or uncontrollable'<sup>1</sup>. I explored the different steps required to avoid mental contamination and argued that, in general, people tend to underestimate their susceptibility to it and to overestimate their capacity for controlling their mental processes. Instead, I suggested a strategy more likely to neutralise those reactions is, as in the case of other forms of contaminants (such as asbestos and waste), to avoid the source of contamination altogether. With the idea of mental contamination in mind, I defended an ecological conception of autonomy. Such a conception acknowledges that, in

---

<sup>1</sup> Wilson and Brekke, 'Mental Contamination and Mental Correction', 119.



many instances, social environments are important sources of mental contamination, rendering individuals less autonomous. Due to the importance autonomy places on mental states having the right structure and the right causal history, as well as agents having a disposition for effective rational reflection, autonomy requires mental contamination to be neutralised in aspects that affect vital interests. A particularly troublesome example of mental contamination is that of aversive racism, the phenomenon by which a person who endorses egalitarian, non-prejudiced views also displays negative feelings, attitudes, and reactions to people from other races which are in part the result of mental contamination<sup>2</sup>.

I follow a familiar strategy in considering problems of freedom of expression. This consists of identifying the interests which are at stake in controversies of free speech, and in cases of conflict assigning due to weight to them.

In Chapter 3 I explored Thomas Scanlon's powerful defence of free speech<sup>3</sup> in which he offers reasons for protecting speech in virtue of audiences' interests. Scanlon argues that the Millian Principle is a 'consequence of a particular idea about the limits of legitimate political authority: namely that the legitimate powers of government are limited to those that can be defended on grounds compatible with [...] the idea that each citizen is sovereign in deciding what to believe and in weighing reasons for action'. The Millian Principle rules 'that the harmful consequences to which [arguments for censorship] appeal cannot count as potential

---

<sup>2</sup> Dovidio, Kawakami, et. al., 'On the Nature of Prejudice' 512.

<sup>3</sup> Scanlon, 'A Theory of Freedom of Expression'.

justification for legal restrictions of expression'<sup>4</sup>. I explored Scanlon's arguments in support of this principle and concluded that there is no plausibility in the view that citizens are primarily interested in arriving at beliefs without governmental interference. They are concerned also with the manner in which those beliefs are reached. Because social environments in which speech is free usually have contaminating features, the Millian Principle fails to protect citizens' autonomy. Instead, I argue, citizens can be reasonably concerned with the social environment and have a motivation to neutralise mental contamination. This conclusion is defended by appealing to an original position-type situation, in which citizens attempt to protect themselves from mental contamination by avoiding exposure to certain kinds of expression, in particular those that threaten their capacity for an effective sense of justice, their capacity to form, revise and pursue a reasonable conception of the good, and their disposition towards effective rational deliberation.

I also concluded, against the notion of discontinuous responsibility, that citizens must share the costs of an autonomy-friendly public ecology, and in this respect it is not unreasonable to make speakers bear the burden of attempts to neutralise sources of mental contamination. It is acceptable, that is, to impose some limits on expression.

The conclusion of Chapter 3, then, claimed that there is a *pro tanto* reason, based on audiences' autonomy, which supports some restrictions on expression. Whether these reasons are conclusive, however, could not be adjudicated upon until the interests of speakers had been considered.

---

<sup>4</sup> Scanlon, 'Freedom of Expression and Categories of Expression', 95.

In Chapter 4, I discussed speakers' interests in free expression. I explored Dworkin's arguments, according to which expressive interests are sufficient to sustain a principle of free speech. I conditionalise the interests of speakers in free speech upon acting in a just manner. In this respect, I claimed that speakers who are prevented from expressing racist views are not wronged in a justice-sensitive sense. I argue that hate speech creates disadvantages in resources for both the victims of hate speech and aversive racists. Because those disadvantages are caused by prejudice, a liberal theory of justice cannot be neutral towards them, as it is towards disadvantages affecting the distribution of resources which result from some differences in talents, ambitions, or lifestyle. I argue that an added difficulty is that the processes by which those inequalities are created are difficult to monitor and control because, in many cases, it is the result of attitudes and processes which are (in part) implicit, automatic or unconscious. The argument concludes, then, that we have reason to neutralise the influences that cause these disadvantages.

A second argument developed in this chapter addressed an apparent conflict between Dworkin's idea of integrity and restrictions of speech. Dworkin claims that restricting the expression of certain views might compromise the integrity of those who endorse such views and who have an interest in voicing their support and living in accordance with them (the principle of authenticity is so discussed in this respect). Because Dworkin thinks that integrity is a parameter of living well, it seems that restrictions on speech threaten the capacity of some citizens to live integrated lives. But, I argue, Dworkin thinks that justice is also a parameter of living well. And, as hate speech creates disadvantages that are unjust, racists who propagate their views live defective lives. I argued that integrity is conditional upon justice, and concluded

that restrictions on racist speech make the lives of those interested in its expression no worse. Accordingly, because there cannot be trade-offs between justice and integrity, a lack of integrity on the part of racist speakers constitutes no wrong from the point of view of justice.

In Chapter 5 I defended the view that content-based restrictions on speech do not imply a commitment to perfectionism. I attempted to dismiss the following objection: because accepting content-based regulations is a violation of liberal neutrality, I could either reject neutrality and keep content regulations or maintain a commitment to the former and reject the latter. In order to rebut this objection, I argued that content regulations are defensible because of the detrimental effects speech has on autonomy. I did not claim that restrictions were justified by an assessment of the intrinsic merits of racist speech. Then, I developed an anti-perfectionist, partly comprehensive view of liberalism, which includes a concern for autonomy. The anti-perfectionist claim is defended by the view that principles of justice must satisfy a Rawlsian full publicity condition, which is a necessary component of the legitimate exercise of power.

