Why Be Reasonable?
Political Liberalism, Moral Pluralism,
and Deep Disagreement

Julia Christiane Gabriela Netter

Christ Church

77644 words

Why Be Reasonable?
Political Liberalism, Moral Pluralism,
and Deep Disagreement

Julia Christiane Gabriela Netter

Christ Church

Abstract

Political liberalism contains a commitment to public justification. The exercise of coercion on the basis of political principles is only thought to be proper if these principles are acceptable to each and every reasonable person. The fact that political liberalism restricts the constituency of public justification, *i.e.,* the constituency of those who are owed justifications, to reasonable people is significant. I argue that, as it stands, this restriction is problematic. Specifically, political liberalism's core commitment to respect for persons as ends in themselves is in conflict with its refusal to justify their exclusion to some individuals who will be coerced. Furthermore, attempts to dispense with the need to provide justifications to the unreasonable seem to resolve that tension, but only at the cost of introducing a second defect: an impoverished and ultimately illiberal conception of the person which refuses to regard individual persons as morally autonomous.

I conclude that political liberalism must justify the very criterion of individuals' exclusion from the constituency of public justification – reasonableness – to those who fail to live up to that standard. The justifications it offers must not loose sight of the liberal commitment to respecting individuals' moral autonomy. To that end, I argue, such justifications must be rooted in the set of reasons and beliefs which individuals can be said to be committed to. Two different kinds of unreasonableness warrant different kinds of justifications: there are those who are fundamentally unreasonable because they reject the core liberal commitment to persons as free and equal, while others merely fail to be fully reasonable when encountering deep moral disagreement in political debates, falling short of the requirement to engage with others in public reason on the basis of shared values. With respect to the former, it can be argued that a commitment to regarding their fellow citizens as free and equal is implicit in their attitude and conduct towards them. The latter individuals can be offered justifications for restraint which are rooted in the character of the very moral convictions they are tempted to draw on in public reason.

Acknowledgements

Contents

Chapter 1

Introduction

Moral pluralism in modern liberal societies is alive and kicking: in recent years, there has been an abundance of political debates that are concerned with the appropriate political status of religious values, the values of cultural identity and cohesion, or with the idea of sanctity. Past years have seen controversies about Catholic employers in the United States being required to subsidize insurance covering contraception, about a Kentucky county clerk's refusal to issue marriage licenses to same-sex couples, or about several cases of business owners denying service to customers as a matter of conscience. There have further been prominent debates questioning the permissibility of flag burning, or of mocking sacred figures, as in the controversy about satirical depictions of the prophet Muhammad.

What unites these debates is that at their heart is the question of which – and whose – values should have weight in determining the political decisions that will shape our societies. In other words, they raise questions of whose voice should be heard, which kind of arguments should carry any weight, and what disqualifies individuals and their reasons from being taken seriously in public political debates on matters of public concern. Should we be concerned if people invoke controversial reasons of religious or moral conscience in opposition to provisions which others view – for potentially equally controversial reasons – as essential matters of health care and basic rights? On what basis should members of a liberal society resolve such disputes? And, not least, how, if at all, should they engage with those whose views fundamentally oppose their core liberal values, such as the freedom and equality of all citizens?

Liberal political theory, and specifically political liberalism, has something to offer to these debates. Rawls's original proposal of this strand of theory specifically seeks to devise liberal principles of justice for societies characterised by the moral pluralism which is generated by liberal political institutions, that is, for precisely the kind of moral pluralism which has come to thrive in Western democracies. It looks at how such principles can emerge and gain legitimacy under conditions of modern pluralism. Consequently, it does not derive its principles from any particular – potentially controversial – moral doctrine, but seeks to anchor them in the plurality of doctrines held by the citizens of liberal societies. In doing so, it raises precisely the question of what grounds we may draw on in determining the content of our

basic political norms and principles. Theories of public justification and public reason are a crucial part of that story. They consider the process by which political principles are selected and ask what individuals owe to each other in that sphere of political deliberation and decision making, examining which kind of reasons citizens should or should not propose to each other. In other words, theories of public reason strive to devise the standards of argument that are supposed to confer legitimacy on the decisions that emerge from that process.

Within the scope of this dissertation, I am not interested in the precise principles and policies supported by various accounts of public reason, but rather in the general viability of the project of political liberalism. I am convinced of its value and importance, given that modern liberal societies are confronted with the kind of debates detailed above. This kind of pluralism calls for a theory which helps us to determine which reasons and proposals deserve our consideration in the public political sphere, and for a theory which explicitly accounts for the diversity of moral reasons and beliefs that we are faced with in modern liberal societies. Hence, there is an important role to play for political liberalism.

1.1 Two defects

The most prominent versions of political liberalism suffer from two serious defects which give rise to tensions within its internal structure. These defects and my attempt to remedy them are the focus of this dissertation.

1.1.1 Excluding the unreasonable

The first defect is rooted in the widely-accepted position that political liberalism need not at all address itself to those who are unconvinced by core liberal ideals and principles. Proponents of this view – held prominently by, for example, John Rawls[1] and Jonathan Quong[2] – argue that liberal political theory as an *ideal* theory does not need to be concerned about whether its principles are justifiable to people whose convictions seem incompatible with core liberal values. According to them, political liberalism should only be concerned with developing principles which are appropriate to structure political interaction among people who are committed to liberal values. Liberal theorists in the Rawlsian tradition usually refer to their intended audience as those who are *reasonable.*

As a substantial moral concept, Rawls associates reasonableness with "first, the willingness to propose and honour fair terms of cooperation, and second, with the willingness to recognize the burdens of judgement and to accept their consequences."[3] Rawls further argues that

---

[1] See John Rawls. *Political Liberalism.* 2nd ed. New York: Columbia University Press, 2005, 63f.

[2] See Jonathan Quong. *Liberalism without Perfection.* Oxford: Oxford University Press, 2011, chapter 10.

[3] Rawls, *Political Liberalism*, p. 49.

> [p]ersons are reasonable in one basic 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. Those norms they view as reasonable for everyone to accept and therefore as justifiable to them; and they are ready to discuss the fair terms that others propose.[4]

In their political conduct, reasonable people are thus conceived as regarding each other as moral equals, who owe each other fair terms of cooperation. Quong argues that restricting the constituency of public justification to those who meet the criterion of reasonableness is not problematic if one adopts an interpretation of political liberalism which he refers to as the "internal conception". The internal conception of political liberalism is concerned with identifying the principles of justice which would be selected in an ideal liberal society of reasonable citizens and would be suitable to govern it.[5] On that account, individuals who find no way to accommodate basic liberal ideals in their moral conceptions have no place in the constituency of public justification, and it should not trouble us that liberal theory does not address itself to them.

This position gives rise to an internal tension within political liberalism: in addressing itself to the restricted constituency of the reasonable, political liberalism fails to respect the moral autonomy of those who are excluded from said constituency. However, it is a core liberal ideal that such respect is owed to all individuals in virtue of their personhood. Failure to offer any kind of justification whatsoever to the unreasonable is in stark contradiction to said ideal. Theories of political liberalism do not live up to some of their core ideals unless this contradiction is resolved.

1.1.2 Impoverished conception of the person

The second defect which crucially affects the viability of political liberalism also relates to the question of who is part of the constituency of public justification. Beyond excluding those individuals who reject fundamental liberal ideals, political liberalism conceives of those who *do* accept them – and to whom, consequently, justifications are owed – in highly idealised terms. They are only reasonable insofar as their conduct in the public political sphere reflects their "willingness to propose and honor fair terms of cooperation and [...] recognise the burdens of judgement and accept their consequences."[6] Accepting the consequences of the burdens of judgement, specifically, means to hold back on offering reasons that one does not expect others to be able to share, irrespective of the depth of one's own conviction. "Attuned to the burdens of judgement, a reasonable person will understand that there is no prospect of

---

[4]Ibid., p. 49.

[5]Quong, *Liberalism without Perfection*, p. 143.

[6]Rawls, *Political Liberalism*, p. 49.

persuading the adherents of disparate comprehensive doctrines to abandon their respective views and convert to her own doctrine."[7] Consequently, each instance in which a person feels unable to suppress her controversial private reasons, and asks for them to be taken into account in the process of political decision making, disqualifies her from the constituency of public justification.

This defect of political liberalism moves into focus when considering those citizens who, while counting as reasonable by all other accounts, fail to live up to that standard in specific cases. Instances of this kind are common in actual liberal societies. They are, for instance, likely to occur in relation to bio-ethical debates, which raise questions about what respect for (human) life and the person demands. Moral convictions on these matters are often rooted in foundational moral beliefs that are crucial to individuals' comprehensive moral doctrines, and sometimes their metaphysical conceptions of the world. Consequently, individuals may find some of these deep moral convictions hard or impossible to discount in some cases, despite being impeccably reasonable in all other respects. Areas of political conflict where such foundational convictions are likely to play a role include such issues as the regulation of genetic engineering, assisted suicide, and abortion. Other examples of political disagreements that may draw on similarly fundamental values concern the morality of torture, the death penalty, and the ethics of war.

Branding individuals who act upon their deep moral convictions in these instances as unreasonable excludes a significant number of individuals from the constituency of public justification. In other words, those who are in said constituency do not reflect the internal moral constitution of many of those people who populate the kind of modern pluralistic societies that political liberalism strives to accommodate. Within the theory, the individual reasonable comprehensive moral doctrines which are supposed to provide support for liberal principles are deprived of the moral force they actually have for individuals in cases such as those mentioned before. Political liberalism does not even acknowledge the possibility of conflict between the substance of citizens' comprehensive moral doctrines and what reasonableness demands of them. Rather, it excludes those doctrines which contain even the slightest potential of such conflict. That is, the moral authority that an individual may consider her convictions to have, as well as her moral autonomy in navigating their potentially complex and conflicting demands is never explored. Hence, political liberalism operates on the basis of an impoverished conception of the person and an unrealistic idea of the doctrines that shape the very society that political liberalism is concerned with. This idealisation might not matter too much, were it not for the fact that it affects crucial, distinguishing elements of what political liberalism purports to be: a theory concerned with people who are capable of exercising their moral autonomy and who are doing so on the basis of a variety of different comprehensive doctrines. Yet, the idealised individuals populating political liberalism's constituency of public justification seem to experience none of the moral force of their comprehensive doctrines and exercise none of the moral autonomy which liberalism values so

---

[7]Matthew H. Kramer. *Liberalism with Excellence*. Oxford: Oxford University Press, 2017, p. 10.

highly. This raises the question whether the principles yielded by the thought experiment of political liberalism are at all suitable for the modern pluralistic societies which are populated by individuals of whom many are likely to experience the potentially divergent forces of their comprehensive doctrines and their liberal commitments. However, these are precisely the societies which political liberalism claims to be concerned with.

1.2   Broadening the constituency of public justification

With both defects, political liberalism fails by its own standards. This is because, from the outset, the constituency of public justification is restricted in a way which is incongruent with the aims of the theory. Hence, there is value in exploring whether the constituency of public justification can be rendered sufficiently broad to evade these charges, and in exploring to which degree that adapted constituency can still support the basic tenets of political liberalism. Doing so may also lead us to the conclusion that those individuals who are currently branded as unreasonable may, after all, justifiably be excluded from the constituency of public justification at the next stage.[8] But for political liberalism to succeed on its own terms – its core commitment to all people as morally autonomous, free and equal, and its purpose of devising political principles for modern pluralistic societies – it must first be prepared to address justifications to *all* members of these societies.

It is not possible for a theoretical argument about justification to avoid some degree of idealisation. In order to get a justificatory argument off the ground, we must make some assumptions about what beliefs and principles are shared by those individuals to whom we address ourselves. But in keeping such assumptions to a minimum, I intend to show that it does not take very much to accept two core liberal principles: the idea of the moral equality of persons and the commitment to coerce others only for reasons they can share. What I am going to offer is a rationalist justification of these two principles which, *(i)* does not require an arsenal of supplementary assumptions for its success, and *(ii)* where it does require such assumptions, draws on elements of individuals' systems of reasons and beliefs that these individuals themselves would be unwilling to deny.

In order to address the first defect and the tensions arising from it, I am going to argue that rejecting the moral equality of one's co-citizens is not only immoral but actually irrational: it entangles the person making that argument in performative contradictions that they cannot rationally resolve. To address the second defect, I am going to argue that individuals have rationally compelling reasons to resolve their cognitive dissonances between the demands of reasonableness and their private moral commitments in favour of the former. These reasons, I argue, are ultimately tied to the foundational character of those private values that give rise to said dissonances in the first place.

---

[8]Depending on the substance and scope of the beliefs they want to see enacted in the political sphere, the scope of their exclusion may equally vary.

1.3 Chapter overview

This dissertation is structured as follows:

- **Chapter 2** explores the two defects in greater detail and concludes that the constituency of public justification must not be restricted without addressing justifications for those restrictions to all.

- **Chapter 3** engages with different conceptions of justification, settling on a weakly internalist approach to identifying sources of reasons for individuals. It further proposes and defends a rationalist standard of inference for emulating a person's reasoning based on these sources.

- **Chapter 4** shows how the commitment to persons as free and equal underlying the demands of reasonableness is almost universally justifiable by drawing on reasons implicit in each and every person's conduct towards her fellow citizens in the sphere of public reason.

- **Chapter 5** explores the concept of deep moral disagreement. It explains why, under these circumstances, invoking reasonableness as defended in the previous section is insufficient to demand restraint in invoking reasons rooted in individuals' private comprehensive doctrines. This is pertinent if political liberalism is to avoid the fallacy of falling prey to an impoverished conception of the morally autonomous person and of misrepresenting the nature of their comprehensive moral doctrines.

- **Chapter 6** argues that the requirement of restraint in invoking reasons arising from private comprehensive doctrines is justifiable to all. I propose a rationalist argument which explores the nature of the foundational beliefs which are at the heart of such doctrines, arguing that their structure makes it impossible for individuals to render them accessible to others.

- **Chapter 7** concludes that rationalist arguments enable us to broaden the scope of the constituency of public justification. It is no longer necessary to filter out all moral conceptions which do not conform to the criteria of reasonableness from the very beginning. As a result, political liberalism can live up to its own aims and principles, offering an internally consistent philosophical framework for addressing challenges posed by morally diverse liberal societies.

Chapter 2

Why reasonableness must be justified

2.1 Introduction

Political liberalism is premised on a commitment to the ideal of persons as ends in themselves, conceiving of them as morally autonomous and self-determining agents. This ideal of respect for persons is the foundation of the idea that political principles and policies backed by the threat of coercion must be publicly justified to those who are subject to them. Within political liberalism, the constituency of those to whom justifications must be offered is restricted to reasonable persons. In this chapter, I will defend the following core premise of this thesis: that this restriction is itself in want of justification to those who are affected it by it. In other words, I will argue that political liberalism must justify the ideal of reasonableness as the criterion for membership in the constituency of public justification to those who are excluded from it because they fail to be reasonable.

2.2 The key challenge: inclusiveness versus pluralism

2.2.1 Moral pluralism and liberal ideals

Liberalism may be said to be rooted in a commitment to universally granting the same moral concern to all persons. A distinguishing feature of liberal political theory consists in its commitment to respecting the individual person, *i.e.* to respecting and protecting the capacities she is endowed with by virtue of her autonomy and capacity to reason, which are the bases of her agency.[1] What liberalism recognises as crucially valuable to an individual is her capacity as a subject to freely determine her actions according to her ends.[2] Rawls argues that

---

[1]See, for example, Jeremy Waldron. *Liberal rights: collected papers, 1981–1991*. Cambridge: Cambridge University Press, 1993, pp. 36, 62.

[2]Ibid., p. 41.

one "respect in which citizens view themselves as free is that they regard themselves as self-authenticating sources of valid claims."[3] Respecting this value could be framed as treating all individuals as ends to one's own self-determination, imposing upon oneself the requirement not to interfere with their process of self-determination without appealing to them to include one's reasons for interfering among their ends. Equal respect for persons requires them to be treated never solely as means, or instruments to the will of others, but always *as ends in themselves*, as Charles Larmore emphasises in a Kantian vein.[4]

This is not to say that coercion, *i.e.* treating individuals as a means to achieve a certain aim by forcing them to behave in a way conducive to the intended outcome, may never be warranted by liberal principles. For us to treat others as ends rather than means merely requires coercive action to be based on ends that they could share,[5] that is, ends that can be said to give them reason to act just as they give us reason to act. In Larmore's words, '[t]o respect another person as an end is to insist that coercive [...] principles be as justifiable to that person as they are to us.'[6] This requirement of justification is also not least a symptom of the liberal respect for individuals' rationality and their capacity to make sense of the world,[7] enabling them to recognise which external claims can or cannot legitimately demand to be included in their realm of ends. Again, emphasising the liberal commitment to the individual as the standard for the adequacy of inter-subjective action, Waldron holds that "the liberal insists that intelligible justifications in social and political life must be available in principle for everyone, for society is to be understood by the individual mind, not by the tradition or sense of a community."[8] Political power being coercive power, political institutions that wield this power hence need to ensure that their measures are acceptable – or could be made acceptable – to all those whose individual freedom is or could be limited as a result. Waldron thus goes on to argue that "[i]f there is some individual to whom a justification cannot be given, then so far as *he* is concerned the social order had better be replaced by other arrangements, for the status quo has made no claim to *his* allegiance."[9] The purposes underpinning state action need to be acceptable to all. In order to ensure that this is the case, liberal political procedures need to address themselves to all citizens (with citizens – somewhat simplistically – being defined as the group of people affected by the coercive domestic power of a state), irrespective of their particular ends, views, and convictions. To provide to all people the justifications they deserve by virtue of their personhood thus requires the constituency of those to whom public justifications are owed – or, in Jonathan Quong's terms, the "constituency of public justification"[10] – to be maximally inclusive.

---

[3]Rawls, *Political Liberalism*, p. 32.

[4]Charles Larmore. "Political Liberalism". In: *Political Theory* 18.3 (1990), p. 348.

[5]Thomas Nagel. "Moral Conflict and Political Legitimacy". In: *Philosophy & Public Affairs* 16.3 (1987), p. 159.

[6]Larmore, "Political Liberalism", p. 349.

[7]Waldron, *Liberal rights: collected papers, 1981–1991*, p. 41.

[8]Ibid., p. 44.

[9]Ibid., p. 44. Original emphasis.

[10]Jonathan Quong. "The Rights of Unreasonable Citizens". In: *Journal of Political Philosophy* 12.3 (2004),

For liberalism, as a doctrine that is primarily concerned with the autonomy, freedom, and capacity for self-determination of individual persons, the promotion of these values takes precedence over a public commitment to implementing more comprehensive conceptions of the good life. It is not least this very emphasis on safeguarding individuals' sphere for autonomous self-determination that invites diversity among citizens in liberal society, enabling them to privately pursue their particular conceptions of the good that liberal theory abstains from promoting. Over the last decades, liberal societies have come to embrace their role as hubs of social, cultural, and religious pluralism.

By steering clear of publicly rooting its core ideals in specific metaphysical foundations, modern liberal political theory aims to widen the scope of its appeal. It thus allows citizens to commit themselves to honouring liberal values in the political sphere for sets of reasons that are specific to each citizen, rooted in their particular private metaphysical, moral or ethical convictions about (the good) life. To put it in Rawlsian terms as set forth in *Political Liberalism*, liberal values draw their legitimacy from being subject to an overlapping consensus, drawing support from a variety of comprehensive doctrines which are compatible with said values.[11] Liberalism thus conceived may – but does not have to – assume the same status as a comprehensive doctrine. Rather, liberal values and principles can be framed as freestanding, publicly detached from potentially divisive metaphysical roots,[12] while remaining individually rooted in a variety of narratives furnishing them with legitimacy for the respective person. For the individual, these values may assume the status of moral truths or commands issuing from a particular metaphysical narrative. Collectively, however, liberal values must only acknowledge their political status as consensually chosen tools for regulating public affairs.

Politicising liberalism is an act of both epistemic humility and pragmatism in the face of the actual diversity of modern societies and the multitude of frameworks of reference which are prone to limit the actual persuasive power of any philosophical theory. Hence, irrespective of whether there is one universally true reason to commit oneself to liberal values, *political* liberalism seeks to provide a somewhat shallow yet fertile bed for liberal ideals to flourish in modern society. This move ultimately privileges preserving its capacity to demonstrate the utility of its core values for political practices in diverse societies over doctrinal depth and unity.

2.2.2 Inclusiveness and exclusiveness in public justification

Despite its commitment to preserving the autonomy, freedom and the capacity for self-determination of all people, political liberalism nevertheless relies on a certain degree of justificatory exclusiveness. Empirically, not all comprehensive doctrines that citizens hold

---

p. 314.

[11]Rawls, *Political Liberalism*, pp. 10-11.

[12]Ibid., p. 10.

can be interpreted in such a way that they are compatible with even a metaphysically unburdened overlapping consensus that supports liberal values.

Although it may be wrong to deprive citizens of all rights and liberties they are granted by liberal political principles, liberals may need to restrict admission to the constituency of public justification and may rightfully do so, as Jonathan Quong argues, with regard to citizens whose views challenge some or all of the very values and principles that are fundamental to the liberal project.[13] Quong holds that the reason based on which those citizens – whom he refers to as *unreasonable* – can and need to be excluded from the constituency of public justification consists in the very conflict of these doctrines with fundamental liberal values:

> unreasonable citizens reject the basic project of public justification that lies at the heart of a liberal, deliberative democracy. In denying that political power should be subject to public justification, they show contempt for the fundamental moral ideal that underlies the project: the idea that all citizens are free and equal. [...] Political liberalism does not address itself to unreasonable citizens because it is a theory about the freedom and equality of citizens. Since unreasonable people by definition reject this premise, their (unreasonable) views are simply of no normative interest in the process of political justification.[14]

Quong thus assumes that the requirement of justification does not apply to unreasonable citizens, because, by virtue of their unreasonable views, they cannot have a genuine interest in taking part in a justificatory public political discourse whose purpose consists in realising distinctly liberal values, both procedurally and substantially. Since they would not be prepared to participate in the process of public justification on these terms, unreasonable citizens are deemed to be beyond the scope of the liberal justificatory project.

However, taken as a justification for the liberal warrant to limit the inclusiveness of the constituency of public justification, this argument suffers from a crucial defect. It draws on the requirement of endorsing the liberal value of regarding citizens as free and equal, as well as the principle of public justification, without also paying due respect to the origin of these ideals: the liberal conception of the person as morally autonomous and self-determining, and the liberal commitment to treating persons as ends – or in Rawls's words, as "self-authenticating sources of valid claims."[15] In other words, Quong's argument claims that the liberal principle of justification only needs to be applicable to those whose world view is already distinctly liberal, while disregarding the fact that the very principle of justification is rooted in the liberal commitment to treating all persons as ends – a commitment which refers to their quality of personhood, not their attitudes to others. Citizens deserve to be offered justifications, not in virtue of their own willingness to do so to others, but owing to their status as persons.

---

[13]Quong, "The Rights of Unreasonable Citizens", p. 314.
[14]Ibid., p. 315.
[15]Rawls, *Political Liberalism*, p. 32.

References to the nature and the content of citizens' unreasonable doctrines themselves are unconvincing arguments for denying them access to the constituency of public justification.

Quong's account furthermore underestimates the potential sophistication of unreasonable attitudes. It is, after all, conceivable that citizens could ultimately reject principles such as public justification or values like the equality of persons, but nevertheless pursue and seek support for their unreasonable ideas within existing liberal democratic institutions and procedures. In this case, their contempt for the idea of public justification may not be the same as that of a person who pursues her unreasonable aims outside the public political process (*e.g.* through violent means), but their instrumental adherence to justificatory principles serves to mask their underlying unreasonable intentions. Hence, even if the above argument were sound, citizens who merely used the public political discourse in order to promote their unreasonable doctrines – pretending to engage in justificatory discourses with others – could at least not be accused of openly rejecting the idea and project of justification. Erin Kelly and Lionel McPherson argue in favour of including such persons in the constituency of public justification despite the fact that they only 'appear to be politically reasonable [and] may accept the political conception as a mere *modus vivendi.*'[16] They hold that the inclusion of those citizens in the constituency of public justification is more compatible with ensuring 'the greatest range of equal rights and liberties for all'[17] than denying to them the right to be offered justifications.[18]

It is, however, doubtful whether the toleration of such attempts to promote illiberal values by liberal means is compatible with the protection of equal rights and liberties in the long run. It is this concern that provides liberals with another, more compelling argument for limiting the inclusiveness of the constituency of public justification. If liberal polities were unable to prevent unreasonable doctrines from being included in the constituency of public justification, they might prove to be incapable of effectively defending their fundamental ideas and principles in public justification. While a prudential consideration in practice,[19] avoiding such a internal conflict is also desirable on the theoretical level: if the principle of justification to all were to be universally and unconditionally valid, it would require illiberal views and proposals to be treated with the same respect as those which are in line with liberal ideals. As a result, liberals could only consistently defend their ideals with respect to the procedures of political decision making, and would be unable ensure that substantive policies live up to basic liberal values. Given that such policies might ultimately also affect

---

[16]Erin Kelly and Lionel McPherson. "On Tolerating the Unreasonable". In: *Journal of Political Philosophy* 9.1 (2001), p. 54, original emphasis.

[17]Ibid., p. 42.

[18]Ibid., p. 55.

[19]Tolerating the expression of illiberal views might allow such proposals to undermine the political procedures based on liberal values. As Marilyn Friedman holds, "if one is seeking fair terms of social cooperation among persons who are free and equal and who are assumed to disagree reasonably on fundamental comprehensive matters, then one must not allow persons who *reject* this goal or these assumptions to hijack the legitimation process." (Marilyn Friedman. *Autonomy, Gender, Politics.* Oxford: Oxford University Press, 2003, p. 169).

procedural principles, liberals would be trapped in a fundamental conflict: true commitment to liberal ideals in terms of both liberal democratic procedures, and the realisation as well as the protection of substantive liberal values within the rights, liberties and benefits that form part of any particular policy, cannot be upheld without imposing boundaries on either the content of individual legislation, or on participation in the constituency of public justification. Without such limits, a commitment to liberal values could ultimately be led *ad absurdum,* with individuals being bound to respect illiberal positions out of liberal motives.

In order to maintain consistency within political liberalism, the criterion for selecting the set of views eligible for entering the public political discourse of a liberal society needs to ensure that the positions which are granted access to the constituency of public justification do not threaten to counteract liberal ideals in the short or long run. Limiting inclusion in the constituency of public justification to the set of moral attitudes and views which are reasonable is one such theoretical safeguard, as I will argue in the following.

2.2.3   The self-referentiality of reasonableness

The idea of *reasonableness* is used as a criterion for regulating access to the constituency of public justification, most notably in John Rawls's *Political Liberalism.*[20] Two dimensions may be distinguished with regard to Rawls's conception of *reasonableness:* a cognitive and a moral one.[21] For Rawls, the cognitive dimension of reasonableness is expressed in individuals' "similar powers of thought and judgement" and their ability to "draw inferences, weigh evidence, and balance competing considerations" – capacities that people share in virtue of a "common human reason".[22]

Going beyond these basic cognitive capacities, Rawls also conceives of reasonableness as a richer moral concept. He associates this moral dimension of reasonableness with "first, the willingness to propose and honor fair terms of cooperation, and second, with the willingness to recognize the burdens of judgement and to accept their consequences."[23] Rawls further argues that

> [p]ersons are reasonable in one basic 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.
> Those norms they view as reasonable for everyone to accept and therefore as

---

[20]See also Barbara Herman. "Pluralism and the Community of Moral Judgment". In: *Toleration. An Elusive Virtue.* Ed. by David Heyd. Princeton: Princeton University Press, 1996, pp. 60–80; Barbara Herman. "Moral pluralism and political consensus". In: *The Idea of Democracy.* Ed. by David Copp, Jean Hampton, and John E. Roemer. Cambridge: Cambridge University Press, 1995, pp. 270–291; and Charles Larmore. *The Morals of Modernity.* Cambridge: Cambridge University Press, 1996, chapters 6 and 8.

[21]In designating these dimensions, I am following Gerald Gaus's terminology as set out in "The Rational, the Reasonable and Justification", p. 234.

[22]Rawls, *Political Liberalism*, p. 55.

[23]Ibid., p. 49.

justifiable to them; and they are ready to discuss the fair terms that others propose.[24]

Crucially, this moral definition of reasonableness contains a commitment to persons as free and equal. Reasonable people are willing to show respect to others in the above sense, because they recognise them as free "in virtue of their two moral powers (a capacity for a sense of justice and for a conception of the good) and the powers of reason (of judgement, thought, and interference connected with these powers)",[25] and equal in being owed fair and reciprocal treatment. In his original treatment of this topic in *A Theory of Justice*, Rawls emphasises that "[t]hose who can give justice are owed justice."[26] Rawls's conception of persons as free and equal is explicitly concerned with moral personality as a potentiality. Citizens' respect for each other as free and equal persons is therefore independent of the degree to which a person realises this potentiality, *i.e.* whether she actually acts – or has a desire to act – in accordance with principles of justice. He holds "that the *capacity* for moral personality is a sufficient condition for being entitled to equal justice",[27] arguing that only very few human beings can be considered to lack this attribute.[28]

This idea of reasonableness serves as the criterion for limiting access to the constituency of public justification in Rawls's political liberalism. This becomes apparent in his discussion of the procedure of political constructivism, which is Rawls's method of choice for the development of political principles. A constructivist political conception draws "the principles of justice from public and *shared* ideas of society as a fair system of cooperation and of citizens as free and equal using the principles of their common practical reason."[29] Rawls states that "[i]n this procedure, [...] rational agents, as representatives of citizens and subject to reasonable conditions, select the public principles of justice to regulate the basic structure of society."[30] These principles of justice are developed "from public and shared ideas of society as a fair system of cooperation and of citizens as free and equal".[31] In other words, these principles are constructed from the idea of a society whose members are reasonable in the Rawlsian sense. It is therefore hardly surprising that Rawls also explicitly limits the range of permissible attitudes and positions within the liberal political discourse to those espoused by reasonable comprehensive doctrines. According to him "[t]hese are the doctrines that [...] political liberalism must address."[32] Unreasonable views, in contrast, do not merit inclusion in the constituency of public justification. In fact, Rawls asserts that their presence in society

---

[24]Ibid., p. 49.

[25]Ibid., p. 19.

[26]John Rawls. *A Theory of Justice*. Cambridge, MA: Belknap Press of Harvard University Press, 1972, p. 510.

[27]Ibid., p. 505.

[28]Ibid., p. 506.

[29]Rawls, *Political Liberalism*, p. 90, emphasis added.

[30]Ibid., p. 90.

[31]Ibid., p. 90.

[32]Ibid., p. 36.

imposes on liberals the "practical task of containing them – like war and disease – so that they do not overturn political justice."[33]

In attributing normative significance in the political realm only to reasonable views, Rawls's approach thus appears to mitigate the potential for tensions yielded by the liberal ideal of maximal justificatory inclusiveness and the subsequent necessity to explicitly exclude those individuals who pose a threat to the realisation of liberal ideals in general. However, these particular restrictions on the access to the constituency of public justification present a new challenge to liberal political theory within pluralistic societies. This challenge arises from the merely self-referential nature of the liberal ideal of reasonableness, with its indifference to rendering itself externally comprehensible and defensible. Political liberalism is unable to explain the normative significance of reasonableness to a neutral external observer who is neither committed to liberal ideals nor rejects them, but merely intends to understand the merit of reasonableness as a norm governing political interaction. Such an observer would ask why individuals should be reasonable, or why it is more appropriate to be reasonable than to be unreasonable when acting in the political domain. This is a question of practical reason. However, within Rawls's account, for a person to reason practically is to ask merely how her reasonable disposition requires her to act. This is because Rawls defines "the principles of [individuals'] common practical reason"[34] in terms of a substantial, moral conception of reasonableness. Rawls asserts that "the principles of practical reason – [consisting of] both reasonable principles and rational principles – and the conceptions of society and person are complementary."[35] The principles of practical reason thus cannot be understood separately from his moral conception of persons and society. This conception, again, is framed in terms of reasonableness, *i.e.* the dispositions which political liberalism considers to be normative in political interaction among free and equal persons. Hence, practical reason in a Rawlsian sense cannot provide an external observer with independent variables that account for the normative significance of reasonableness, precisely because it is itself defined in terms of reasonableness.

As a result, reasonableness is basically incontestable from within the framework of political liberalism, yet apart from self-referential affirmation, remains unable to provide reasons to those who, so far, do not endorse the liberal ideals it contains. A commitment to the these ideals, thus remains a prerequisite for being able to accept the normative significance of reasonableness in the political sphere.

2.2.4   Addressing justifications to the unreasonable

The preceding, self-referential argument for the normative significance of reasonableness turns to be problematic as soon as liberal political theory is confronted with a plurality of

---

[33]Rawls, *Political Liberalism*, p. 64.
[34]Ibid., p. 90.
[35]Ibid., p. 107.

moral frameworks, as David Estlund points out. He argues that

> political liberalism must find some way to penetrate this plurality of insular groups.[36] This is where it must appeal [...] not to reasonableness alone. The difficulty cannot be avoided by saying that "we the reasonable" should just carry on and ignore the other views about the authoritative group rather than insisting that they are false. [...] For if they were not mistaken, they would be the ones with rejection rights and we would not. The question is [...] whether we can or cannot suspend judgement on it. We cannot, since suspending judgement would leave us with a plurality of insular groups, none evidently having a better claim to be authoritative than any other.[37]

In other words, if reasonableness cannot prove its validity beyond the realm of its adherents by criteria external to its own standard, its normative authority as a criterion for limiting access to the constituency of public justification will remain elusive to people outside the liberal community. This elusiveness should strike us as problematic for political liberalism once we consider the roots of the liberal commitment to justification. Justifications are owed to all in virtue of their personhood and the status as ends in themselves that it confers on them. To echo Waldron's statement, the lack of a justificatory intention deprives liberal polities of a legitimate claim to the allegiance of those citizens who are outside of the insular group of the reasonable.[38]

That said, in the ideal liberal society which Rawls has in mind when framing the conditions for an overlapping consensus on just political principles, the choice of reasonableness as the criterion for inclusion in the constituency of public justification may not violate the broader liberal requirement to provide justifications to all individuals. In such an ideal society which consists only of reasonable people, there would be no need to exclude anyone from the constituency of public justification. However, in non-ideal societies in which some people do in fact hold unreasonable doctrines, the failure to provide these citizens with reasons for their coercion – specifically, reasons that they can accept as valid – violates liberalism's universalist commitment to justification based on its inclusion of all people in the realm of ends. The latter commitment is also affirmed by Rawls, who recognises all human beings as free and equal merely in virtue of their *potential* for moral personhood.[39]

However, given this conflict, how can liberalism cope with the condition of pluralism in modern societies? Modern liberal, but non-ideal, morally pluralistic societies are empirically highly unlikely to consist solely of reasonable citizens. If liberal ideals are to be of any

---

[36]Groups whose members' consent is the only prerequisite for including a doctrine in the process of political justification.