In the last chapter I turned my attention to the issue of free association, defences of which often resemble those of speech in important aspects. I identified a set of ‘associational goods’ and illustrated how they might be employed by three examples of racist associations. I critically assessed Nancy Rosenblum’s theory of free association, the ‘moral uses of pluralism’, and concluded that she fails to accommodate sufficiently for the impact associations have in society because her account considers only short-term effects. Instead, I argued, we should expand our

considerations and focus also on the public ecology of associations. There is significant evidence to suggest that attitudes developed in the private will spillover into the public area, thus jeopardising citizens' two moral powers. In order to neutralise those effects, I suggested, associations should have to satisfy a criterion of reasonableness. I also claim that the moral development facilitated by free association is conditional upon them being reasonable.

The aim of the argument developed here is in one respect, modest. I provided arguments contending that content-based restrictions on hate speech (and association) are compatible with autonomy and justice within a liberal framework. I do not argue, however, that those restrictions are, all things considered, acceptable. There may be practical issues, such as maintaining coherence in legislation, that militate against such restrictions. I have not provided guidance on how those restrictions should be implemented, or on who should authorise and execute them. There may be other alternatives to state censorship. Perhaps, independent commissions with some form of democratic accountability would be better. Or, perhaps the costs of regulation are too high. Similarly, it may be difficult to devise legislation that regulates instances of hate speech, but which leaves sufficient free expression to promote citizens' autonomy. Or, perhaps there are other reasons for protecting speech which are independent of autonomy. Nevertheless, given the arguments developed here these shortcomings are detrimental to autonomy and not justifiable from an account of justice. This might be yet another example of the breach separating an ideal theory of justice from our non-ideal world.

The thesis is less modest in another respect. I argue that political theorists have neglected developments in other disciplines, and that such oversight has

important consequences. In order to organise institutions that promote and protect liberal values, we need a clearer understanding of how our minds function. In particular, we need to develop a deeper sense of the relationship between rationality and the environments in which we operate. Developments on Social and Cognitive Psychology (and other areas of research) can help liberals to understand better the associations between the institutions they favour and the values which motivate them. Failing to review and revise these knowledge or connections might threaten liberalism's capacity to inculcate in its citizens the motivations and the development of moral powers ultimately necessary to render itself stable over time.

## BIBLIOGRAPHY

- Alexander, L., 'Low Value Speech', *Northwestern University Law Review*, 83 (1989), 547-54.
- . *Is There a Right of Freedom of Expression?* (Cambridge: Cambridge University Press, 2005).
- Altman, A., 'Liberalism and Campus Hate Speech: A Philosophical Examination', *Ethics*, 103 (1990), 302-17.
- Amdur, R., 'Scanlon on Freedom of Expression', *Philosophy and Public Affairs*, 9 (1980), 287-300.
- Arneson, R., 'Autonomy and Preference Formation', in Coleman, J., and A. Buchanan, (eds.) *In Harm's Way: Essays in Honour of Joel Feinberg* (New York: Cambridge University Press, 1994), 42-75.
- . 'Perfectionism and Politics', *Ethics*, 111 (2000), 37-63.
- . 'Cracked Foundations of Liberal Equality', in Burley, J. (ed.) *Dworkin and his Critics* (Oxford: Blackwell, 2004), 79-98.
- Arthur, J., 'Recent Work in Freedom of Speech', *Philosophical Books*, xxxviii (1997), 225-34.
- Austin, J., *How to Do Things with Words?* (Oxford: Clarendon Press, 1962).
- Bargh, J., 'The Automaticity of Everyday Life', *Advances in Social Cognition*, 10 (1997), 1-61.
- . 'The Cognitive Monster: The Case against the Controllability of Automatic Stereotype Effects', in Chaiken, S., and Y. Trope (eds.) *Dual-Process Theories in Social Psychology* (New York: Guilford Press, 1999), 361-82.
- . 'Bypassing the Will: Toward Demystifying the Nonconscious Control of Social Behaviour', in Hassin, R., J. Uleman and J. Bargh (eds.) *The New Unconscious*. (New York: Oxford University Press, 2005), 37-58.
- . 'What Have We Been Priming all these Years? On the Development, Mechanisms, and Ecology of Nonconscious Social Behavior', *European Journal of Social Psychology*, 36 (2006), 147-68.

- Bargh, J., and T. Chartrand 'The Unbearable Automaticity of Being', *American Psychologist*, 54. (1999), 462-79.
- Bargh, J., M. Chen, and L. Burrows 'Automaticity of Social Behavior: Direct Effects on Trait Construct and Stereotype Activation on Action', *Journal of Personality and Social Psychology*, 71 (1996), 230-44.
- Bargh, J. and M. Ferguson 'Beyond Behaviorism: On the Automaticity of Higher Mental Processes', *Psychological Bulletin*, 126 (2000), 925-45.
- Bargh, J., P. Gollwitzer, et. al., 'The Automated Will: Nonconscious Activation and Pursuit of Behavioral Goals', *Journal of Personality and Social Psychology*, 81(2001), 1014-27.
- Bargh, J., and E. Williams 'The Automaticity of Social Life', *Current Directions in Psychological Science*, 15 (2006), 1-4.
- Bayles, M., 'Review of Joel Feinberg, *Harm to Others*', *Law and Philosophy*, 4 (1985), 423-32.
- Benson, P., 'Autonomy and Oppressive Socialization', *Social Theory and Practice*, 17 (1991), 385-408.
- . 'Feminist Intuitions and the Normative Substance of Autonomy', in Taylor, J. (ed.) *Personal Autonomy: New Essays on Personal Autonomy and Its Roles in Contemporary Moral Philosophy* (Cambridge: Cambridge University Press, 2005), 124-42.
- Berlin, I., *Liberty* (New York: Oxford University Press, 2002).
- Blackmore, S., *The Meme Machine* (Oxford: Oxford University Press, 1999).
- Boettcher, J. 'What is Reasonableness', *Philosophy and Social Criticism*, 30 (2004), 597-621.
- Bou-Habib, P., 'A Theory of Religious Accommodation', *Journal of Applied Philosophy*, 23 (2006), 109-26.
- Boy Scouts of America et. al. v Dale*(99-699) 530 U.S. 640 (2000).
- Brandenburg v. Ohio*, 395 U.S. 444 (1969).
- Brink, D., 'Millian Principles, Freedom of Expression and Hate Speech', *Legal Theory*, 7 (2001), 119-57.
- Brison, S., 'The Autonomy Defense of Free Speech', *Ethics*, 108 (1998), 312-39.
- . 'Speech, Harm, and the Mind-Body Problem in First Amendment Jurisprudence', *Legal Theory*, 4 (1998), 39-61.



- . 'Relational Autonomy and Freedom of Expression', in Mackenzie, C., and N. Stoljar (eds.) *Relational Autonomy* (New York: Oxford University Press, 2000), 280-99.
- . 'Review of Alan Haworth *Free Speech*', *Mind*, 113 (2004), 351-7.
- Buss, S., 'Personal Autonomy', in Zalta E. (ed.) *The Stanford Encyclopedia of Philosophy* (2002), <http://www.plato.stanford.edu/archives/win2002/entries/personal-autonomy/> (accessed 5<sup>th</sup> January 2004).
- Carruthers, P., 'Simple Heuristics Meet Massive Modularity', in Carruthers, P., S. Laurence, and S. Stich (eds.) *The Innate Mind II: Culture and Cognition* (Oxford: Oxford University Press, 2006), 181-98.
- Carter, S., *A Measure of Freedom* (Oxford: Oxford University Press, 1999).
- Chan, J., 'Legitimacy, Unanimity, and Perfectionism', *Philosophy and Public Affairs*, 29 (2000), 5-42.
- Chartrand, T., and J. Bargh, 'Automatic Activation of Impression Formation and Memorization Goals: Nonconscious Goal Priming Reproduces Effects of Explicit Task Instructions', *Journal of Personality and Social Psychology*, 71 (1996), 464-78.
- . 'The Chameleon Effect: The Perception-Behavior Link and Social Interaction', *Journal of Personality and Social Psychology*, 76 (1999), 893-910.
- Chase, V., R. Hertwig, and G. Gigerenzer, 'Visions of Rationality', *Trends in Cognitive Sciences*, 2 (1998), 206-14.
- Chater, N., and M. Oaksford, 'Human Rationality and the Psychology of Reasoning: Where Do We Go from Here?', *British Journal of Psychology*, 92 (2001), 193-216.
- Chen, M., and J. Bargh, 'Nonconscious Behavioral Confirmation Processes: The Self-Fulfilling Consequences of Automatic Stereotype Activation', *Journal of Experimental Social Psychology*, 33 (1997), 541-60.
- Christiano, T., 'The Authority of Democracy', *The Journal of Political Philosophy*, 12 (2004), 266-90.
- Christman, J., 'Autonomy and Personal History', *Canadian Journal of Philosophy*, 21 (1991), 1-24.
- . 'Autonomy in Moral and Political Philosophy', in Zalta, E. (ed) *The Stanford Encyclopedia of Philosophy*, (2003) <http://www.plato.stanford.edu/archives/fall2003/entries/autonomy-moral/> (accessed 1<sup>st</sup> December 2004).
- . 'Relational Autonomy, Liberal Individualism, and the Social Constitution of Selves',

- Philosophical Studies*, 117 (2004), 143-64.
- . 'Autonomy, Self-Knowledge, and Liberal Legitimacy', in Christman, J., and J. Anderson (eds.) *Autonomy and the Challenges to Liberalism* (Cambridge: Cambridge University Press, 2005), 330-58.
- . 'Procedural Autonomy and Liberal Legitimacy', in Taylor, J. (ed.) *Personal Autonomy: New Essays on Personal Autonomy and Its Roles in Contemporary Moral Philosophy*. (Cambridge: Cambridge University Press, 2005), 277-98.
- Clayton, M., 'White on Autonomy, Neutrality and Well-Being', *Journal of Philosophy of Education*, 27 (1993), 101-12.
- . 'Liberal Equality and Ethics', *Ethics* (2002), 113, 8-22.
- . 'A Puzzle about Justice, Ethics and the Sacred'. In Burley, J. (ed.) *Dworkin and his Critics* (Oxford: Blackwell, 2004), 99-110.
- . *Justice and Legitimacy in Upbringing* (Oxford: Oxford University Press, 2006).
- Cohen, G. A., 'On the Currency of Egalitarian Justice', *Ethics*, 99 (1989), 906-44.
- . 'Incentives, Inequality, and Community', in Darwall, S. (ed.) *Equal Freedom* (Ann Harbor: Michigan University Press, 1995), 331-97.
- . 'The Pareto Argument for Inequality', *Social Philosophy and Policy*, 12 (1995), 160-85.
- . 'Where the Action Is: On the Site of Distributive Justice', *Philosophy and Public Affairs*, 26 (1997), 3-30.
- . *If You're an Egalitarian, How Come You're so Rich?* (Cambridge, MA: Harvard University Press, 2000).
- . 'Expensive Tastes Ride Again', in Burley, J. (ed) *Dworkin and His Critics*, (Oxford: Blackwell, 2004), 3-29.
- Cohen, J., 'Okin on Justice, Gender, and Family', *Canadian Journal of Philosophy*, 22 (1992), 263-86.
- . 'Moral Pluralism and Political Consensus', in Copp, D., J. Hampton, and J. Roemer (eds.) *The Idea of Democracy* (Cambridge: Cambridge University Press, 1993), 270-91.
- . 'Freedom of Expression', *Philosophy and Public Affairs*, 22 (1993), 207-63.
- Cohen-Eliya, M., and Y. Hammer, 'Advertisement, Stereotypes and Freedom of Expression'. *Journal of Social Philosophy*, 35 (2004), 165-87.