[37]David Estlund. "The Insularity of the Reasonable: Why Political Liberalism Must Admit the Truth". In: *Ethics* 108.2 (1998), p. 262.

[38]Waldron, *Liberal rights: collected papers, 1981–1991*, p. 44.

[39]As Rawls argues in both *Political Liberalism* (p. 19) and *A Theory of Justice* (§77).

relevance to these societies, political liberalism must be revised in such a way that allows it to effectively defend its values, without in doing so failing to live up to some of them. This is something political liberalism cannot avoid in its current state, revealing a fundamental internal tension: Rawls's political liberalism is defined by the limits it sets to the inclusiveness of the constituency of public justification, barring all unreasonable doctrines and their proponents from participating in the public political discourse. At the same time, the fundamental liberal commitment to treating all persons as ends imposes upon liberals a requirement to justify state action to all individuals who are coerced by it. This latter norm, requiring the constituency of public justification to be maximally inclusive, however, is in conflict with imposing limitations on the membership in the constituency of public justification, specifically because the limiting criterion – reasonableness – is merely self-referentially justified. As I have argued in this section, said criterion is not backed by reasons which appeal to anyone outside the group of reasonable people as to why it should be furnished with the authority to regulate admission to the constituency of public justification. Yet, this is exactly what the liberal norm of universal justification requires in virtue of its concern for treating all persons as ends. Consequently, political liberalism is caught in a dilemma between a normative requirement for the constituency of public justification to be universally inclusive and the fact that the way in which it establishes its criterion for membership in said constituency violates the former norm.

This dilemma can be resolved if political liberalism can be revised such that its restrictions on the access to the constituency of public justification can be said to be justified to all: political liberalism needs to provide unreasonable people with a justification for their exclusion from all further justifications. In other words, it must be able to demonstrate that unreasonable people have reason to accept their exclusion from the constituency of public justification. Not to do so would, as I have argued, constitute an expression of disrespect for their equal moral value – for their entitlement to be treated as ends in themselves, rather than means to the purpose of developing and establishing liberal principles. Yet, if unreasonable people *themselves* recognised reasonableness as a valid criterion for selecting those whose views are to be eligible to be taken into account in the process of public justification – *i.e.,* if their coercive exclusion could be considered to be justifiable *to them* – their coercion would no longer amount to them being treated merely as means. In accepting their coercion as justified, they would also accept the end of their coercion as an end for themselves. Justifying to unreasonable people the grounds, or the criterion for their exclusion from the constituency of public justification, is thus a way for political liberalism to live up to its commitment to universal justification.

Working within the framework of *political* liberalism, we should refrain from drawing on any comprehensive moral doctrines when looking for justificatory reasons that might induce unreasonable individuals to recognise reasonableness – or the values it incorporates – as the only adequate attitude within liberal political discourses and processes. Reasonableness needs to be justifiable to unreasonable people from a *political* standpoint. One such source

of reasons capable of justifying reasonableness as the criterion for participation in public justification lies within the very practice of justification itself. As liberals, I will argue in the following chapters, we need to search for reasons that the inclusion of unreasonable individuals in justificatory procedures would give them to be reasonable (or, more precisely, to adopt only reasonable positions) in their capacity as members of the constituency of public justification. An argument for the justifiability of the exclusion of unreasonable people from public justification can hence be based on the following conditional: *if* unreasonable people sought admission to the constituency of public justification,[40] and *if* they participated in a justificatory discourse with all other citizens, their performance of these actions would provide them with a source of reasons to consider reasonableness as a necessary criterion for inclusion in said constituency. I will return to this idea later. In the remainder of this chapter, I will consider – and ultimately reject – an alternative approach to addressing the challenge presented by the unreasonable, provided by Jonathan Quong's internal conception of political liberalism.

2.2.5 The internal conception of political liberalism: a means to resolve the tension?

It is a legitimate concern to question whether or not the self-referential justification of reasonableness must by all means be considered a vice that can only be remedied by constructing justifications for asserting reasonableness that even unreasonable people can relate to. The answer ultimately hinges on the question of how to best interpret the purpose of political liberalism. The justificatory circle that sustains reasonableness as the criterion for entering the constituency of justification is obviously a vicious feature if one conceives of the principal purpose of political liberalism as striving to accommodate the broadest possible constituency of individuals. Quong refers to this interpretation of liberalism in which said constituency imposes external constraints on the content of justifiable principles as the *external conception*.[41]

> On one view, the fact of pluralism or disagreement is a fact about the world to which liberalism must accommodate itself if it is to be considered a sound position in political theory. Liberalism's most foundational norms and principles, on this approach, lack an adequate grounding if they cannot be justified to the diverse constituency of persons that currently inhabit modern liberal societies. If political liberalism cannot be shown to work – if no version of liberal justice

---

[40]This is the only case we need to be concerned with, as those who are not even willing to put forward their unreasonable positions in a justificatory discourse – thus rendering it impossible to take their interests into account in the public justification of coercive measures – can hardly claim to be treated unjustly if, as a result, they are being coerced by measures which they do not consider to be justifiable to them. As a result, I am also only concerned with citizens whose unreasonableness is sophisticated enough for them to intend to make use of existing public political procedures in order to promote and seek support for their unreasonable doctrines.

[41]Quong, *Liberalism without Perfection*, p. 139.

can be publicly justified in light of the pluralism that is a feature of contemporary liberal democratic societies – this would demonstrate that liberal rights and institutions are not justifiable, and thus lack legitimacy.[42]

Given this interpretation of political liberalism, the failure to render the core virtue of conduct in the public political discourse – reasonableness – accessible to this maximally inclusive constituency of justification does indeed expose a tension within the theory. A merely self-referential justification of reasonableness cannot expect to be acceptable to all in light of the diverse set of views present in actual liberal societies.

However, the self-referential justification of reasonableness loses its vicious character in the context of the alternative interpretation of political liberalism which Quong advocates. His *internal conception*[43] focuses on reasonableness not as the criterion regulating access to the constituency of justification, but rather as the pervasive feature of the kind of society that political liberalism ought to be concerned with.

> Because the internal conception sees the fact of reasonable pluralism as an internal challenge and not an external constraint on liberal theory, it has a correspondingly different way of defining the constituency of reasonable citizens. The internal conception does not aim to achieve public justification or normative stability in current liberal democratic societies – its aim is instead to work out the content and structure of liberal political justification given the fact that any well-ordered society will be characterised by reasonable pluralism.[44]

Hence, the internal conception declares that limiting the scope of the constituency of public justification to reasonable people does not itself require justification, because the project of political liberalism is only concerned with working out appropriate political principles for societies which are characterised by *reasonable* pluralism. The role reasonableness plays within the internal conception does not merely fail to provide unreasonable people with justifications – either for specific political decisions or their exclusion from the constituency of public justification – but rather dismisses them as objects of liberal concern altogether.

To summarise, the approach taken by the internal conception preserves the status of reasonableness as a self-sustaining criterion, insulating itself against challenges that demand that the justificatory scope for the standard of reasonableness be extended. Within the internal conception, all citizens are first and foremost reasonable, committed to prioritising their pledge to treat others as equals in their effort of mutual political cooperation, while honouring the burdens of judgement. It is this idealised conception of the reasonable citizen that, as I will argue in the following section, is the key fallacy of the internal conception, which

---

[42]Quong, *Liberalism without Perfection*, pp. 138-139.
[43]Ibid., p. 139.
[44]Ibid., p. 143.

threatens its *internal* consistency upon confrontation with a further challenge: deep moral disagreement. In this context, it will become obvious that the internal conception's theoretical conception of the person is in danger of erasing from individual persons the very traits liberal theory is committed to promoting and protecting: their moral autonomy as well as their capacity for self-determination which does not least include their capacity for a sense of justice and their capacity to develop and revise a conception of the good. As a result, by depriving its hypothetical citizens of the capacities necessary for conducting themselves as morally autonomous beings, the internal conception calls into question its credentials as a liberal theory. It is for this reason, as I will argue in the remainder of this chapter, that the internal conception of political liberalism ultimately cannot overcome the need to justify reasonableness.

2.3 A defect in the internal conception

Moving away from the tension arising from the presence of fundamentally unreasonable people who are excluded from the constituency of public justification, the internal conception of liberalism seems to be faced with a challenge from *within* said constituency. As I will argue in this section, even if all members of the constituency of public justification are indeed reasonable, the internal conception of liberalism might still be in conflict with the fundamental liberal commitment to honour the autonomy and capacity for self-determination of all, fundamentally misconceiving their relation to their individual comprehensive moral doctrines, as well as the complexity of the latter. This challenge to the internal conception, I intend to demonstrate, is the result of the reductive and static conception of the ideally reasonable person that the internal conception relies on for the sake of theoretical self-sufficiency.

2.3.1 Public reason and reasonable disagreement

The internal conception regards as its sole object of concern the constituency of justification of an ideal liberal society characterised by reasonable pluralism. As reasonable individuals, members of said constituency are defined in terms of their commitment to proposing and realising a framework of mutual cooperation and accepting the burdens of judgement. These reasonable citizens are assumed to expect political decisions to take place in a shared justificatory framework in which all participants are required to frame their proposals in terms of reasons that are endorsed by all other citizens. This is the core idea of the practice of public reason. According to Quong, Rawls's "idea of public reason is best seen as a practical ideal derived from the liberal principle of legitimacy."[45] This principle requires that for political power to be legitimate, it must be "exercised in accordance with a constitution the essentials of which all citizens as free and equal may reasonably be expected to endorse in light of

---

[45]Ibid., p. 43.

principles and ideals acceptable to their common human reason."[46] This restricts the content of public reason to those claims that all reasonable citizens endorse.[47] In other words, when engaged in public reason, a person must only refer to values that are shared by *all,* rather than offering others reasons merely because she deems them to be valid within the context of her private comprehensive moral doctrine. The reasons based on such shared values gain the status of public reasons because they are subject to an overlapping consensus of the comprehensive moral doctrines of all members of the constituency. Offering only public reasons expresses respect for individual moral autonomy by ensuring that the public justification for all coercive measures is rooted in a reason that every individual citizen can accept. The internal conception thus assumes that the reasonable citizens populating the constituency of public justification of an ideal liberal society will draw on such public reasons in support of any policy under discussion in public reason.

Within such a framework of public justification in a liberal society, reasonable disagreement about matters which are subject to the exercise of political power is possible and likely to occur. As Quong points out, "[r]easonable people [...] disagree about the principles of justice just as much as they do about conceptions of the good life. All we need to do is look at current political debates over issues like abortion, taxation, capital punishment, health care, or freedom of expression to see that many reasonable people disagree in deep and seemingly intractable ways about justice."[48] Such disagreement is "a fact about the exercise of rationality under liberal conditions"[49] and thus a facet of the very fact of reasonable pluralism, which is the object the internal conception.

Reasonable disagreement of this kind – in which all arguments duly refer to shared public values – may occur simply because some citizens assign different priorities to particular shared values than others do. According to Quong, whichever way we decide in such a case of reasonable disagreement, the decision would be justified, given that it is premised on a shared value.[50] The fact that to some a given value might not appear to be the most salient one does not damage the legitimacy of the decision in question. In this case, the proponents of the argument that was ultimately rejected cannot claim that they have *no* reason to reject the decision as unjustified, since the decision is supported by a value they share and recognise as appropriate grounds for action in the public sphere. For a decision to be considered justified to a person, said decision does not have be the one that best fits her perception of priorities among shared values in any given case. In other words, for it to be publicly justified, the decision does not have to be the one she would have considered *most* justified. Hence, under the conditions set by the internal conception, reasonable disagreement is no obstacle to the legitimate exercise of coercive power in the public sphere.

---

[46]Rawls, *Political Liberalism*, p. 137.
[47]Quong, *Liberalism without Perfection*, p. 41.
[48]Ibid., p. 192.
[49]Ibid., p. 142.
[50]See ibid., pp. 205-207.

And yet, the internal conception crucially assumes that among reasonable citizens, all disagreement on matters of public policy arises from them merely assigning different priorities to the relevant shared values, and can thus be resolved with reference to the sharedness of public reasons alone. This assumption overlooks the potential complexity of their overall set of reasons and beliefs, including their comprehensive moral doctrines. Given that for many individuals the values that are publicly shared do not exhaust their private sets of reasons and beliefs, it is not inconceivable that citizens' sincere commitment to particular shared, public values may sometimes come into conflict with private moral convictions they also hold. This may be the case despite the fact that those citizens recognise that their nonpublic reasons are inadmissible as arguments in public justification due to their controversial status in the public political sphere (*i.e.,* the fact that they are not subject to an overlapping consensus). After all, the material scope of an individual's private convictions which are rooted in her comprehensive doctrine is not restricted to issues outside the scope of the political. For instance, a person's religious convictions may recommend to her a different position on policies on issues like abortion or genetic engineering from the one that derives from her public reasons. The internal conception assumes that, as a reasonable person, she is committed accepting the consequences of the burdens of judgement and thus realises that she cannot expect others to embrace the reasons rooted in her private convictions. Her reasonableness is assumed to prevent her from demanding state-sanctioned action based merely on the content of her private comprehensive moral doctrines, since she has no reason to assume that others can equally recognise them as sources of valid reasons.

Reasonable people, it must thus be assumed, will always prioritise their commitment to public reasons over their potentially conflicting nonpublic ones. However, as I will argue in the following section, this assumption idealises reasonable citizens in a way that its incompatible with the core liberal commitment to respecting the moral autonomy of persons. If we take that commitment seriously, we must acknowledge that individuals may be genuinely torn between the demands of their public and nonpublic reasons. Depending on the status a person assigns to the private values that conflict with her public reasons, the former may not be simply be overridden by the latter, leaving her in a state of cognitive dissonance with regard to the question of how to act in the public political sphere. The following subsection is concerned with showing that such instances of cognitive dissonance are the result of the exercise of individual moral autonomy which a liberal theory must not ignore.

2.3.2 Lack of respect for the moral autonomy of citizens

Quong defends the asymmetric treatment of disagreements about political questions of justice and questions of the good life. He does so by discerning between justificatory disagreement and foundational disagreement. The former case is characterised by the existence of

shared premises which can serve as the bases of justification,[51] while in the latter case such shared premises are unavailable. As explained in the previous section, the existence of such shared premises is thought to ensure that whichever policy is ultimately enacted, it is justi-fied to all, thus rendering state interference legitimate with regard to questions of justice. The absence of such shared premises in the case of disagreements about the good life means that state action cannot be expected to be justifiable to all, thus restricting its legitimate scope to the realm of political questions of justice. Importantly, for my present argument, Quong as-serts that reasonable disagreements about justice are *always* justificatory in nature.[52] In other words, the only reasons a reasonable person deems to be relevant to questions of justice are public reasons.

At the same time, Quong concedes that "disagreements about justice are frequently just as profound as disagreements about the good life."[53] But that does not, in Quong's view, render them foundational. Individual citizens may indeed firmly defend their positions by means of a particular public reason, while heavily disputing the relevance of other public reasons to the case in question. But among reasonable citizens, such disagreement is rooted in shared reasons alone, no matter its depth. Nonpublic reasons do not enter into individuals' assessment of questions of justice.

Quong's highly idealised conception of the reasonable person whose position on questions of justice cannot, by definition, draw on nonpublic reasons rooted in her comprehensive moral doctrine is a far cry from the way in which we would ordinarily expect individuals to relate to matters of justice, which ultimately remain moral questions. He assumes a rather simplistic view of the nature of individuals' understanding of the relation between their publicly shared values on the one hand, and their comprehensive moral doctrines on the other. Once a value is recognised as a public reason, it is assumed that it will necessarily supervene on any nonpublic reasons. It is also assumed that a person's comprehensive moral views will not at all affect her assessment of whether a particular shared value constitutes a valid source of reasons with regard to a particular question. Consider, for example, disagreements on legislation which hinge upon the appropriate definitions of the beginning or end of life (*e.g.* legislation concerning abortion rights, genetic engineering, or organ donation). A person whose comprehensive doctrine leads her to believe that life begins at conception may well have a different view of the strength of shared reasons drawing on values such as women's bodily autonomy, prolonging life or alleviating suffering compared to people with no or different beliefs on that matter. At worst, her privately held beliefs may shed doubt on whether or not the position to which she owes allegiance qua her sincere commitment to her public reasons is truly justifiable to her in this case. In any case, she is bound to experience

---

[51]More precisely, justificatory disagreement is characterised by the fact that the disagreeing parties "share a justificatory framework" (Quong, *Liberalism without Perfection*, p. 207) which is defined as follows: "*(a)* all the parties must be sincere, *(b)* the conflicting positions must be grounded in free standing political values, and *(c)* the conflicting arguments must represent a plausible balance of political values" (ibid., p. 207).

[52]See ibid., p. 193.

[53]Ibid., p. 212.

some degree of cognitive dissonance when considering the demands of her comprehensive moral doctrine versus those of the set of publicly shared values.

This complexity of her moral reasoning is not acknowledged within the internal conception of political liberalism. According to Quong, the right response to a person whose private comprehensive doctrines lead her to consider a pro-life position as in the example above is to ask whether she

> is being offered reasons she can reasonably be expected to endorse. The answer is clearly affirmative. She may not believe that the right to control your own body outweighs the foetus' right to life, but she cannot plausibly say that the pro-choice position is an unreasonable one. Most pro-choice arguments appeal to political values that she could be reasonably expected to endorse, such as the right to control your own body [...].[54]

But for a person who encounters the kind of moral experience I described above, this response only begs the question. There is no doubt that she is committed to the respective shared political values. Yet these values have lost the very normative force Quong assumes them to have, precisely because, from the perspective of the person in question, their validity is affected by moral commitments rooted within her comprehensive doctrine.