- Cosmides, L., and J. Tooby, 'Cognitive Adaptations for Social Exchanges', in Bakow, J., L. Cosmides and J. Tooby (eds.) *The Adapted Mind* (New York: Oxford University Press, 1992), 163-228.
- Dahl, R., *Dilemmas of Pluralist Democracy: Autonomy vs. Control* (New Haven: Yale University Press, 1982).
- Davidson, D., *Essays on Actions and Events* (Oxford: Oxford University Press, 2001).
- Dawkins, R., *The Selfish Gene* (Oxford: Oxford University Press, 1989).
- Dennett, D., *Darwin's Dangerous Idea: Evolution and the Meanings of Life* (London: Penguin, 1995).
- . 'The Evolution of Culture', *The Monist*, 84 (2001), 305-24.
- Devine, P., 'Stereotypes and Prejudice: Their Automatic and Controlled Components' *Journal of Personality and Social Psychology*, 56 (1989), 5-18.
- Dijksterhuis, A. and J. Bargh, 'The Perception-Behavior Expressway: Automatic Effects of Social Perception on Social Behavior', *Advances in Experimental Social Psychology*, 33 (2001), 1-40.
- Dovidio, J., N. Evans, and R. Tyler 'Racial Stereotypes: The Contents of Their Cognitive Representations', *Journal of Experimental Social Psychology*, 22 (1986), 22-37.
- Dovidio, J., and S. Gaertner 'Aversive Racism'. *Advances in Experimental Social Psychology*, 36 (2004), 1-52.
- Dovidio, J., K. Kawakami, et. al., 'On the Nature of Prejudice: Automatic and Controlled Processes', *Journal of Experimental Social Psychology*, 33 (1997), 510-40.
- Dreben, B., 'On Rawls and Political Liberalism', in Freeman, S. (ed.) *The Cambridge Companion to Rawls* (Cambridge: Cambridge University Press, 2003), 316-46.
- Dworkin, G. *The Theory and Practice of Autonomy* (Cambridge: Cambridge University Press, 1988).
- Dworkin, R. *Taking Rights Seriously* (London: Ducksworth, 1977).
- . 'Liberalism', in Hampshire, S. (ed.) *Public and Private Morality* (Cambridge: Cambridge University Press, 1978), 113-43.
- . 'Do We Have a Right to Pornography?', in *A Matter of Principle* (Cambridge, MA: Harvard University Press, 1985), 335-72.

- . 'Foundations of Liberal Equality', in Darwall, S. (ed.) *Equal Freedom: Selected Tanner Lectures on Human Values* (Ann Arbor: University of Michigan Press, 1995), 190-306.
- . *Freedom's Law* (Cambridge, MA: Harvard University Press, 1996).
- . *Sovereign Virtue: The Theory and Practice of Equality* (Cambridge, MA: Harvard University Press, 2000).
- . 'Sovereign Virtue Revisited', *Ethics*, 113 (2002), 106-43.
- . 'Ronald Dworkin Replies', in Burley, J. (ed.) *Dworkin and his Critics* (Oxford: Blackwell, 2004), 339-95.
- . *Is Democracy Possible Here?* (Princeton: Princeton University Press, 2006).
- Pincus, F., and H. Ehrlich, (eds.) *Race and Ethnoviolence* (Boulder: Westview Press, 1994).
- Elster, J. *Sour Grapes: Studies in the Subversion of Rationality* (Cambridge: Cambridge University Press, 1983).
- . *Ulysses and the Sirens* (Cambridge: Cambridge University Press, 1984).
- . 'The Nature and Scope of Rational Choice Explanation' in LePore, E., and B. McLaughlin (eds.) *Actions and Events: Perspectives on the Philosophy of Donald Davidson* (Oxford: Blackwell, 1985), 60-72.
- . *Rational Choice Theory* (Oxford: Blackwell, 1986).
- . 'Some Unresolved Problems in the Theory of Rational Behaviour', *Acta Sociologica*, 36 (1993), 179-90.
- Evans, R., *Lying About Hitler: History, the Holocaust and the David Irving Trial* (New York: Basic Books, 2001).
- FCC v. Pacifica Foundation*, 438 U.S. 726 (1978)
- Feinberg, J., *The Moral Limits of the Criminal Law: Harm to Others* (Oxford: Oxford University Press, 1984).
- . *The Moral Limits of the Criminal Law: Offense to Others* (Oxford: Oxford University Press, 1985).
- . *The Moral Limits of the Criminal Law: Harm to Self* (Oxford: Oxford University Press, 1986).

- Ferguson, M., J. Bargh, and D. Nayak, 'After-Affects: How Automatic Evaluations Influence the Interpretation of Subsequent, Unrelated Stimuli', *Journal of Experimental Social Psychology*, 41 (2005), 182-91.
- Fiss, O., *The Irony of Free Speech* (Cambridge, MA: Harvard University Press, 1996).
- . *Liberalism Divided* (Boulder: Westview Press, 1996)
- Fleischacker, S., 'Insignificant Communities', in. Gutmann, A. (ed.) *Freedom of Association* (Princeton: Princeton University Press, 1998) 273-313.
- Frankfurt, H., 'Alternate Possibilities and Moral Responsibility', *The Journal of Philosophy*, 66 (1969), 829-39.
- . 'Freedom of the Will and the Concept of a Person', *The Journal of Philosophy*, 68 (1971), 5-20.
- Gaertner, S. and J. Dovidio, 'The Subtlety of White Racism, Arousal, and Helping Behavior', *Journal of Personality and Social Psychology*, 35 (1977), 691-707.
- Galston, W., *Liberal Pluralism: The Implications of Value Pluralism for Political Theory and Practice* (Cambridge: Cambridge University Press, 2002)
- Gaus, G., 'The Place of Autonomy within Liberalism', in. Christman, J. and J. Anderson, (eds.) *Autonomy and the Challenges to Liberalism* (Cambridge: Cambridge University Press, 2005), 272-306.
- Gigerenzer, G. *Adaptive Thinking: Rationality in the Real World* (New York: Oxford University Press, 2000).
- . 'The Adaptive Toolbox', in Gigerenzer, G., and R. Selten, (eds.) *Bounded Rationality: The Adaptive Toolbox* (Cambridge, MA: MIT Press, 2001), 37-50.
- . *Reckoning with Risk* (London: Penguin Books, 2002).
- Gigerenzer, G., and K. Hug, 'Domain-Specific Reasoning: Social Contracts, Cheating, and Perspective Change', *Cognition*, 43 (1992), 127-71.
- Gigerenzer, G., and P. Todd, *Simple Heuristics that Make Us Smart* (New York: Oxford University Press, 1999).
- Gilbert, D. 'How Mental Systems Believe', *American Psychologist*, 46 (1991), 107-19.
- . 'The Assessment of Man: Representation and the Control of Belief', in Wegner, D., and J. Pennebaker (eds.) *Handbook of Mental Control* (Englewood Cliffs, New Jersey: Prentice Hall, 1993), 57-87.

- Gilbert, D., R. Tatarodi and P. Malone, 'You Can't Not Believe Everything You Read', *Journal of Personality and Social Psychology*, 65 (1993), 221-33.
- Gilovich, T., D. Griffin, and D. Kahneman, (eds.) *Heuristics and Biases: The Psychology of Intuitive Judgment* (Cambridge: Cambridge University Press, 2002).
- Goldstein, D., and G. Gigerenzer, 'Models of Ecological Rationality: The Recognition Heuristic', *Psychological Review*, 109 (2002), 75-90.
- Goodin, R., *No Smoking: The Ethical Issues* (Chicago: University of Chicago Press, 1989).
- Greenawalt, K., *Speech, Crime, and the Uses of Language* (Oxford: Oxford University Press, 1989).
- . *Private Consciences and Public Reasons* (New York: Oxford University Press, 1995).
- . 'Freedom of Association and Religious Association', in Gutmann, A. (ed.) *Freedom of Association* (Princeton: Princeton University Press, 1998), 109-44.
- Greenwald, A., and M. Banaji 'Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes', *Psychological Review*, 102 (1995), 4-27.
- Greenwald, A., B. Nosek, and M. Banaji, 'Understanding and Using the Implicit Association Test: I. An Improved Scoring Algorithm', *Journal of Personality and Social Psychology*, 85 (2003), 197-217.
- Gutmann, A., 'Freedom of Association', in Gutmann, A., (ed.) *Freedom of Association* (Princeton: Princeton University Press, 1998), 3-32.
- Hacker, A., *Two Nations: Black and White, Separate, Hostile, Unequal* (New York: Ballantine Books, 1995).
- Hargreaves-Heap, S., M. Hollis, et. al., *The Theory of Choice: A Critical Guide* (Oxford: Blackwell, 1992).
- Haworth, A., 'On Mill, Infallibility, and Freedom of Expression', *Res Publica*, 13 (2007), 77-100.
- Hirschfeld, L., *Race in the Making: Cognition, Culture, and the Child's Construction of Human Kinds* (Cambridge, MA: MIT Press, 1996).
- Hornsby, J., and R. Langton, 'Free Speech and Illocution', *Legal Theory*, 4 (1998), 21-37.
- Hurka, T., 'Indirect Perfectionism: Kymlicka on Neutrality', *The Journal of Political Philosophy*, 3 (1995), 36-57.

- Hurley, S., *Natural Reasons: Personality and Polity* (Oxford: Oxford University Press, 1989).
- . 'Cognitivism in Political Philosophy', in Crisp, R., and B. Hooker (eds.) *Well-Being and Morality: Essays in Honour of James Griffin* (Oxford: Clarendon, 2000), 77-208.
- . *Justice, Luck, Knowledge* (Cambridge, MA: Harvard University Press, 2003).
- . 'Social Heuristics that Make Us Smarter', *Philosophical Psychology*, 18 (2005), 585-612.
- . 'Bypassing Conscious Control: Unconscious Imitation, Media Violence, and Freedom of Speech', in Pockett, S., W. Banks, and S. Gallagher, (eds.) *Does Consciousness Cause Behavior* (Cambridge, MA: MIT Press, 2006), 301-37.
- . 'The Public Ecology of Responsibility', (Unpublished Manuscript, 2006).
- Jacobson, D. 'Freedom of Speech Acts? A Response to Langton', *Philosophy and Public Affairs*, 24, (1995) 64-79.
- . 'Speech and Action: Replies to Hornsby and Langton', *Legal Theory*, 7, (2001) 179-201.
- Jones, P., 'Equality, Recognition, Difference', *Critical Review of International Social and Political Philosophy*, 9, (2006) 23-46.
- . 'Toleration, Recognition and Identity', *The Journal of Political Philosophy*, 14, (2006) 123-43.
- Kagan, S., *The Limits of Morality* (Oxford: Oxford University Press, 1989).
- Kahneman, D., and A. Tversky, 'Judgment under Uncertainty: Heuristics and Biases' in Kahneman, D., P. Slovic, and A. Tversky, (eds.) *Judgment under Uncertainty: Heuristics and Biases* (Cambridge: Cambridge University Press, 1982).
- . (eds.) *Choices, Values and Frames* (Cambridge: Cambridge University Press, 2000).
- Kateb, G. 'The Value of Association', in Gutmann, Amy (ed) *Freedom of Association* (Princeton: Princeton University Press, 1998), 35-63.
- Katz, I. and G. Hass, 'Racial Ambivalence and American Conflict Value: Correlational and Priming Studies of Dual Cognitive Structures' *Journal of Personality and Social Psychology*, 55, (1988) 893-905.
- Kawakami, K., J. Dovidio, and A. Dijksterhuis, 'Effect of Social Category Priming on Personal Attitudes', *Psychological Science*, 14, (2003) 315-9.
- Kawakami, K., J. Dovidio, et. al., 'Just Say NO (to Stereotyping): Effects of Training in the Negation of Stereotypic Associations on Stereotype Activation', *Journal of Personality and*