Quong does not seriously consider this scenario. That said, the objective of the internal conception of political liberalism does not seem to require addressing this kind of conflict. It does not need to concern itself with the question of whether and why citizens would always be required prioritise their commitments to reason only on the basis of shared, public values. This is because it is only concerned with persons who are defined by the very fact that they are prepared to honour said commitments. The internal conception merely aims to explore the content and structure of public justification within this specifically idealised constituency.[55] In this light, it is not surprising that Quong's discussion of justificatory disagreement does not provide a more nuanced account of how individuals in a less idealised context navigate the relations between their public and their nonpublic reasons.

However, I will argue that the failure to account for these individual epistemic dilemmas introduces another defect into political liberalism. Again, it is the liberal commitment to persons as morally autonomous beings capable of self-determination that provides a principled argument to challenge the assumption that it is sufficient to define a reasonable citizen solely in terms of their disposition to prioritise shared values over individual private convictions: from a liberal perspective, it is not unreasonable to ask whether the ideal of honouring any person's moral autonomy should not be reflected in the model of the ideal reasonable citizen who populates the constituency of public justification of the internal conception of political liberalism.

---

[54]Ibid., p. 213.
[55]See ibid., p. 153.

Quong frames the ambition of the internal conception as follows: by "showing that liberal theory is internally coherent [...] the internal conception offers important help with regard to one of political philosophy's main roles: probing the limits of practical possibility."[56] The fact that the internal conception strives to test whether liberal theory is viable in what may be considered the most favourable of circumstances may legitimately account for some degree of abstraction and idealisation. However, once this idealised experimental set-up and the model individuals interacting within it no longer remotely resemble the way in which liberalism usually conceives of persons, it is questionable why we should believe that the conclusions drawn from it are at all relevant to the study of liberal theory. If the internal conception is ultimately committed to the *practical* aim of developing *liberal* principles, it must shape the individuals who populate its idealised constituency of public justification in a way that does not conflict with a liberal conception of the person.

A key component of such a conception, as I have argued earlier in this chapter, is individual moral autonomy. A theory which is committed to protecting said autonomy cannot just conceive of its ideally reasonable citizens as static and free-standing entities whose reasonableness is neither rooted in nor involved in an ongoing process of moral deliberation. Even if individuals are ideally reasonable – in the sense the internal conception requires them to be – they are reasonable for a reason, or a variety of reasons. This may sound trivial, yet it is not. Their reasonableness is the result of the exercise of their moral autonomy: in order for us as external observers to conclude that the norms governing the constituency of public justification are acceptable to them, they must have been able to freely choose to endorse them in the first place. But any such theory that pays respect to individuals' moral autonomy in this regard must also allow for the fact that even after entering the constituency of public justification, they will continue to reflect on the standards for the justification of political coercion in the light of the comprehensive moral doctrines they hold. Hence, individuals who exercise their moral autonomy will disagree with Rawls's claim that "[t]he express contents of these doctrines have no normative role in public justification."[57] There is no reason why we should assume that a realistically conceived, morally autonomous person would draw on these doctrines only once in order to assess whether the demands of reasonableness are justified in light of their comprehensive moral commitments. But that also means that we cannot disregard and discount how all of their other substantive content affects a person's perspective on the entirety of moral reasoning that happens in the political realm, including the validity and relevance of public reasons in a given situation. It is the *possibility* of such an ongoing process of moral reasoning in which a person's moral autonomy manifests itself. A theory of public justification that perceives the need to curtail it or deny its necessity deprives its subjects of the capacity to exercise said autonomy.

Thus the argument put forward by the internal conception of political liberalism, *viz.,* that a liberal theory with a suitably limited scope may be framed around highly idealised reasonable

---

[56]Quong, *Liberalism without Perfection*, p. 138.

[57]John Rawls. "Political liberalism: reply to Habermas". In: *The Journal of Philosophy* 92.3 (1995), p. 144.

citizens, can only succeed at the expense of depriving these hypothetical citizens of any moral autonomy. Such a move, however, is testament to the fact that the internal conception does not perceive persons in a truly liberal way, *i.e.,* as ends in themselves, fully capable of and entitled to exercising their moral autonomy.[58] Thus, the internal conception of political liberalism fails to live up to one of the central moral commitments of liberal theory.

2.3.3 Failure to account for moral pluralism

As I have argued so far in this chapter, the internal conception's reductive account of the person masks the complexity of real citizens' relation to their various moral commitments – specifically their comprehensive moral doctrines and the political elements within them. As a result, the idealised constituency of public justification that forms the basis of the internal conception does not reflect the diversity of actual modern pluralistic societies. One may argue that the internal conception does not aspire to reflect that diversity, but instead aims to explore what political principles could be justified in an ideally liberal society, as Quong does.[59] However, it would be a mistake to accept that framing of the liberal project unless one were also prepared to accept that the conclusions drawn from it are likewise restricted to societies which are populated by citizens resembling the ideally reasonable model. It is not the case that once we have worked out the political principles which apply to such a model society, we can take them as a basic blueprint for less idealised societies without further argument. This is because the challenges arising under more realistic conditions of moral pluralism fundamentally affect the process of public reason that furnishes these principles with legitimacy: for those citizens who experience the kind of cognitive dissonance which has been the subject of the previous sections, the priority of shared, public reasons is not necessarily always obviously mandated by their comprehensive doctrines. Actual comprehensive doctrines present in modern pluralistic societies do not necessarily operate as political liberalism expects them to operate.

It is possible to argue in response that individuals who are broadly committed to core liberal principles may have reason to prioritise public reasons and should therefore amend their comprehensive doctrines in that respect. In fact, this is precisely the kind of argument which I will propose in the further course of this thesis. But it is crucial to note that a significant argumentative step is required if we want to claim that political principles which are justified in an idealised constituency of public justification may also claim validity in less idealised pluralistic societies. This is a step which must not be avoided if the political principles which are ultimately recommended by political liberalism are to be relevant to the political and moral context of modern pluralistic societies. A theory which has its eyes set on addressing the challenges of moral pluralism in liberal societies cannot ignore this.

---

[58] As discussed in section 2.2.1.

[59] See Quong, *Liberalism without Perfection*, pp. 6, 143.

At least the original Rawlsian project of political liberalism is committed to devising princi-
ples for pluralistic societies in more than the extremely narrow sense of a pluralism of highly
idealised moral doctrines. This is not to disregard the fact that Rawls clearly envisages politi-
cal principles to be supported not by all comprehensive doctrines, but merely by those which
are reasonable. Throughout *Political Liberalism,* Rawls emphasises that "the problem of po-
litical liberalism is to work out a political conception of political justice [...] that a plurality
of *reasonable* doctrines [...] may endorse".[60] Nevertheless, his project is clearly motivated
by a concern about conflicts between actual moral doctrines in the world we live in:

> the historical origin of political liberalism (and of liberalism more generally)
> is the Reformation and its aftermath, with the long controversies over religious
> toleration in the sixteenth and seventeenth centuries. [...] [T]he fact of religious
> division remains. For this reason, political liberalism assumes the fact of reason-
> able pluralism of comprehensive doctrines, including both religious and nonre-
> ligious doctrines. This pluralism is not seen as disaster but rather as the natural
> outcome of the activities of human reason under enduring free institutions.[61]

The comprehensive doctrines populating the pluralistic societies that Rawls is concerned
with are no artificial constructs. They are the very doctrines we encounter in daily life,
generated by "the activities of human reason under enduring free institutions."[62] Rawls's
assumption of the fact of *reasonable* pluralism is not equivalent to an assumption of an ide-
alised version of these doctrines. He simply assumes that many of the doctrines present in
a free society meet the criteria of reasonableness. The upshot of this is that the political
principles determined within political liberalism ought to be seen as principles that strive
to accommodate a pluralism of comprehensive moral doctrines as found in free societies.
Therefore, it matters if the moral doctrines present in the idealised constituency of justifica-
tion differ from the latter in such a way that the conditions which lend support to political
principles in the idealised constituency cannot be assumed to obtain in actual pluralistic so-
cieties.

Quong's position explicitly rejects the idea that political liberalism should accommodate the
actual kind of pluralism which is produced by free societies. He argues that political liberal-
ism only aspires "to clarify what kinds of reasons *liberals* can offer to one another".[63] In this
vein, the aim of political liberalism "is to understand how the fact of reasonable pluralism
generated by liberal principles and institutions constrains the kinds of political arguments
that liberal citizens can coherently offer to one another, and thereby constrains the kind of
policies that can be legitimately imposed in liberal democratic societies."[64] But Quong is

---

[60]Rawls, *Political Liberalism*, xxxix, my emphasis.
[61]Ibid., p. xxiv.
[62]Ibid., p. xxiv.
[63]Quong, *Liberalism without Perfection*, 6, my emphasis.
[64]Ibid., p. 6.

wrong to assume that the pluralism generated by liberal principles and institutions is necessarily of the ideally reasonable kind that he imagines. He agrees with Rawls that reasonable pluralism is "*a fact about liberalism.*"[65] Like Rawls, he argues that "disagreement between rational and well-intentioned people about morality, religion, and other matters concerning human flourishing is to be expected, indeed is inevitable, within a liberal society."[66] It is unclear why we should assume that these circumstances may not also produce instances of cognitive dissonance of the kind I described in section 2.3.1. After all, it is likely that "allowing citizens to think, speak, and associate freely"[67] may result in some people affirming comprehensive moral doctrines whose norms do not align precisely with the demands of reasonableness at all times. Therefore, Quong's presentation of the circumstances which produce reasonable pluralism could be taken to commit him to a more nuanced view of such pluralism and its constituent moral doctrines than he actually affirms.

But even if we grant that such a commitment is not part of Quong's internal conception of political liberalism, it opens itself to criticism precisely on that account. Idealising persons to the extent that we do not expect their comprehensive moral doctrines to possess the complexity and potential for internal conflict described in the previous section is simply an illiberal thing to do. Taking the liberal commitment to respecting individuals' moral autonomy seriously is incompatible with a degree of idealisation in which any complexity of comprehensive moral doctrines that results from individuals' use of their moral autonomy under free institutions is absent. This is not a pluralism that would arise among persons as liberalism conceives them. To frame reasonable pluralism as narrowly as the internal conception of political liberalism is in conflict with the liberal conception of the person as such. So even if Quong's internal conception seems to reject a more nuanced view of the reasonable pluralism present in liberal societies, we can argue that, as a liberal theory, it should not do so.

One may object that I seem to be criticising the internal conception of political liberalism for failing to live up to a standard it never aspired to fulfil. However, I want to stress that we do not need to be beholden to the internal conception's interpretation of the kind of pluralism it is concerned with in order to argue that it is not an internally consistent liberal theory. We can say that, for a liberal theory, it is wrong to frame its ambitions the way the internal conception does. We should therefore interpret said ambition, that is, accommodating the pluralism generated by liberal institutions, in the more expansive sense that I proposed, and which I believe is actually implicit in a consistent interpretation of Rawls's and Quong's rendering of the origins of liberal pluralism. As a theory which is motivated by the fact of (reasonable) pluralism produced by free societies, political liberalism should at least have the ambition to engage with and strive to accommodate the doctrines that define said pluralism in their full complexity.

---

[65]Ibid., 6, original emphasis.
[66]Ibid., p. 6.
[67]Ibid., p. 6.

In conclusion, it is thus possible to state that the internal conception of political liberalism is defective in the following sense: its conception of the person is inconsistent with liberal commitments, as is – as a result – its view of the pluralism produced by liberal societies. The internal conception is therefore an unsatisfying answer to the question raised in the first half of this chapter: that of whether political liberalism needs to address justifications to a constituency which extends beyond reasonable people, specifically justifications for henceforth excluding those who are unreasonable from the constituency of public justification. If we want political liberalism to live up to core liberal ideas about the person, we must look for means other than idealisation to satisfy the requirement of justifying the standard of reasonableness to those subject to its coercive potential.

2.4   Conclusion

In this chapter, I made the case that the core liberal commitment to persons as ends in themselves who are free and equal, endowed with moral autonomy and a capacity for self-determination establishes the need to justify political principles *to* those individuals who might be coerced by them. I argued that restricting the constituency of public justification to reasonable persons ultimately constitutes a coercive act with respect to those who are excluded from it – that is, the unreasonable. This is because it deprives them of their entitlement to be provided with justifications for other potentially coercive policies. However, justifications for the standard of reasonableness in Rawlsian political liberalism are self-referential: reasonable persons have reason to accept the requirements of reasonableness. Those who are unreasonable are beyond the scope of liberal concern, given that political liberalism seeks to establish political principles for societies which are ultimately built upon liberal values. But this argumentative strategy, I argued, is in conflict with the key liberal commitment to treating *all* persons as ends in themselves. A political liberalism which does not provide justifications to all on the initial level of determining who should henceforth be entitled to justifications fails to live up to its own core values. This is a significant defect.

Quong's interpretation of the project of political liberalism remedies this defect at the cost of introducing another. Within his internal conception, justifications do not need to be provided to all, given that the ambition of political liberalism as he sees it is to identify what kind of reasons ideally reasonable people can offer each other in their search for political principles. However, this approach is in tension with the liberal conception of the person that crucially emphasises individuals' endowment with moral autonomy. We are thus left with the question of how political liberalism can ensure that it lives up to this ideal while also upholding reasonableness as the criterion for membership in the constituency of public justification. How can it accommodate a realistic rather than over-idealised and therefore impoverished liberal conception of the person and the pluralism generated in a society populated by such persons?

The answer does not consist in imagining that individuals do not have to navigate conflicting moral demands between their liberal commitments and other elements of their comprehensive moral doctrines. It consists in exploring if, and on which basis, individuals who do experience these conflicts might be prepared to act as an ideally reasonable person is expected to act. In other words, we must look for a way to justify the requirements of reasonableness to individuals in light of their agency and genuine commitment to their comprehensive moral doctrines. While remaining mindful of the diversity of these commitments between different people, we will only be able to get a general philosophical argument off the ground if we look for justifications in things they do share. Hence, we must look for them in traits whose prevalence among individuals can reasonably be assumed to be maximal, if not universal. Traits that I believe to be promising candidates in this regard are human rationality, individuals' beliefs implicit in their actions within the public political sphere, and their possession of a comprehensive moral doctrine as such. In the remainder of this dissertation, I will argue that these traits can provide individuals with reasons to act as reasonableness requires when confronting others with their political demands, thus delivering a non-circular justification of the status of reasonableness within liberal theory.

In this chapter, I established the requirement to justify reasonableness for the sake of preserving internal consistency within political liberalism, ensuring that it lives up to its commitment to the basic liberal value of respect for persons as ends in themselves and associated ideals such as the preservation of their moral autonomy. I will propose two arguments justifying reasonableness to the unreasonable, in chapters 4 and 6. But first, we must explore the standards such a procedure of justification must itself meet in order for it to live up to the core ideal of respecting persons as ends. This will be the subject of the following chapter.

Chapter 3

A liberal conception of justification

3.1   Introduction

In the previous chapter, I argued that political liberalism must justify the choice of reasonableness as a criterion for regulating the access to the constituency of public justification to all people affected by it. For any such attempt, it is, however, imperative to avoid succumbing to the same fallacies as the approaches I criticised in the previous chapter. In particular, a conception of justification that we can draw on in order to address the defects in political liberalism which ultimately arise from it failing to live up to the basic liberal commitment of treating persons in a way that respects them as ends in themselves must itself satisfy that very standard. In this chapter, I intend to develop and defend a conception of justification that stays true to this distinctly liberal ideal.

This chapter is concerned with the character of the reasons which we may draw on when arguing that citizens *have reason* to accept reasonableness as the criterion for regulating access to the constituency of public justification. As these reasons are being attributed externally, it is important to ask about the standard for judging when such reasons can actually be considered to apply to a person and thus be a reason *for her*. In other words, what criteria, both in terms of the sources and the standards of inference, does an alleged reason need to fulfil for us to be able to conclude that an individual can actually be said to *have* a reason? Does it need to correspond to facts which may be said to obtain objectively? For instance, can citizens be said to have reason to be reasonable in the political sphere, because *it is the case* that reasonableness is the only appropriate attitude in this context? Or must the reasons a person is supposed to have bear some relation to a person's internal mental constitution? In the latter case, we might only be justified in coercively excluding people from the constituency of public justification if the reasons we offer them can be shown to be supported by other beliefs they hold.

Justification, as I understand it in the context of validating reasonableness as an appropriate entry criterion to the constituency of public justification, is a process of attribution. More

specifically, I conceive of it as a process of attributing reasons to individuals who are in danger of being framed as unreasonable – reasons for them to recognise the norms implicit in the ideal of the reasonable person as legitimately authoritative in the context of a liberal democracy. The attribution of reasons in the context of justifying reasonableness is ultimately a matter of external judgement. As an attempt to preserve the internal theoretical consistency of political liberalism, it does not seek the actual approval of real people and does not require their judgement. I do not propose a referendum on reasonableness, but a theoretical analysis of its legitimacy within liberal theory, thus merely emulating individuals' judgement. In the absence of a person's granting or refusing of consent as a means of conferring their authority, the emulation of judgement requires standards for deriving a person's choice of acceptable reasons from a set of (interrelated) cognitions, such as her ends, beliefs, or desires. For the sake of theoretical consistency, it is crucial that the standards governing the way in which the attribution of reasons for reasonableness may legitimately proceed – and thus determining which reasons we are, as a result, warranted to attribute others – must reflect the ideal whose protection gave rise to the need to justify reasonableness in the first place: the respect for persons as ends in themselves.