- Social Psychology*, 78, (2000) 871-88.
- Kelly, E., and L. McPherson, 'On Tolerating the Unreasonable', *The Journal of Political Philosophy*, 9, (2001) 38-55.
- Kukathas, C., *The Liberal Archipelago: A Theory of Diversity and Freedom* (Oxford: Oxford University Press, 2003).
- Kymlicka, W. *Liberalism, Community and Culture*, (Oxford: Clarendon Press, 1989).
- . 'Liberal Individualism and Liberal Neutrality' *Ethics*, 99, (1989) 883-905.
- . *Contemporary Political Philosophy: An Introduction* (Oxford: Clarendon Press, 1990).
- Langton, R., 'Whose Rights? Ronald Dworkin, Women, and Pornographers', *Philosophy and Public Affairs*, 19, (1990) 311-59.
- . 'Speech Acts and Unspeakable Acts', *Philosophy and Public Affairs*, 22, (1993) 293-330.
- Langton, R., and C. West, 'Scorekeeping in a Pornographic Language Game', *Australasian Journal of Philosophy*, 77, (1999) 303-19.
- Larmore, C. 'Political Liberalism', *Political Theory*, 18, (1990) 339-60.
- . 'Public Reason', in Freeman, S. (ed.) *The Cambridge Companion to Rawls* (Cambridge: Cambridge University Press, 2003), 368-93.
- Lessig, L., 'The Regulation of Social Meaning', *The University of Chicago Law Review*, 62 (1995), 943-1045.
- Levine, M., and T. Pataki, (eds.) *Racism in Mind* (Ithaca: Cornell University Press, 2004).
- Lipstadt, D., *Denying the Holocaust: The Growing Assault on Truth and Memory* (London: Penguin, 1994).
- Lord, C., L. Ross, and M. Lepper, 'Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence', *Journal of Personality and Social Psychology*, 37, (1979) 2098-109.
- MacKinnon, C., *Only Words* (Cambridge, MA: Harvard University Press, 1993).
- Mackenzie, C., and N. Stoljar, 'Introduction: Autonomy Reconfigured', in Mackenzie, C., and N. Stoljar (eds.) *Relational Autonomy* (New York: Oxford University Press, 2000), 3-31.
- de Marneffe, P., 'Liberalism, Liberty, and Neutrality' *Philosophy and Public Affairs*, 19, (1990) 253-74.



- . 'Rights, Reasons and Freedom of Association', in Gutmann, A. (ed.) *Freedom of Association* (Princeton: Princeton University Press, 1998), 145-73.
- Mason, A., *Levelling the Playing Field: The Idea of Equal Opportunity and its Place in Egalitarian Thought* (Oxford: Oxford University Press, 2006).
- Matsuda, M., C. Lawrence III, et. al., (eds.) *Words that Wound: Critical Race Theory, Assaultive Speech and the First Amendment* (Boulder: Westview Press, 1993).
- McGowan, M., 'Conversational Exercitives and the Force of Pornography', *Philosophy and Public Affairs*, 31, (2003) 155-89.
- McKinnon, C., *Toleration: A Critical Introduction* (London: Routledge, 2006).
- . 'Should We Tolerate Holocaust Denial', *Res Publica*, 13 (2007), 9-28.
- Meckled-Garcia, S., 'Toleration and Neutrality: Incompatible Ideals?', *Res Publica*, 7 (2001), 293-313.
- Mill, J. S., *On Liberty and Other Essays* (Oxford: Oxford University Press, 1991).
- Miller v. California*, 413 U.S. 15 (1972).
- Miller v. California*, 413 U.S. 15 (1973).
- Moles, A., 'Autonomy, Free Speech and Automatic Behaviour', *Res Publica*, 13 (2007), 53-75.
- Nagel, T. 'Rawls on Justice', *The Philosophical Review*, 82 (1973), 220-34.
- . *Equality and Partiality* (Oxford: Oxford University Press, 1991).
- . 'Personal Rights and Public Sphere' in *Concealment and Exposure* (New York: Oxford University Press, 2002), 31-52.
- . 'Rawls and Liberalism', in Freeman, S., (ed) *The Cambridge Companion to Rawls* (Cambridge: Cambridge University Press, 2003), 62-85.
- Nolan, D., and C. West, 'Liberalism and Mental Mediation', *Journal of Value Inquiry*, 38, (2004) 186-202.
- Nozick, R., *Anarchy, State, and Utopia* (Oxford: Blackwell, 1974).
- Nussbaum, M., *Sex and Social Justice* (Oxford: Oxford University Press, 1999).
- N.Y. City of Ancient Order of Hibernian v. Dinkins*, 814 F. Supp. 358 (1993).

- Okin, S., *Justice, Gender, and the Family* (New York: Basic Books, 1989).
- . 'Gender, the Public and the Private', in Held, D., (ed.) *Political Theory Today* (Cambridge: Polity Press, 1991), 67-90.
- Oshana, M., 'How Much Should We Value Autonomy?', *Social Philosophy and Policy*, 20 (2003), 99-126.
- Parents Involved in Community Schools v. Seattle School District No. 1 et. al.*, 551 U.S. \_\_\_\_ (2007)
- Pauer-Studer, H., 'Liberalism, Perfectionism, and Civic Virtue', *Philosophical Explorations*, 3 (2001), 174-92.
- Pettit, P., *A Theory of Freedom* (Cambridge: Polity Press, 2001).
- Pratto, F., and J. Bargh, 'Stereotyping Based on Apparently Individuating Information: Trait and Global Components of Sex Stereotypes Under Attention Overload', *Journal of Experimental Social Psychology*, 27 (1991), 26-47.
- Qizilbash, M., 'Well-Being, Adaptation and Human Limitations', *Royal Institute of Philosophy Supplements*, 81 (2006), 83-110.
- . 'Capability, Happiness and Adaptation in Sen and J. S. Mill', *Utilitas*, 18 (2006), 20-32.
- Quong, J., 'Disputed Practices and Reasonable Pluralism', *Res Publica*, 10 (2004), 43-67.
- . 'The Scope of Public Reason', *Political Studies*, 52 (2004), 233-50.
- . 'The Rights of Unreasonable Citizens', *The Journal of Political Philosophy*, 12 (2004), 314-35.
- Rattansi, A., *Racism: A very short introduction* (Oxford: Oxford University Press, 2006).
- R.A.V. v. City of St. Paul, Minnesota*, 505 U.S. 377 (1992).
- Rawls, J., *Political Liberalism* (New York: Columbia University Press, 1996).
- . *A Theory of Justice* (Cambridge, MA: Harvard University Press, 1999).
- . *Collected Papers* (Cambridge, MA: Harvard University Press, 1999).
- . *The Law of People with "The Idea of Public Reason Revisited"* (Cambridge, MA: Harvard University Press, 1999).
- . *Justice as Fairness: A Restatement* (Cambridge, MA: Belknap Press, 2001).

- Raz, J., *The Morality of Freedom* (Oxford: Clarendon Press, 1986).
- . 'Facing Up: A Reply', *Southern California Law Review*, 62 (1989), 1153-235.
- . *Practical Reasons and Norms* (Oxford: Oxford University Press, 1990).
- . 'Facing Diversity: The Case of Epistemic Abstinence', *Philosophy and Public Affairs*, 19 (1990), 3-46.
- . 'Free Expression and Personal Identification', in his *Ethics in the Public Domain* (Oxford: Clarendon Press, 1994), 146-69.
- Reidy, D., 'Rawls's Wide View of Public Reason: Not Wide Enough', *Res Publica*, 6, (2000) 49-72.
- Roberts v. Jaycees*, 468 U.S. 409, (1984).
- Rosenblum, N., 'Compelled Association: Public Standing, Self-Respect and the Dynamics of Exclusion', in Gutmann, A. (ed.) *Freedom of Association* (Princeton: Princeton University Press, 1998), 75-108.
- . *Membership and Morals: The Personal Uses of Pluralism in America* (Princeton: Princeton University Press, 1998).
- Rousseau, J-J., *The Social Contract and Discourses* (London: Everyman, 1993).
- Runyon v. McCrary*, 427 U.S. 160 (1976).
- Russell, P., 'Critical Notice of Fischer, J. M. and M. Ravizza *Responsibility and Control*', *Canadian Journal of Philosophy*, 32 (2002), 587-606.
- Ryle, G., *The Concept of Mind* (London: Penguin, 1963).
- Sandel, M., *Liberalism and the Limits of Justice* (Cambridge: Cambridge University Press, 1998).
- Scanlon, Thomas (1998) *What We Owe to Each Other*. Cambridge: MA: Harvard University Press.
- . *The Difficulty of Tolerance: Essays in Political Philosophy* (Cambridge, MA: Harvard University Press, 2003).
- . 'Rawls on Justification', in Freeman, S., (ed.) *The Cambridge Companion to Rawls* (Cambridge: Cambridge University Press, 2003), 139-67.
- Schauer, F., *Free Speech: A Philosophical Enquiry* (Cambridge: Cambridge University Press, 1982).