The first part of this chapter is dedicated to an analysis of the ideal of respecting persons as ends in themselves. I argue that there are specific dimensions of this ideal whose protection is implicit in the ideal of justification itself. Drawing on these insights, section 3.2.1 develops a conception of persons as ends which can serve as a point of reference for identifying appropriate standards for externally attributing justifications which do not infringe upon the very values that justification implicitly protects. Crucially, I will argue in section 3.2.2 that it is implicit in a conception of respecting persons as ends in themselves in the process of justification to ascribe to them a reflective attitude towards their ends. In section 3.2.3, I will consider a possible objection to my proposed account of the respect for persons as ends in themselves in justification, namely, that it cannot not ensure that the actions of an attributor of reasons are morally right. I will finally proceed to draw on the conception of persons as ends in themselves proposed in the first part of this chapter as a standard for identifying both suitable sources and epistemic standards for the process of justification in sections 3.3.1 to 3.3.4.

3.2 Justification and respect for persons as ends

3.2.1 Respect for persons as respect for self-authorship

If we conceive of justification as a moral ideal and if we approach the question of what kind of reasons we must offer each other in the domain of the political from a liberal perspective, we are prompted to ask what the appropriate liberal principle to determine the standards for processes of justification might be. At the beginning of the previous chapter, I argued that

liberalism in general, and the justificatory branch of liberalism in particular, is ultimately motivated by a respect for individual persons as ends in themselves.[1] I argued that in order to live up to its core commitments and to treat *all* persons as ends, liberalism owes justifications to those it excludes from its constituency of public justification: it owes to them justifications for their exclusion, which amount to justifications for drawing on reasonableness as the criterion for access to the constituency of public justification. If political liberalism is crucially committed to respecting persons as ends, surely the process of justification, of proposing justifying reasons to others, must at least live up to this very commitment. In other words, liberals who are committed to treating others as ends in themselves and who, as I argued, must therefore re-examine their conduct towards those who have so far not been offered justifications, must apply a no less demanding standard to their conduct towards members of the latter group when addressing them with justifications for their exclusion, lest they violate the principle that compelled them to address them with justifications in the first place.

In this section, I intend to elaborate on the relation between the ideal of justification and the respect for persons as ends, developing a specific conception of persons as ends in themselves as the appropriate guiding principle for determining the standards of a liberal conception of justification. In particular, I intend to highlight a specific dimension of the ideal of persons as ends which renders the requirement of justification so vital to the achievement of respect for such persons: autonomy as self-authorship. To that end, I will discuss and contrast different related liberal ideals which are arguably derivative and supportive of the ideal of respecting persons as ends, approaching them with the guiding question of whether they would require acts of (potential) coercion to be justified to the affected individuals. The rationale behind this explorative strategy is as follows: if treating others as ends requires us to provide them with justifications for their coercion, we should be aware of the element within said ideal that gives rise to the need for justifications in the first place, in order to ensure that the standards we apply to the process of justification preserve that specific element. This is to ensure that any future argument showing that reasonableness as the criterion regulating access to the constituency of public justification can be justified to all people is not undermined by the very standards of justification applied in the attribution of reasons to them.

3.2.1.1 Non-interference

The Kantian origin of the moral ideal of respect for persons as formulated in the *Groundwork of the Metaphysics of Morals* demands "that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means".[2] In conjunction with Kant's definition of a *mere means* as something that exists "for the discretionary use for this or that will",[3] the ideal can be interpreted as affirming the agency

---

[1]See section 2.2.1.

[2]Immanuel Kant. *Groundwork of the Metaphysics of Morals*. Ed. by Mary Gregor and Jens Timmermann. Cambridge Texts in the History of Philosophy. Cambridge: Cambridge University Press, 2012, 4:429.

[3]Ibid., 4:428.

of individual persons.[4] Not to use someone at one's discretion – or, that is, at the discretion of some third party – is to accord that discretion to the person in question. It is thus upon her to decide whether or not to consider her treatment by others acceptable and what treatment to subject herself to in the pursuit of her own ends.

The most straightforward way of ensuring that she is able to act solely at her own discretion is to ensure a maximal amount (negative) freedom, granting her the largest possible realm of individual self-determination compatible with the same provision for all, and prosecuting any intrusion in or external interference with said realm. A political theory that assumes that respect for persons as ends in themselves merely amounts to non-interference (as far as possible) with the pursuit of their ends does not require justificatory procedures. This is not a critique of theories centred around the value of negative freedom emphasising the significance of non-interference. As I highlighted at the beginning of this section, it is not my concern at this point to defend justificatory interpretations of liberal theory. The purpose of my analysis merely consists in uncovering which dimensions and interpretations of the ideal of persons as ends is of significance to and necessitates acts of justification. The interpretation of respect for persons as ends that amounts to preventing them from being treated as mere means by protecting their discretion to act – through limiting interference with their actions – does not require us to address individuals with reasons. It merely requires that they are not constrained in their capacity to act as they see fit.

Thus, if we assume that respect for persons as ends in themselves merely requires non-interference, there is no need for liberal theory to consider offering justifications to individuals in order to ensure that it is respectful of their status as ends in themselves. If respecting persons as ends in themselves is merely conceived as the need to ensure that they are not treated as mere means, thus according to them the sole discretion of how to act and be acted upon, a conception of justification does not need to be part of a political theory committed to that purpose.

3.2.1.2   Respect as respect for ends

Let us now turn to an interpretation of the ideal of treating of persons as ends which is more substantive than the command that they must not be treated as mere means. The latter focuses on protecting a person's agency itself, rather than the objects of that agency, that is, a person's individual ends. In according primacy to the procedural dimension, doctrines of non-intervention discount the value of the substance of a specific person's ends. The possibility that her ends may, for instance, not necessarily be limited to those that can be realised alone or by means of private cooperation between individuals, or that she might pursue moral ends which would require more extensive intervention in others' private realms of self-determination, appears to be less important than the imperative that she must not be

---

[4]See Onora O'Neill. "Kant: Rationality as Practical Reason". In: *The Oxford Handbook of Rationality*. Ed. by Piers Rawling and Alfred R. Mele. Oxford: Oxford University Press, 2004, pp. 93–109, p. 106.

used at others' discretion for the pursuit of their ends. Again, this is not to be interpreted as an argument against the conception of respect as non-interference, but rather as an illustration of ways in which it contrasts with other conceptions of persons as ends that may eventually be of relevance to justification.

Interpretations of the respect for persons as ends in themselves which emphasise not only its procedural but also its substantive dimension, thus attributing some value to the objects of individual persons' agency, do not need to relax the condition that persons must not be used as mere means. Maximising the realm in which agency can be realised without interference, thus ensuring that, within said realm, a person cannot legitimately become subject to another person's pursuit of her aims at all, is not the only way of ensuring that she is not treated as a mere means to the fulfilment of others' ends. Preventing others from disrespecting both her agency as such *and* the objects of her agency may be achieved by procedures that ensure that all coercive action is legitimate only if it (ultimately) also satisfies – or at least supports and protects – her pursuit of her ends. This ensures that her (potential) coercion in the name of a particular policy does not turn her into a *mere* means to the fulfilment of others' ends, all the while its meaning is not exhausted by defending the individual in question against the latter. Rather, the focus on protecting and fostering the pursuit of individual persons' ends (or, in alternative terminology, preferences or goods) conceives of legitimate coercion as a force for protecting what a person *has determined* to be of value to her, and thus, by proxy, respecting her agency.

Martha Nussbaum gives expression to this idea – though framed in terms of meaning rather than agency – in highlighting how the absence of respect for persons' chosen doctrines of the good life translates into a lack of respect for persons themselves.

> Of course it remains the case that respect is for persons, not for their doctrines. But these doctrines are so deeply a part of people's search for the meaning of life that public governmental denigration of those doctrines puts those people at a disadvantage, suggesting that they are less worthy than other citizens, and, in effect, not treating them as fully equal ends in themselves.[5]

Non-perfectionist liberal theories can also be interpreted as being committed to respecting persons as ends through the satisfaction or facilitation of their ends. Respect for citizens' individual conceptions of the good life expresses equal respect for all citizens by treating their diverse and particular ends with equal concern in virtue of their quality as products of citizens' exercise of their autonomy,[6] thus recognising the significance of the latter for respecting the equal status of each citizen.

---

[5]Martha C. Nussbaum. "Perfectionist Liberalism and Political Liberalism". In: *Philosophy & Public Affairs* 39.1 (2011), p. 22.

[6]See, for example, Will Kymlicka's suggestion that "no life goes better by being led from the outside according to values the person doesn't endorse. My life only goes better if I'm leading it from the inside, according to my beliefs about value." (Will Kymlicka. *Liberalism, Community, and Culture*. Oxford: Oxford University Press, 1989, p. 12) Similarly, Lawrence Haworth asks "what is there to value in a community of shared

Such theories ultimately – though not necessarily consciously or explicitly – draw on the Kantian notion that a person's status as an end in itself is tied to her capacity to determine her own ends and act according to reason.[7] I do not intend to engage in an exegesis of Kant's account of the relation between human reason and the status of human beings in ethical theory. The account of respect for persons as ends in themselves which I rely on – which is, like other contemporary accounts of liberal political theory, committed to the respect for persons as its basic value – is not Kant's but merely Kantian in its emphasis on the objective worth of the individual as a being conceiving of herself as capable of thinking and acting according to reasons.[8]

We do not have to engage in a debate over why human beings' capacity to set ends according to reason is of significance to their status as ends in themselves in order to conclude that a conception of that status which disregarded their ends would be void. What we are presented with in inter-subjective interaction with a person are expressions of her agency. Any conscious and voluntary action communicates her ends broadly defined (including her preferences and desires). In this regard, respecting a person amounts to respecting what she does, intends, desires or, more generally, wills, since being confronted with a person, apart from the physical dimension, ultimately amounts to being confronted with her ends.[9] Respecting a person by respecting her ends is to respect her as the *author* of her ends.[10]

It certainly does not follow that any end set by a person is of (equal) objective worth just because it has been devised by a being who we are committed to treating as an objective end. Neither is any end a person may want to be realised by political means a legitimate object of political power,[11] nor are all legitimately political ends of equal importance.[12] Nevertheless, enabling or satisfying a person's political ends as far as possible given other citizens' ends,

---

values when the members of the community are automata? The fact of sharing or of being mutually devoted to a transcendent or collective good commands respect only to the extent that the members of the community participate autonomously – to the extent that each [...] is living his own life, pursuing his (procedurally) own conception of the good". (Lawrence Haworth. *Autonomy: An Essay in Philosophical Psychology and Ethics*. New Haven: Yale University Press, 1986, p. 208).

[7]For an argument linking human reason and the mere capacity of setting ends to the status of the end in itself, see Allen W. Wood. *Kant's Ethical Thought*. Cambridge: Cambridge University Press, 1999, p. 121.

[8]See, for example, Larmore's discussion of detaching his defence of political neutrality from a comprehensive commitment to Kant's ideal of autonomy in Larmore, *The Morals of Modernity*, pp. 138–139.

[9]See, for example, Stanley Benn's characterisation of an agent's projects as "an exteriorization of himself, projections, indeed, of himself into the world; his identity as a person, qualifying for respect not only from others but also from himself, depends on his sense that they are indeed his own [...] One may believe the other's project quite worthless in itself. Its claim to respect rests not on its being valuable and worthy of one's concern [...] but simply on its being a person's project". (Stanley I. Benn. *A Theory of Freedom*. Cambridge: Cambridge University Press, 1988, p. 107).

[10]This conception may be interpreted as the antithesis of Raz's perfectionist proposition that the value of a person's autonomy is a function of the value of her autonomous choices (see Joseph Raz. *The Morality of Freedom*. Oxford: Oxford University Press, 1986, p. 417). On the contrary, according to an account which conceives of the respect for persons in terms of respect for their authorship of their ends and, more generally, choices, a person's choice possesses value in virtue of the value that is attributed to a person's autonomy (or self-authorship).

[11]Consider, for example, Alice's end to become a more accomplished player of the Ukulele than Bob.

[12]Compare, for example, Alice's desire to direct more public funds at improving the situation of the homeless in her city with Bob's campaign for re-landscaping the city's public parks.

as well as physical or economic constraints may be a measure of the legitimacy of a political process that is rooted in a commitment to respecting persons as ends.

Do political theories that draw on some conception of respect for persons as the satisfaction of their ends or preferences *have* to rely on justificatory procedures? The assumption that they do is not far-fetched. After all, to acknowledge that a person's potential coercion is admissible because it is executed in the name of an end which is an end *for* that very person can be seen as equivalent to acknowledging that said person has a reason to accept the respective policy. In other words, it would seem that a policy is justifiable to a person in virtue of its ties to some of her ends or preferences which act as reasons for her to endorse it.

However, the conception of justification which those political theories that define respect for persons as ends in terms of the satisfaction of their ends could be found to (implicitly) rely on is shallow at best. At no point does this process require the individual in question to be *addressed,* that is, to be *presented with* the proposal which is supposedly justifiable to her in virtue of its capacity to satisfy her ends. It does not require her to be asked for her active judgement. It is her specific ends that are addressed, rather than the person herself. At all times, this process may remain external to the individual.

One could object to this diagnosis by arguing that all those justificatory theories which do not insist on a requirement of actual, empirical consent must ultimately resort to judging a proposal's justifiability to an individual from an external perspective, merely assessing whether it satisfies the ends and preferences exhibited by, or ascribed to, the respective individual. To some degree, one could argue, a justificatory theory is no more able to address the recipients of justifications, and can much less rely on *their* judgement, than a theory which merely requires their ends and preferences to be satisfied. In other words, since any theory is unable to replicate actual judgement, justificatory processes must ultimately rely on, and could even be said to collapse into, the mere satisfaction of citizens' ends.

However, this objection does not fairly represent genuine efforts of justification in theory. While unable to emulate individuals' actual judgement, a justificatory theory is able to enter a hypothetical dialogue with an imagined interlocutor to whom justifications are being presented. Rather than taking that person's ends and preferences at face value, it engages with them, assuming that they can – but *do not necessarily* – represent that person's final response to any proposition she is presented with.[13] By contrast, a theory which regards the satisfaction of her current ends and preferences as sufficient for respecting her as an end in herself does not require an elaborate process of justification that enters a (hypothetical) dialogue with said person. It merely needs to "read" her ends and preferences.

---

[13]If a person's ends were assumed to exhaust her agency and the mere satisfaction of such an end were sufficient to ascribe to a person a reason for accepting a given proposal, we would be forced to discount the potentially complex interactions between her different ends with regard to a given proposal: that a person's end is satisfied by the proposal in question does not imply that she would have authorised it. She might well possess some other end that defeats or, at least, diminishes the value of the former in this specific context, thus rendering her judgement far less certain than any isolated assessment of her ends might suggest.

3.2.1.3 Self-authorship

A shallow conception of justification that conceives of a person's reasons in terms of her actual ends, thus emphasising the outcome rather than the act of judgement, would only exhaust the ideal of justification if we equated respect for persons as authors of their ends with merely respecting the outcomes of their self-authorship.[14] Just as theories of non-interference may be interpreted as relying on a conception of respect for persons as ends which is exhausted by the prescription not to treat them as mere means, the conception emphasising the mere satisfaction of ends assumes that persons as ends are exhausted by the ends they exhibit, pursue, or want to see protected in the political process. Justification, as I have understood the ideal so far, goes beyond drawing on a person's expression of ends, but is instead focused on and intends to preserve her agency in the process of *judging* which ends are sufficient to warrant her coercion. By putting emphasis on the process of addressing a person, an act of justification *asks* for her judgement on a specific matter as opposed to merely inferring to it on the basis of her proclaimed ends. It is implicit in a justificatory procedure so understood that she must authorise *the proposed policy* – thus asking for an active expression of agency with reference to the proposal– rather than merely pointing to a set of ends whose proposed satisfaction might warrant the assumption that she might consider said policy acceptable.

The ideal of justification emphasises and ensures that a person is truly the *author of the acts* she authorises. Rainer Forst expresses a similar idea when he notes that justifying reasons must be such that "a moral person must be able to *take responsibility for* his or her actions before affected others and also generally."[15] In other words, a person must truly be able to conceive of the actions which have been furnished with coercive power in her name as actions she can claim to have consciously chosen, as opposed to them having been chosen merely in her name. Unlike Forst, I have so far conceived of self-authorship as a purely intrinsic value, rather than a trait a person must exhibit in order to conduct herself respectfully in public justification. This difference in terms of purpose, however, does not detract from the validity of the observation that, crucially, justification can be interpreted as a device directed at preserving individual self-authorship.

But is it not a contradiction for a theoretical conception of justification, which must ultimately attribute reasons to individuals, to define respect for a person as an end in herself in terms of her active authorisation of a given coercive act? After all, the very necessity to

---

[14]Rawls, for instance, affirms a perspective which does not only reject the equivalence of a person's ends with her autonomy, but rather conceives of a person's self as prior to her ends: "It is not our aims that primarily reveal our nature but rather the principles that we would acknowledge to govern the background conditions under which these aims are to be formed and the manner in which they are to be pursued" (Rawls, *A Theory of Justice*, p. 560). For a contrasting view that does indeed regard at least some commitments of a person as constitutive of the self which cannot be framed as reversible objects of the latter's judgement lest the person looses her ability to rationally acquire any further commitments at all, see Michael J. Sandel. *Liberalism and the Limits of Justice*. Cambridge: Cambridge University Press, 1982, p. 54.

[15]Rainer Forst. *The Right to Justification: Elements of a Constructivist Theory of Justice*. New York: Columbia University Press, 2012, 19. Original emphasis.

attribute reasons in the theoretical realm that I am concerned with seems to imply that there is no and can be no final act of judgement on the part of those to whom reasons are being attributed. If justification must merely succeed at attributing hypothetical assent rather than requiring actual endorsement of a given proposal in order for the latter to gain legitimacy for an individual, is it not merely a means for circumventing the judgemental authority of the individual person? By attributing reasons to individuals, do we not consequently break the very link between the direct authorisation of coercive acts and the respect for persons which I have claimed to be the very purpose of justification?

In the theoretical realm I currently operate in, attributing reasons does not intend to circumvent any actual judgements by real persons. The hypothetical reasoning I am engaged in for the purpose of developing a general normative argument for the justifiability of reasonableness as a standard in political liberalism is populated merely by hypothetical persons with hypothetical judgements. Not to take a hypothetical person's proclaimed ends and preferences to necessarily be their final word on the subject matter is, in this context, not to be considered an expression of disrespect for said ends. Rather, it is the only way to engage in a dialogue with the hypothetical agent in question, akin to asking a real person to state her reasons for adopting a certain end and entering a debate on their respective strength in a given context – and thus ultimately asking whether she might indeed be persuaded to accept justifications which refer to a different end, having identified new compelling reasons. Attributing reasons in an analogous hypothetical dialogue is merely to predict the outcome of such a dialogue based on the perceived (theoretical) strength of the arguments presented.[16] Justification in the theoretical realm can only emulate a person's reasoning and judgement, yet it emulates it with respect to a specific proposal, asking the hypothetical person whether she is prepared to lend her authority to the pursuit of a specific end. The attribution of reasons that result from this process of emulation affirms, rather than negates, the status of the hypothetical agent as the actual author of her choices.

In this section, my aim has been to develop a conception of the respect for persons as ends in themselves which is congruent to an understanding of attributive justification as a dialogue that emphasises the need to justify a given proposal *to* individuals. Such a requirement for *addressing* persons, as I have argued, cannot be understood if one adopts a conception of persons as ends in themselves which merely amounts to the prohibition of treating them as mere means. Neither does a conception of the respect for persons as ends in themselves framed in terms of respect for merely the outcomes of an agent's determination of her ends necessitate the degree of activation and confrontation which is implicit in the ideal of justification as I have understood it so far. Justification which, by hypothetically addressing persons, seeks to

---

[16]In the realm of real politics, the outcome of such a dialogue between real citizens is of course uncertain (the denial of assent even against an apparently objectively good reason is still possible). Certain standards are necessary for deciding to which extent a person can legitimately refuse to accept justifications, which by all means appear to give her good reason to endorse a given policy. This is not my concern at this point; however, the discussion in sections 3.3.3 and 3.3.4 on legitimate epistemic constraints in theoretical justification might offer some suggestions for reasonable epistemic standards for justification.

enter into a dialogue and asks for their judgement respects their status as ends in themselves in the sense that it respects their capacity to set and judge their own ends. A conception of respect for persons as ends which is congruent to the intentions of justification thus conceived should hence emphasise respect for the procedural dimension of persons' agency which is their capacity to *determine* their own ends. It is not their ends as such which justification considers to be primarily worthy of respect, but rather the process in virtue of which they have gained this status: the fact that they have been determined by individuals themselves. Justification respects the person as the legitimate author of her ends by aiming to ensure that all legitimate political (and thus potentially coercive) action to which she is subject can still be considered a result of self-authorship.

3.2.2 Respect and reflection

Having made a case for highlighting the respect for individual self-authorship as the crucial dimension of the principle of respecting persons as ends in themselves as manifested in the ideal of justification, I will now turn to discussing the implications of this understanding of the respect for persons for the standards regulating the process of justification. More specifically, I intend to develop a framework that allows us to determine which kinds of cognitions and inferences we may draw on in attributing reasons to individual persons and which (hypothetical) arguments on their part we can reject without failing to live up to the kind of respect we owe them as ends that is implicit in the very ideal of justification. In this section, I will argue that our focus on protecting a person's self-authorship in justification allows us to expect our interlocutors to adopt a particular epistemic norm in considering the reasons they are presented with – a norm which, given our conception of them as ends in themselves, we can sensibly assume to be rational for them in terms of instrumental rationality.

Consider the following claim: if we were required to assume that any conceivable end of a given person to whom justifications are being presented was of equal significance to her, we would be required to recognise any of her arguments against a proposed reason as equally valid, since it originated in an individual whose ends we are committed to respect in virtue of them being acts of self-authorship. The following example helps to illustrate this point: Alice presents Bob with a variety of reasons for buying a particular red, reasonably-priced, environmentally friendly second-hand car as their family car – reasons which she expects Bob to recognise since they satisfy two of his ends: to act in a financially responsible manner and to reduce his impact on the environment as far as he reasonably can. Bob, however, rejects her claim that there are indeed good reasons for him to consent to the purchase, arguing that to do so would violate his commitment to avoid owning any object of red colour. Since Alice so far lacks the means to question or normatively differentiate between Bob's ends in terms of their quality, she must take Bob's judgement of prioritising his desire not to possess any red objects over his other ends at face value and refrain from challenging his rejection of her justifications.

This is not to say that any such challenge would necessarily yield a different outcome. Yet, in the absence of any legitimate reason – that is a reason that is rooted in the respect for persons as ends in themselves – for gauging the quality of a person's ends, all acts of attributive justification that involved such a challenge would be rendered impermissible. The person presenting potential reasons would have no *mandate* to consider the ends which her reasons claim to satisfy to be superior to the ones which her interlocutor draws on for rejecting her proposal. As long as, in virtue of their status as results of a person's acts of self-determination, all her ends are created equal in the eyes of the attributor of reasons, the latter has no valid normative reason to assume that *the person in question* should share her verdict on the appropriate priority of ends.

If we take seriously the conception of respect for persons as ends in themselves developed in the previous section, it appears that we might have to accept this conclusion, thus conceding defeat for justification as an attributive exercise. Respect for persons' capacity to determine and judge their own ends might be considered to require us to adopt a permissive stance towards the standards they apply – consciously or unconsciously – in the course of epistemic self-determination. To demand that a person accounts for her choice and ranking of ends could be interpreted as defeating the idea that the act of determining her ends must not serve any ends other than the end in itself, that is the respective person herself. Accountability requires some shared frame of reference, thus calling for epistemic standards in this context which are legitimately applicable to all. Demanding such standards to be applied to a person's reasoning about her choice of ends for the sake of its amenability to external judgement would, however, seem to deprive the determination of ends of their quality as unmediated acts of self-authorship.

This danger could be defused – that is, the imposition of standards on a person's reasoning could be shown to be compatible with my current interpretation of the commitment to respecting any given person as an end in herself – if such a standard could be shown to be implicit in the conception of a person as an end in herself. In other words, legitimate external expectations about epistemic standards could be considered compatible with individual self-authorship if it can be demonstrated that their application itself would reflect an act of self-authorship.

The context of justification provides us with a setting for an argument that allows us to justify a (relatively weak) standard of reflectiveness, which is rooted in on our commitment to protecting a person's capacity to retain her agency and avoid being treated as a mere means when being presented with justifications. To that end, I argue, we can expect a person to reflect on her ends and to *consider* their relative status *for her*, not primarily for the sake of external intelligibility and accountability, but for the sake of retaining control over her agency.

This idea is best illustrated by an example that highlights the consequences of wilful deception within a more permissive conception of the determination of ends. Let us assume that

Alice has adopted the end of improving her physical fitness and has, as a result, decided to exercise on a regular basis. Aware of the fact that her lack of discipline might threaten that goal, she seeks to put herself under social pressure to stick to her commitment and thus forms a plan to persuade Bob to start exercising with her. Knowing that Bob does not share her intrinsic motivation of improving his physical fitness, Alice is looking for reasons which might sway Bob to accompany her to the gym in his leisure time, rather than hanging out at the pub with his friends, which she knows he enjoys a lot. Knowing Bob well, deceptive Alice is also aware of the fact that her explicit appeals to his fears about the negative impact that his inactivity might have on his health as well as feelings of guilt about neglecting his health are likely – not least due to their emotional nature – to sway Bob into agreeing to join her in her exercise regime right away, without further evaluating the weight of the reasons she has presented in comparison to any of his other ends, such as spending time with his friends.

Has Bob been treated as a mere means to Alice ends? Possibly. Alice specifically took advantage of the fact that she could circumvent Bob's comprehensive assessment of his ends – which might have endangered his assent to her proposal – by appealing to reasons she expected to elicit in him an instinctive and immediate reaction. In doing so, Alice used the structural power of her position in order to *direct* Bob's cognitions according to her convenience. The fact that his gut reaction to her proposal might have led Bob to give up a cherished end for the sake of a lesser one, or even one, which, in comparison to others, turns out to be of very little value at all despite its initial appeal, is not of primary relevance in this case. It is even possible that Bob's assent to her proposal may reflect a higher-order preference[17] for his health, in which case Alice's treatment of Bob would indeed have been in alignment with Bobs ends. What is relevant is that Alice did not appeal to said end *because* she expected it to be his actual primary end in this context. In other words, she did not *appeal to* his reason, but rather used her awareness of the efficacy of an appeal to a particular end in order to *direct* Bob's reason, irrespective of his considered judgement on his ends. Alice certainly did not prevent him from making such an assessment prior to accepting her proposal. However, given the circumstances, she willingly accepted the possibility that Bob might act to the detriment of some of his ends which he might otherwise have considered to be highly significant. It was due to the fact that, when presented with reasons intended to justify his assent to a proposal, Bob was not required to assess and reflect upon his ends and their relative merit, that he could be treated as a mere means to Alice's end at all. After reflection, he might still have decided to grant precedence to the same ends as prior to reflection. The quality of his ends is not at stake here. The crucial property of a requirement of reflection consists in its capacity to prevent him from becoming subject to

---

[17]Bob may, for instance, have a higher-order desire to act in accordance with what kind of person he wants to be – a person who acts towards what is objectively good for him (his health) rather than what *merely* gives him pleasure (time with friends at the pub) – rendering him susceptible to respond to reasons which, when compared to other lower-order desires, would be unlikely to prevail. See, for example, George Sher. *Beyond Neutrality: Perfectionism and Politics*. Cambridge: Cambridge University Press, 1997, p. 52.

Alice's deception.

What this example intended to illustrate is that, in order to ensure that an individual remains in charge of judging her ends in a given situation – a concern which is wholly supported by a conception of respect for persons as ends in themselves which emphasises self-authorship – it is worth asking what a person would have to do in order to ensure that those acts to which she confers her authority truly reflect the ends she has set for herself. Requiring the addressees of reasons to reflect upon the relative merit of these reasons and the ends they appeal to is a way to address the concern that they might become subject to deception. This argument relies on a view which equates the potential that a person might be deceived about the best or most relevant reason for her to act on in a given situation with that person being treated as a mere means to others' ends despite having formally performed an act of self-authorship.

In the context of the previous section, the claim that to deceive a person *intentionally* is disrespectful to her as an agent should not be contentious. In Waldron's words, "[m]essing with the options that one faces, changing one's pay-offs can be seen as manipulation [...] If it is done intentionally, it also takes on the insulting aspect of manipulation, for it treats the agent as someone incapable of making independent moral decisions on the merits of the case."[18] More importantly, to consciously refrain from drawing on what one believes to be a person's best reason supporting or defeating a particular proposal in the hope to gain her assent treats her not only as *incapable* of making an independent decision but as *unentitled* to do so. A person who intends to deceive her interlocutor does not honestly assume that the latter is incapable of making an independent decision, but rather fails to recognise the moral claim implicit in the quality of agency. As Christine Korsgaard notes, deception treats persons as mere means no less than (non-consensual) physical coercion does. Korsgaard characterises acts of deception as follows:

> Your reason is worked, like a machine: the deceiver tries to determine what levers to pull to get the desired results from you. Physical coercion treats some-one's person as a tool; lying treats someone's *reason* as a tool.[19]

The fact that a person shares the ends which are presented to her in an attempt to gain her consent to a given proposal does not automatically render her treatment respectful. As an object of a deceptive act, her commitment to her ends in general is awarded no independent significance. This is implicit in the concept of deception. The only reason the deceiver has to conceal *her* ends, referring instead to specific ends of the person whose consent she intends and needs to gain by means of deception, is that she fears the latter might withhold her endorsement if she was aware of the actual end pursued by the deceiver. As O'Neill

---

[18]Jeremy Waldron. "Autonomy and Perfectionism in Raz's Morality of Freedom". In: *Southern California Law Review* 62 (1989), pp. 1145–1146.

[19]Christine M. Korsgaard. *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press, 1996, 141. Original emphasis.

points out, "deception standardly works by revealing subsidiary aspects of action, which misleadingly point to some underlying maxim to which consent *can* be given. Deception works only when the underlying proposal is kept obscure. The deceiver's actual maxim therefore cannot be consented to."[20] By distorting the context in which a person is supposed to choose to consent to or dissent from a given proposal by withholding the actual motives of the individual seeking said person's agreement, intentional deception hence clearly seeks to interfere with the latter's process of genuine self-determination. Her reason is *directed* for the purpose of securing a certain end desired by the deceiver, rather than being allowed to assess and be the ultimate judge on whether the realisation of that end is desirable to her – and thus whether she would want to facilitate it – in the broader context of her ends.[21] It is their failure to appeal to a person's independent and informed judgement which warrants the conclusion that efforts of intentional deception express a lack of respect for a person's agency.

Is the same conclusion warranted when the person being addressed with reasons is not *intentionally* deceived by her interlocutor, but merely fails to recognise, at this point in time, that the end the reasons relate to would otherwise be considered by her to be inferior to one or several of her other ends? In this case, at the time she endorses the reason she is presented with she has no reason to question its status as a product of her self-authorship. (The same is true, however, in the previous case of intentional deception.) For the pre-reflective agent, her actions based on said reasons can reasonably be said to constitute acts of autonomous self-determination. The fact that her reflective self might have rejected the same reasons does not affect the status of the pre-reflective agent's choice as autonomous, given that she conceives of them as autonomous, as long as agency is not defined in terms of any objective condition that requires individuals to act upon their most important or most relevant ends. None such condition has been imposed by now. So far, we have not established that a person must not deceive *herself* (consciously or unconsciously).

However, *from an external perspective,* the *possibility* of external, intentional deceit may be reason enough to ascribe to the addressees of reasons an interest in avoiding such manipulation of their process of self-determination in so far as they are assumed to conceive of themselves as ends in themselves. We may have to concede that a person may deceive herself about – what at post-reflection stage would be – her best reasons for accepting a given proposal without being turned into a mere means to others' ends. But, as attributors of reasons, our commitment to respecting the person we address as an end in herself prevents us from intentionally deceiving her by selectively presenting her with reasons supporting our

---

[20]Onora O'Neill. *Constructions of Reason: Explorations of Kant's Practical Philosophy*. Cambridge: Cambridge University Press, 1989, p. 113.

[21]In our example, Bob *might* reconsider his decision to join Alice in her effort to exercise more despite his fear of ill-health if he knew that Alice only wanted him to join her in order to put herself under social pressure to keep up her exercise regime. For instance, that could be the case if he disapproved of the fact that Alice appears to value his presence merely for the sake of its motivational effect on her, rather than expressing any intrinsic appreciation of him as an individual.

position which we believe would resonate with her, despite our awareness that some of her other ends would, upon reflection, defeat the ends our reasons relate to.

That said, our commitment not to *intentionally* deceive said person may not be enough to prevent us from doing so unintentionally. Attributors of reasons ultimately aim to gain their interlocutor's support for the proposal they seek to justify and may hence be likely to focus disproportionately on those of their addressee's reasons that validate their particular objective. Such a bias lacks all the insulting notions of actively exploiting individuals' imperfections in reasoning about their ends, and merely reflects that the attributors of reasons are potentially as imperfect in their reasoning and assessment of their addressees' set of ends as the latter themselves.

Nevertheless, the position of the attributor is inherently more powerful than that of her interlocutor in the process of justification, since the former is capable and required to frame the initial choice situation for the latter, presenting her with one or more potential reasons. Despite her best – though most likely imperfect – efforts to focus on gaining her interlocutor's assent on the basis of reasons she considers to be sustainable for the latter, her position as the agenda setter might ultimately confer an advantage to her proposal which may not necessarily reflect the actual strength of its supporting reasons for the addressee. By presenting a particular reason to her interlocutor, the attributor is able to selectively direct the latter's reasoning towards a particular consideration, or at least frame the context in which she considers the proposal, ultimately increasing the likelihood that she will deem the perspective she is presented with to be salient. This effect has been widely documented, in particular in the context of political communication and the impact of news media reports on audiences' evaluations of certain issues,[22] and has sometimes been linked,[23] in terms of its psychological mode of action, to the availability heuristic. This refers to the empirical phenomenon that individuals' judgement is biased towards instances which, in terms of memory, are easily available to them.[24] The attributor can affect "which bits of [...] memory can most easily be brought to mind"[25] within her addressee, hence raising the possibility that her interlocutor's positive judgement has been elicited in virtue of circumstantial psychological factors, rather than the substantive merit of the proposed reason in the wider context of the latter's system of ends.