- Scheffler, S., 'Is the Basic Structure Basic?', in Sypnowich, C. (ed.) *The Egalitarian Conscience* (Oxford: Oxford University Press, 2006), 102-29.
- Searle, J., *Speech Acts: An Essay in the Philosophy of Language* (Cambridge: Cambridge University Press, 1969).
- . *Rationality in Action* (Cambridge, MA: MIT Press, 2001).
- Selten, R., 'What is Bounded Rationality', in Gigerenzer, G., and R. Selten (eds.) *Bounded Rationality: The Adaptive Toolbox* (Cambridge, MA: MIT Press, 2001), 13-36.
- Sen, A., 'Equality of What?', in Darwall, S., (ed.) *Equal Freedom* (Ann Harbor: Michigan University Press, 1985), 307-30.
- . *On Ethics and Economics* (Oxford: Basil Blackwell, 1987).
- Sher, G., *Beyond Neutrality: Perfectionism in Politics* (Cambridge: Cambridge University Press, 1997).
- Shermer, M., and A. Grobman., *Denying History: Who Says the Holocaust Never Happened and Why They Say It?* (Berkeley: University of California Press, 2000).
- Smith, W., 'Democracy, Deliberation and Disobedience', *Res Publica*, 10 (2004), 353-77.
- Sperber, D., *Explaining Culture: A Naturalistic Approach* (Oxford: Blackwell, 1996).
- . 'Modularity and Relevance: How Can a Massively Modular Mind be Flexible and Context-Sensitive?', in Carruthers, P., S. Laurence, and S. Stich (eds.) *The Innate Mind I: Structure and Contents* (Oxford: Oxford University Press, 2005), 53-68.
- Sperber, D., and L. Hirschfeld, 'Culture and Modularity', in Carruthers, P., S. Laurence, and S. Stich (eds.) *The Innate Mind II: Culture and Cognition* (Oxford: Oxford University Press, 2006), 149-64
- St. Paul Bias-Motivated Crime Ordinance, St. Paul, Minn., Legis. Code 292.02 (1990)
- Sterelny, K., *Thought in a Hostile World: The Evolution of Human Cognition* (Oxford: Blackwell, 2003).
- Stoljar, N., 'Autonomy and the Feminist Intuition', in Mackenzie, C., and N. Stoljar (eds.) *Relational Autonomy* (Oxford: Oxford University Press, 2000), 94-111.
- Strauss, D., 'Persuasion, Autonomy and Freedom of Expression', *Columbia Law Review*, 91 (1991), 334-71.
- Sunstein, C., 'Preference and Politics', *Philosophy and Public Affairs*, 20 (1991), 3-34.

- . *Democracy and the Problem of Free Speech* (New York: Free Press, 1995).
- Tamir, Y., 'Revisiting the Public Sphere', in Gutmann, A. (ed.) *Freedom of Association* (Princeton: Princeton University Press, 1998) 214-38.
- Taylor, C., *Philosophical Papers II: Philosophy and the Human Sciences* (Cambridge: Cambridge University Press, 1985).
- Tolstoy, L., *Anna Karenina* (Oxford: Oxford University Press, 1995).
- Tony & Susan Alamo Foundation v. Secretary of Labor*, 471 U.S. 290, (1985).
- Tooby, J., and L. Cosmides 'The Psychological Foundations of Culture', in Bakow, J., L. Cosmides, and J. Tooby (eds.) *The Adapted Mind* (New York: Oxford University Press, 1992), 19-136.
- Uleman, J., 'Introduction: Becoming Aware of the New Unconscious', in Hassin, R., J. Uleman, and J. Bargh (ed.) *The New Unconscious* (New York: Oxford University Press, 2005), 3-15.
- Wall, S., 'Perfectionism, Public Reasons, and Religious Accommodation', *Social Theory and Practice*, 31 (2005), 281-304.
- Walzer, M., 'On Involuntary Association', in Gutmann, A. (ed.) *Freedom of Association* (Princeton: Princeton University Press, 1998) 64-74.
- Watson, G., 'Free Agency', *Journal of Philosophy*, 72 (1975), 205-20.
- Wegner, D., 'You Can't Always Think What You Want: Problems in the Suppression of Unwanted Thoughts', *Advances in Experimental Social Psychology*, 25 (1992), 193-225.
- . 'Ironic Processes of Mental Control', *Psychological Review*, 101 (1994), 34-52.
- Wegner, D., G. Coulton, and R. Wenzlaff 'The Transparency of Denial: Briefing in the Debriefing Paradigm', *Journal of Personality and Social Psychology*, 49 (1985), 338-46.
- Welsh v. Boy Scouts*, 993 F.2d 1267 (1993).
- Wheeler, S. C., W. Jarvis, and R. Petty, 'Think Unto Others: The Self-Destructive Impact of Negative Racial Stereotypes', *Journal of Experimental Social Psychology*, 37 (2001), 173-80.
- White, S., 'Freedom of Association and the Right to Exclude', *Journal of Political Philosophy*, 5 (1997), 373-91.
- . 'Trade Unionism in a Liberal State', in Gutmann, A. (ed.) *Freedom of Association* (Princeton: Princeton University Press, 1998), 330-56.

- Whitney v. California*, 274 U.S. 357 (1927).
- Williams, A., 'Incentives, Inequality and Publicity' *Philosophy and Public Affairs*, 27 (1998), 225-47.
- . 'The Alleged Incompleteness of Public Reason', *Res Publica*, 6 (2000), 199-211.
- . 'Liberty, Equality, and Property', in Dryzek, J., B. Honig, and A. Philips (eds.) *The Oxford Handbook of Political Theory* (Oxford: Oxford University Press, 2006), 488-506.
- Wilson, T., *Strangers to Ourselves: Discovering the Adaptive Unconscious* (New York: Belknap, 2002).
- Wilson, T., and N. Brekke, 'Mental Contamination and Mental Correction: Unwanted Influences on Judgments and Evaluations', *Psychological Bulletin*, 116, (1994) 117-42.
- Wolf, S., 'Sanity and the Metaphysics of Responsibility', in Schoeman, F. (ed.) *Responsibility, Character, and the Emotions* (Cambridge: Cambridge University Press, 1987), 46-62.
- . *Freedom Within Reason* (Oxford: Oxford University Press, 1990).
- Young, I., *Justice and the Politics of Difference* (Princeton: Princeton University Press, 1990).
- . 'Rawls's *Political Liberalism*', *The Journal of Political Philosophy*, 3 (1995), 181-90.
- . 'Taking the Basic Structure Seriously', *Perspectives on Politics*, 4 (2006), 91-7.
- Zack, N., 'Review of Nancy L. Rosenblum, *Membership and Morals: The Personal Uses of Pluralism in America*', *The Journal of Value Enquiry*, 34 (2000), 111-15.
- . 'Race and Racial Discrimination', in LaFollette, H., (ed.) *The Oxford Handbook of Practical Ethics* (Oxford: Oxford University Press, 2003), 245-71.