Such power structures inherent in the process of justification are not *intended* to take advantage of the recipient of justifications. Nevertheless, their deployment by a given individual for the legitimisation of a proposal which realises one of her ends, coupled with the structural

---

[22]See, for example, Shanto Iyengar, Mark Peters, and Donald R. Kinder. "Experimental Demonstration of the Not-So-Minimal Consequences of Television News Programs". In: *American Political Science Review* 76 (1982), or Lawrence Jacobs and Robert Shapiro. *Politicians Don't Pander: Political Manipulation and the Loss of Democratic Responsiveness*. Chicago: University of Chicago Press, 2000.

[23]See Donald R. Kinder. "Communication and Opinion". In: *Annual Review of Political Science* 1 (1998).

[24]See, for example, Amos Tversky and Daniel Kahneman. "Availability: A Heuristic for Judging Frequency and Probability". In: *Cognitive Psychology* 41 (1973).

[25]Kinder, "Communication and Opinion", p. 180.

bias in her favour could, as a whole, be seen as limiting her addressees' morally inviolable capacity to actively determine their ends in a given situation. Consequently, the latter, merely due to their position in the structure of the process of justification, may be considered at risk of being turned into mere means to the attributor's ends.

Given that, in the argumentative framework I have established so far, the very purpose of justification in the realm of political power and coercion is assumed to consist in maintaining the respect for persons as ends in themselves, this risk must not be tolerated. Attributors of justifications may hence be considered to have a duty of care to avert that threat by actively enabling their addressees' capacity for self-determination. In other words, they must ensure that their interlocutors act upon reasons (of whichever quality), rather than merely responding to the pressures (structural power) of the situation they are in. In virtue of both their commitment to respecting persons as ends in themselves and of their advantageous position in the structure of justification, the onus is on them to ensure that the latter are not misled by the forces at play in this context. Respect for their interlocutors as ends in themselves thus requires them to protect their addressees' entitlement to actively determine their ends according to their will. They therefore have reason to ensure that the latter have scrutinised the reasons based on which they confer their authorisation. Demanding that they reflect upon their ends and their relative merit in the respective context prior to accepting a reason they have been presented with hence counters the threat of structural deceit to their agency.[26]

This measure appears to be open to the charge of perfectionism. If we were genuinely concerned for our interlocutors' autonomous self-determination, should we not object to the requirement of reflection on the grounds that infringes upon their autonomy, given that it reflects *our* specific conception of their good and how best to protect it? The perfectionism implicit in the requirement of reflection is, however, weak at best. Requiring an individual to reflect upon her ends and potential choices does not commit her to make a particular choice. An agent's actual choice of ends in response to being presented with justifications may still appear entirely unintelligible – or even blatantly wrong – to external observers, including the attributor.[27] As Steven Wall argues, the requirement of reflection does not "favour [...] valuable pursuits over less valuable ones."[28] At most, imposing upon others a requirement of reflection may be said to be perfectionist in the sense of promoting autonomy as an ideal.

---

[26]It is important to note that the critical-reflection requirement I propose here is not rooted in the belief that an individual's choice can only be deemed to be truly autonomous if it is the outcome of critical reflection. (For descriptions of such an account of autonomy which does link critical reflection to the autonomous quality of a decision, see, for example, Sher, *Beyond Neutrality: Perfectionism and Politics*, p. 47, or William E. Connolly. *The Terms of Political Discourse*. Princeton: Princeton University Press, 1983, pp. 150–151.) The critical-reflection requirement I have discussed so far does not primarily aim to improve the internal quality of a person's reasoning. Rather, its purpose is to deflect any *external* interference with said person's autonomous choice.

[27]Recall Bob's decision to reject Alice's reasons for buying a particular environmentally friendly car on the grounds of its colour. Requiring Bob to reflect on his ends *does not guarantee* that his judgement will be any different post reflection. In principle, he could still insist on prioritising his commitment not to own any red objects.

[28]Steven Wall. *Liberalism, Perfectionism and Restraint*. Cambridge: Cambridge University Press, 1998, p. 197.

This latter kind of perfectionist action is not unequivocally considered to be inconsistent with a commitment to respecting individuals' autonomy, and is accepted by political philosophers who reject the former kind.[29]

However, even the charge of weak perfectionism with regard to the requirement of reflection can be rejected as unfounded. The motivation behind the requirement of reflection, in this line of argument, is ultimately prudential. In so far as we conceive of a person as the sole legitimate author of her ends, we have to consider it to be in her interest to retain her self-authorship and hence to erase the potential for manipulation and exploitation brought about by the very structure of justification. Reflecting upon the ends that might be used as reasons in justification, in order to ensure that, within this process, she is not, without a conscious choice, swayed to make judgements which do not reflect her weightiest ends in that context, ensures that her judgement remains a conscious act of self-authorship. For the imagined addressee of reasons, the requirement of reflection may indeed be considered to be prudentially rational. Yet, given the attributor's conception of the respect for persons as ends in themselves, her ascription of prudential reasoning to her interlocutor becomes a normative prescription for the attributor herself. After all, the protection of a person's agency in justification should not reflect what *the attributor* believes should be of value to her interlocutor, but rather what she must consider the addressee to want – or, more precisely, to be rationally interested in – as someone who conceives of herself as an end in herself.

Finally, the requirement of reflection turns into a more substantial means of epistemic accountability in the case of attributing reasons to hypothetical persons. Lacking the real judgement of a real person, we may only emulate a person's reasoning, replicating the way in which she could likely be expected to reflect upon on her ends. Without the requirement of reflection, any indication of her ends would have been sufficient for attributing to her a reason to accept or reject a given proposal. Now, we are under a duty to consider not only what she might consider herself to have reason to assent to, given all of her ends, but rather, what she might consider to have most reason to assent to. The requirement of reflection ensures that, in the case of attributing justifications to hypothetical persons whose judgement we have to emulate, we may and must examine and evaluate their ends comprehensively, rather than being allowed to draw on any given end they may be said to have as a source for attributing reasons without further scrutiny.

3.2.3 Objection: respect and rightness?

Before moving on, in the second half of this chapter, to assessing the compatibility of different standards and sources for attributing reasons to others with the conception of respect for persons as ends in themselves that I have developed, in this section I will discuss an objection to the central claim I have defended so far: the claim that attributive justification, in virtue of

---

[29]For a more elaborate discussion of autonomy-promoting perfectionism, see ibid., p. 198.

its commitment to emulating individuals' judgement, is indeed a means of treating persons with respect, that is, as ends in themselves, rather than as mere means. In general terms, the objection I mean to discuss – and eventually reject – can be framed as follows:

> To ensure that a person's treatment is aligned with that person's judgement on her relevant ends, and that, hence, she is not treated as a mere means to the ends of others, is of no consequence to the question whether we act in a way which is morally right, both
>
>   1. in general, and
>   2. towards her.

This objection captures the following sentiment: what is the moral value of justification, if its proclaimed commitment to refrain from treating individuals as mere means may well fail to *ensure* that people treat others in a way that can be said to be morally right?

I begin by assessing the more general claim that treating an individual with respect in this sense is not equivalent to acting rightly in objective moral terms. In other words, this variant of the objection contends that acting in such a way that a person is not treated as a mere means but as an end in herself does not ensure that in doing so we are doing what is morally right.

Consider the following example: non-swimmer Bob is about to drown in a pool. Non-swimmer Alice, and Eve, who is able to swim, are standing at the edge of the pool. Alice recognises that the only way to save Bob would be to ask and, if necessary, coerce Eve to jump into the pool and drag Bob out of the water. Alice's moral assessment of the situation concludes that to do so would be right, since Bob's life could easily be saved at, what appears to Alice, little cost – the relatively minor inconvenience of jumping into a cold pool – and almost no risk to Eve. Alice is, however, aware that Eve's assessment of her situation differs from Alice's in that Eve attaches a much higher value to her convenience than to Bob's life. If Alice were to coerce Eve to save Bob's life, she would not only treat her as a means to achieving her end of rescuing Bob, but, given that in doing so she would act contrary to Eve's ends, Eve would be treated as a *mere* means. Alice's commitment to acting in a way that does not treat Eve as a mere means thus forces her to act in a way that is morally wrong.

As this example intended to illustrate, it seems, the overall rightness of a person's action and the respect for persons express different dimensions of morality. Derek Parfit captures this distinction when he concludes that while "[i]t is wrong to *regard* anyone merely as a means [. . .] the wrongness of our *acts* never or hardly ever depends on whether we are treating people merely as a means."[30] It is doubtful whether this distinction remains as strong once we consider the commitment to refrain from treating individual persons as mere means in a wider context of justification.

---

[30]Derek Parfit. *On What Matters*. Vol. 1. Oxford: Oxford University Press, 2011, 232, original emphasis.

Ultimately, coercion must be justifiable to all, *i.e.* no individual must be treated as a mere means. If, furthermore, we adopt a broader interpretation of "treating a person as a mere means" that extends the maxim of respect not only to persons we could *use* as a means but also includes those whose ends we would need to disrespect as a result of our inaction for the sake of respecting others, the result of justification might gravitate closer towards the more comprehensive moral assessment of the situation. Parfit defends such a broader interpretation of treating persons as mere means:

> That can sometimes be done [...] by revising Kant's claim so that it also condemns acts that are *close* to treating people merely as a means. [...] When Bad Samaritans ignore someone who needs urgent help, they do not treat this person as a mere means. But they treat this person as a *mere thing,* something that has no importance, like a stone or heap of rags lying by the road. That, we could claim, is just as bad.[31]

To treat a person as one would treat a *mere thing* is to express disdain for their agency. As Onora O'Neill highlights, "[t]hings, unlike persons, are neither free nor rational; they lack the capacities required for agency. They can only be props or implements, never sharers or collaborators in any project."[32] A person who is treated as a mere thing in the course of another person's pursuit of her ends (*e.g* becoming subject to harm which is incidental to another person's actions or inaction), without technically being a *means* to the achievement of said ends, is thus treated as if she did not possess the quality of agency. In this case, she is not treated as a mere means, yet she is treated *as one would treat* a mere means. Consequently, failing to respect a person's agency does not require that we *act on* that person in a way that restricts her capacity for determining and acting upon her ends *in order* to bring about the ends we wish to realise. In the previous example, the situation is such that Alice's refusal to coerce Eve results in Bob being treated not as a mere *means* – his death does not achieve anything – but with the same attitude one would adopt towards a mere means, that is, with no respect for his entitlement to act upon his ends.

Hence, we can, as I believe, rightly conclude that being committed to the ideal of justification does not render us entirely oblivious to moral rightness. A single, isolated act of justification that refrains from treating a person as a mere means may indeed appear to require a person to act in a way, which, all things considered, would be wrong. However, if we take seriously the ideal of respecting individual persons as ends in themselves, we may well be required to adopt a less myopic conception of justification: while every single act of justification is only concerned with the legitimacy of a (potentially coercive) relation between two parties (the attributor of reasons and her addressee), we must not loose sight of the consequences that the success or failure to justify coercion to some individuals may have on others and

---

[31] Ibid., 227. Original emphasis.
[32] O'Neill, *Constructions of Reason: Explorations of Kant's Practical Philosophy*, p. 138.

our obligations to them. Thus, in reality, the relationship between individual or collective action and the capacity of individuals to act according to their ends is complex enough to warrant the conclusion that, in order to avoid otherwise inevitable deadlock, an individual's initial rejection of the justifications she is presented with does not necessarily have to be taken to be absolute. Rather, it should initiate an inquiry into her set of ends, asking which appropriate conclusions – that is reasons – can be derived from them in a given context. Such an inquiry may well uncover reasons for considering the overall moral requirements from a more objective, rather than a purely self-interested perspective like Eve's attitude in our example. Nothing less than the performance of such an inquiry is the purpose of attributive justification, as I will argue in detail at a later point in this section.

So far, I have argued that justification with its commitment to refrain from treating persons as mere means is neither independent of wider considerations of moral rightness, nor is any partial justificatory act sufficient to ensure that the attributor of reasons acts in a way that is morally right all things considered. But even though it may be insufficient in this regard to consider individual acts of justification in isolation, we may still insist that to address individual persons with justifications and thus treating them as ends in themselves is to do right *by them*. This claim, however, may also be challenged, in the context of attributive justification, by the two following objections which cast doubt on the moral link between the respect for persons as ends in themselves and the moral rightness of our actions *towards* these persons:

1. *Alignment of ends:* we can treat someone as a mere means and still do right by them, that is, act in a way they would find acceptable.

2. *Mistaken attribution:* we can refrain from treating someone as a mere means and still do wrong by them, that is, act in a way that they would, in fact, not find acceptable.

The first of these objections is fairly intuitive, if we consider that, from the perspective of an addressee of reasons, respect for her agency is expressed by ensuring that coercive action must be acceptable to her in light of the ends she designated as relevant in the respective context. It is possible to imagine a case in which she perceives the treatment she received to have met the above standard of justification, while being oblivious to the attitude and intentions of the person who addressed her with justifications. It is very well possible that the latter intended to use her as a means – and would have done so irrespective of her consent – only that their respective ends happened to be aligned. In this case, the addressee of reasons is *under the impression* of being treated with respect, despite the fact that the attributor conceives of her as a mere means.

To illustrate this idea, consider the following example: medical scientists Alice and Eve are looking for human subjects for the first clinical trial of a recently developed vaccine. Convinced of the importance of their research and the need to get the vaccine approved as

soon as possible, Bob volunteers to be injected with the still untested vaccine, thus accepting potential risks to his health. In this case, Bob is a means to both Alice's and Eve's end of continuing their research, yet he does not appear to be a *mere* means, since his end (to support their research) is in alignment with theirs. To Bob, his treatment (being injected with the experimental vaccine) is acceptable. Now consider the following difference in Alice's and Eve's attitude towards Bob. For Alice, who is committed to refrain from treating others, in this case Bob, as a mere means, his consent to participate in the trial is decisive for her decision whether or not to inject Bob with the vaccine. For Eve, however, who does not care whether or not she treats her interlocutor as a mere means, Bob's consent is irrelevant to her decision of administering the vaccine to him. In other words, Eve would have been prepared to coerce Bob if his ends had not been in alignment with hers.[33] For Bob, however, her attitude towards him – *i.e.,* the fact that she would be prepared to use him as a *mere* means – does not make a difference, at least in this isolated instance. The fact that Eve is, in fact, treating him as a mere means does not affect his impression that both Alice and Eve acted towards him in a way that he considers to be right.

As this example highlights, respect for persons as ends is also reflected in the attributor's attitude, not only in the outcome for the addressee of reasons. According to Parfit,

> [w]hether we are treating someone *as a means* depends only on what we are intentionally doing. Whether we ware treating someone *merely* as a means depends also, I believe, on our underlying attitudes or policies. And that is in part a matter of what we would have done, if the facts had been different.[34]

But why should we care about the attributor's attitude at all? At best, where the attributor's and the addressee's ends are in alignment, the former's attitude does not affect the latter's moral assessment of the outcome. Likewise, where there is no such alignment, an attitude of respect does not automatically translate into morally right action as far as the addressee is concerned. This leads us to the second challenge of *mistaken attribution:* we can refrain from treating persons as a mere means and still do wrong by them. How is this possible?

Returning to our previous example, let us assume that Alice is still committed to refrain from treating Bob as a mere means to her end of testing her experimental vaccine. Bob, however, after some reflection on his ends, has changed his mind about his willingness to participate in the trial of the vaccine and has withdrawn his consent at the very last minute. Let us further assume, that due to an administrative error, Alice has not received that information in time, and, assuming that Bob is still willing to receive the experimental vaccine, she administers it to him.

It is fair to say that, in this case, despite living up to her commitment not to treat Bob as a mere means, Alice treated Bob in a way he would not consider acceptable. His coercion

---

[33]This illustration is modelled on a similar example by Parfit. See Parfit, *On What Matters*, p. 215.
[34]Ibid., p. 215, original emphasis.

turned out to be unjustified. From Bob's perspective, he has *become* a mere means to Alice's end, despite the fact that she acted in best faith according to her maxim to respect Bob as an end in himself. Alice's behaviour may have been blameless, but can we still conclude that she treated Bob with respect?

In order to address this question, it is important to examine which perspective – that of the attributor of reasons, or that of the addressee – is significant for our judgement on whether a person has been treated with respect. In other words, can we still conclude that a person has been treated with respect if it turns out that she has, in fact, become a means to others' ends? I will consider the two following replies:

1. Yes, that person has been treated with respect because the attributor *has treated* her with respect when she based her decision to act on her honest judgement that said action can be expected to be in alignment with the ends of her addressee.

2. No, that person has not been treated with respect because she *has been coerced* for the sake of others' ends on grounds she did not actually deem acceptable.

Based on my line of argument in the previous sections, which emphasised the importance of respecting a person's judgement, as opposed to the mere satisfaction of her ends as the crucial dimension of respecting her as an agent, it seems that we should be inclined to accept *2*. Since her judgement has not been emulated correctly, it seems that the person who is being coerced can rightly claim that the individual attributing reasons to her has *failed* to respect her, irrespective of whether she is to blame for that failure. If we were to accept this reply, we might indeed have to conclude that our confidence in the ideal of justification as an instrument for maintaining respect for persons as ends in themselves is misplaced.

The question we need to answer in order to judge whether this failure *in treating* others as ends in themselves in this crucial dimension – that is, the fact that a person might indeed find herself to have *become* a mere means to others' ends – is of greater significance to the legitimacy of a justificatory attribution of reasons than the blamelessness of the attributor. To that end, it is worth highlighting some differences in terminology between *1* and *2*. *2* highlights that the person in question *has been coerced* on unacceptable grounds, thus emphasising the significance of her perspective as the one who is subject to coercion. *1* assesses the role of the attributor from a deontological perspective, highlighting that she *has treated* her interlocutors with respect, having done everything she could to treat the latter not as a mere means but as ends in themselves.

These differences in terminology are more than mere rhetorical finesse. Instead, they do reflect actual qualitative distinctions. Recognising these distinctions alone, however, does not provide us with an argument in favour of either *1* or *2*. Our emphasis on respecting

individual persons' judgement already provided us with an argument for accepting *2*. It remains to be shown if we can defend *1*. Why should we pay attention to the deontological sufficiency of a procedure which does not achieve with certainty what it claims to be its main purpose: to coerce people only on the basis of reasons they would be prepared to act on themselves?

In short, the answer is: because nothing can. In the theoretical environment in which we *have to attribute* reasons, that is, in an environment in which we do not have access to individuals' actual judgements, there is no process that could *ensure* that people are ever only coerced on the basis of what they would actually judge to be acceptable to them. This is both due to the natural limits of the attributors' capacity to perceive and assess their interlocutors' sets of ends, and to the fact that some of the ends espoused by different people are likely to be in such fundamental conflict that any political action (or indeed inaction) would fail to respect some of them. If we moulded hypothetical citizens – those we attribute reasons to – on the basis of real citizens' actual concerns and ends, only the requirement of actual consent could ensure that all coercive action has in fact been judged to be acceptable by and to all citizens. The likely result of this requirement is legislative deadlock, which – to the extent that even political inaction may have unwarranted coercive effects on some citizens – also fails to respect all citizens.

Theories that draw on attributive justification as a tool aim to resolve this very deadlock by means of identifying suitable arguments which are deemed capable of eventually arguing that some citizens ought to amend their actual judgements about the acceptability of a given policy. Such arguments are ultimately a product of the application of practical reason on the part of the attributor, albeit with other people's ends as its object. They take into account a person's ends as a matter of respecting her, but they do not have to pander to them, since, ultimately, they strive to make the best case for a given proposal that, as they are convinced, *should* be acceptable to her and to others. Individual people may still reject that argument. Normative philosophical arguments cannot *compel* individual's actual judgement. This is no reason for concern, given that their empirical judgement is not the standard for judging the validity of such arguments.[35] At the same time, arguments about what is justifiable to a person are not infallible and may sometimes fail to do justice to that person's own assessment of her ends. As I argued, the purpose of attributive justification lies in identifying arguments that the theorist is able to claim, in best faith, *can become acceptable* to all citizens upon reflection, rather than being acceptable to the latter in light of their current assessment of their ends. Given that, a person's claim that a reason has been falsely attributed to her does not as such defeat the validity of the attribution. What it does warrant is a re-examination of the reasoning leading to the attribution and the potential validity of the person's claims.

---

[35]In Ronald Dworkin's words, "[v]alue judgements are true, when they are true, not in virtue of any *matching* but in view of the substantive case that can be made for them. The moral realm is the realm of argument, not brute, raw fact." Ronald Dworkin. *Justice for Hedgehogs*. Cambridge, MA: Harvard University Press, 2011, 11, my emphasis.

This argument sheds some light on the significance of the deontological interpretation of respecting persons as ends in themselves and our question why the attributor's intention of not *treating* a person as a mere means should be decisive in our assessment of what it means to treat a person with respect in the context of attributive justification. Said context, as described above, neither enables nor requires the theorist – the attributor of reasons – to actually accommodate all ends whose satisfaction individuals claim to be the condition of their legitimate coercion. A person's (hypothetical) claim that she is treated wrongfully is of no immediate normative significance for the overall soundness of an argument that draws on attributive justification. (Hence, we can reject *2*.) The theorist must, however, always be aware in her exploration of (alternative) reasons to be attributed to a person that said reasons must aspire to be acceptable to the latter. In other words, she must find reasons that are aligned with that person's ends, according to her best-faith assessment.

Since the purpose of justification, as I see it, ultimately consists in *convincing* individuals to assess their judgements about what they deem to be acceptable to them, the actual alignment of a given justificatory argument with a person's current assessment of her ends is of no value as a criterion for respect on this theoretical level. Hence, we should accept reply *1,* rejecting an interpretation of respect for persons that ties the conclusion that they have been treated as ends in themselves rather than as mere means to a criterion which is irrelevant in the theoretical context just described. Rather, we should endorse a conception of respecting persons as ends in themselves which rewards blameless intentions and best-faith conduct on the part of the attributor of reasons.

3.3   Standards of liberal justification

In the second half of this chapter, I will proceed to draw on the conception of respect for persons as ends in themselves that I developed in the previous sections in order to identify standards that the justifications we attribute must meet – both in terms of the sources of reasons and the epistemic norms of inference we may draw on – in order for them to qualify as legitimate grounds for coercive action. I begin my analysis of appropriate sources for the external attribution of reasons by considering a strongly externalist conception of reasons, ultimately rejecting it due its failure to provide the addressees of justifications with the means to assess their merit. Acknowledging the importance of individuals' ability to relate to the reasons which are supposed to be applicable to them, I turn to a weakly externalist account of reasons which draws on the possibility of inferring them from within a person's existing, internal set of reasons and beliefs as the standard for justificatory reasons. I will subsequently discuss whether a person can validly reject such external inferences as a means of attributing reasons to her by questioning the sharedness of the epistemic norms underlying such inferences, concluding that this is not the case because the norms are implicit in all acts that aim for inter-subjective intelligibility. Finally, I will consider a further type of objection

to rationality as a standard of inference which does not deny its status as a shared epistemic norm, but rejects the assumption that human beings are able to and can normally be expected to strive to comprehensively apply rational standards to their reasoning.

3.3.1 Strong externalism of reasons

I begin my enquiry into the sources we may draw on in order to infer to reasons we may attribute to others with an assessment of strong externalism of reasons. According to this perspective, whether or not a person, Alice, can be said to have a reason R to endorse a belief X is no matter of the relation between X and possible elements of Alice's individual internal constitution, but is instead determined by the correspondence of X to a fact that obtains objectively and independently from Alice's individual perspective and is therefore entirely external to Alice.[36] As Gerald Gaus explains in his discussion of externalist justifications "[w]hether Alf is justified in believing β [...] [*i.e.,* whether he has reason to believe β] ultimately depends on whether there simply *are* good reasons for believing β."[37] An externalist perspective of this kind stresses that our understanding of what constitutes a good reason should not be detached from what we may consider to be true facts about the world. As Joseph Raz notes, "[i]t should be remembered that reasons are used to guide behaviour, and people are to be guided by *what is the case,* not by what they *believe to be the case.*"[38]

This emphasis on a necessary correspondence of facts and reasons certainly captures an important element of our intuitions about the nature of good reasons. Given the situation in which we know[39] X to be the case, we consider the fact that X obtains – or, in short, the truth of X – to give us a good reason to believe X. Surely, most would agree that the fact that the chair in front of me is blue is a good reason to believe that the chair in front of me is blue. Now, it is hardly surprising that, in virtue of this knowledge ("X is true"), we will question the quality of other reasons for beliefs about X. If I know Alice's belief "the chairs in the library are red" to be untrue – since I am sitting in the library she is referring to, I know that they are in fact blue – I can sensibly consider her reason for believing the chairs to be red ("My friend Bob told me that the chairs in the library are red.") to be a bad one, since it produced a belief that I know to be false. Yet, does this also warrant the conclusion that the fact that the chairs in the library are blue is a reason for Alice to believe that they are blue – even though, at present, there is no way for her to become aware of this fact? Indeed, Joseph Raz believes it does:

> To be sure, in order to be guided by what is the case a person must come to believe that it is the case. Nevertheless it is the fact and not his belief in it which

---

[36]See Gerald F. Gaus. *Justificatory Liberalism: An Essay on Epistemology and Political Theory.* Oxford, New York: Oxford University Press, 1996, p. 32.

[37]Ibid., 33, my emphasis

[38]Joseph Raz. *Practical Reasons and Norms.* London: Hutchinson, 1975, p. 17, emphasis added.

[39]This is to be understood as knowledge as it is commonly referred to, and not as knowledge in a more demanding, epistemological sense as justified true belief.

should guide him and which is a reason. If p is the case, then the fact that I do
not believe p does not establish that p is not a reason for me to perform some
action. *The fact that I am not aware of any reason does not show that there is
none.* If reasons are to serve for guiding and evaluating behaviour not all reasons
are beliefs. If p is the case, then the fact that I do not believe p does not establish
that p is not a reason for me to perform some action.[40]

Relying on such an externalist perspective, it may be perfectly acceptable to claim that there
is a reason for people to believe that the chairs in the library are blue, irrespective of their
potential lack of awareness of this fact. However, does it also allow us to claim that a partic-
ular person such as Alice *has* a reason to believe that the chairs are blue, and thus to *attribute*
such a reason to her?[41] From a liberal perspective, this distinction is not merely a matter of
linguistic sophistry. It reflects the conflict between what others believe or even know to be
a relevant fact to which a person should adapt her actions, and what she herself is able to
recognise as relevant in this respect.

In Alice's case, it is certainly rational for her to consider an objectively obtaining fact to be
relevant to her deliberations on choosing to believe one thing or another about the colour of
the chairs in the library. When attributing to Alice a strongly externalist reason, however,
I do not present her with the objectively obtaining fact itself (*i.e.,* the fact that the chairs
in the library are blue), but rather with my testimony of what I have recognised to be the
case. So, even if she were inclined to consider the content of my statement to constitute a
reason for her to adapt her present belief, the relevance of my statement about her reason is
not necessarily obvious to her. By simply claiming that she has reason to believe that the
chairs in the library are blue because they are, in fact, blue, I did not present her with any
evidence, apart from my testimony, for the truth of my statement and thus for its relevance
to *her* choice of belief about the true colour of the chairs in question.

To strong externalists it is unproblematic that the reasons they attribute to others are actually
inaccessible to the latter in the sense that they cannot account for their acceptability on their
own terms. In light of the conception of respect for persons as ends in themselves that I have
developed in the first part of this chapter, however, this is a deeply problematic stance to take.
I have argued that, if to treat a person as an end in herself in justification is to recognise her
as the sole legitimate author of her judgements, we must ensure that the reasons we attribute
to her are the result of faithful emulations of her judgement. In other words, they must reflect
a judgement which we can imagine her to have actually made.

A reason that has been attributed to a person can, however, hardly claim to be a faithful em-
ulation of a possible act of self-authorship on the part of said person if it bears no relation

---

[40]Raz, *Practical Reasons and Norms*, p. 17. Emphasis added.

[41]The distinction between the reasons that exist and the reasons one possesses is, for instance, also affirmed
by Robert Audi (*The Architecture of Reason: The Substance and Structure of Rationality*, pp. 53-55) and Gaus
(*The Order of Public Reason: A Theory of Freedom and Morality in a Diverse and Bounded World*, pp. 232-
235).

to any of her cognitions – *i.e.* beliefs or ends – by means of which she could have derived it herself. To consider an attributed reason to be acceptable to a person without her being able to reflect upon whether she herself can accept it as valid basis for guiding her actions is to expect her to assent to potentially being treated as a means to the ends of those who, based on their attribution, consider said treatment to be justifiable to her. A strongly externalist account of reasons is indifferent to her being able or unable to *self-determinedly* recognise the validity of the reason she is attributed. However, to respect her as an end in her own right is to refrain from interfering with her self-determined choices, unless she could recognise the reason for the intervention as a valid end to herself. Since *her* capacity to judge the acceptability of reasons to her is the standard for the legitimacy of coercion, her judgement on the validity of the reason justifying said coercion cannot be dispensed with. To be sure, a theoretical argument assessing the reasons a person may be said to have can only try to emulate as closely as possible the reflective process leading to her judgement on the validity of a given reason and its implications for her choices. Strong externalism of reasons, however, eliminates the possibility of an alleged reason to be assessed within such a process from the outset.

Enabling people to evaluate the validity of the reasons they are confronted with requires reasons to be *accessible* in a way strongly externalist reasons cannot be. In this context, accessibility is to be understood as the condition enabling people to relate the reasons in question to other elements of their individual mental constitution – the context which provides the criteria based on which individuals judge the validity of a reason. Strictly externalist attributions of reasons cannot insist on such relations as a necessary condition for a person to be considered to have a reason. As soon as the reason for a person to reject or endorse a given statement consists in *her* recognition of the validity of that statement and not in its truth independent of her judgement, it is no longer strictly external.

However, proponents of a strongly externalist conception of reasons deny that treating other people with respect requires that the reasons they are offered are accessible to them. William Galston holds that

> [t]o treat an individual as a person rather than an object is to offer him an explanation. [...] I would suggest [...] that we show others respect when we offer them, as an explanation, what we take to be our true and best reasons for acting as we do.[42]

An explanation, however, is monological. Its standard is whether it aptly reflects the *speaker's* reasons. For an explanation to succeed, it merely needs to be intelligible, rather than accessible, to the speaker's interlocutors. It may, but does not need to, take into account the perspective of those who are addressed by the speaker. Therefore, providing an explanation

---

[42]William A. Galston. *Liberal Purposes: Goods, Virtues, and Diversity in the Liberal State*. Cambridge: Cambridge University Press, 1991, p. 109.

for our actions is not (necessarily) an expression of respect.[43] As Gaus remarks, "murderers often explain *their* reasons, but we can hardly take this as showing respect for others."[44] Explanation is merely an effort to set out one's own reasons for believing or acting, whereas engaging in a process of justification requires us to ask whether our interlocutors have reason to consider a given belief or action to be acceptable to them. Although strong externalism of reasons can be sustained within an explanation, it is, as I attempted to show, incompatible with genuine efforts of justification, since strongly external reasons do not need to be *accessible* to all individuals to whom they are supposed to apply.

It is for this reason that strong externalism of reasons must be rejected for the purpose of my project: in my efforts to show that unreasonable people have reason to endorse the justifications they are offered upon their exclusion from the constituency of public justification, relying on a strongly externalist conception of reasons would be contrary to the liberal ideals that commit us to engage in justification in the first place: the respect for persons as ends in themselves who are endowed with moral autonomy, as manifested in their ability to determine and reason about their own ends.

3.3.2   Weak externalism of reasons

Rejecting strong externalism of reasons, I concluded that for one to be considered to *have* a reason to believe *X*, one must be able to access, that is, relate to, said reason in order to judge for oneself whether a supposedly objective claim about what is the case is to be trusted. If, as individuals, we were not to judge for ourselves whether a purportedly objective reason is likely to be an apt reflection of what is, in fact, the case, we would be committed to accepting any claim of the form "*X* is an objective reason for believing *Y*", irrespective of how far removed from our individual understanding of the world this claim might be.

Imagine, for the purpose of this example, that Alice grew up and still lives in a society which, until recently, used to be isolated from all other civilisations, so most of its members have never been in contact with foreigners. Imagine further that, however unlikely, this society has, so far been unable to produce blue colourants, so there have never existed any manufactured goods of blue colour. Only recently, the library acquired some blue chairs that were imported from outside the country. Having seen those chairs, I am telling Alice that she has a reason to believe that the chairs in the library are blue, since they are, in fact, blue, while Bob still claims that they are red. Obviously, Bob's statement is consistent with Alice's current set of beliefs about the world (including the belief that there cannot be any manufactured goods of blue colour, as her society is unable to produce blue colourants), whereas mine contradicts this very assumption which, so far, has never been contested. A strongly externalist account of reasons would claim that Alice still has a reason to believe

---

[43]See Gaus, *Justificatory Liberalism: An Essay on Epistemology and Political Theory*, p. 141.
[44]Ibid., p. 141.

that the chairs are blue, even though this belief would violate a long-held and, so far, valid assumption of hers. If, as I argued, Alice should not be prevented from evaluating the reasons that, supposedly, apply to her, can we reproach her for drawing on the set of assumptions and beliefs that, so far, have provided her with an adequate understanding of (and relation to) her environment? Assuming that Alice lacks any further information, we would not consider Alice's evaluation to be sound if she decided to believe what is actually the case, *i.e.,* that the chairs are blue. All things being equal, from her point of view there is nothing that counts in favour of believing my statement rather than Bob's, since its core claim contradicts something that has so far been entirely sound for Alice to believe.

This rather crude example is supposed to illustrate the intuition that a sound attempt to evaluate one's beliefs and the reasons one is presumed to have for holding them relies on the set of beliefs that currently provide us with an understanding of our environment. As Gaus notes, "to have justified belief [a belief one has reason to hold], one has to make the connection between the belief and the relevant considerations."[45] Such considerations are beliefs which, as individuals, we have already come to recognise as plausible *explanantia* of other phenomena, and which, therefore, can be considered to be beliefs that apply to us insofar as they provide the underpinnings of our understanding of the world: we have thus internalised them. In turn, they enable us to scrutinise beliefs we are asked to accept and the reasons we are offered for doing so, by providing points of comparison that have already proven to be reliable indicators for accurate beliefs in the past.

In the preceding section, I argued that strong externalism of reasons is incompatible with the idea that, for someone to be said to have a reason, she must be able to assess the beliefs and reasons that are supposed to apply to her. This requirement also reflects the argument I presented in section 3.2.2. In order to defend herself against being treated as a mere means in the process of justification, a person must be allowed to retain the capacity to scrutinise any attribution of reasons, ensuring that she can recognise her alleged endorsement of them as a genuine act of self-authorship. Now that we have considered the idea and process of such scrutiny more closely, we have gained a point of reference for the attribution of beliefs and reasons. Given that individuals can only sensibly evaluate the validity of such beliefs or reasons with reference to the set of those beliefs they have already come to accept, attributing to them beliefs and reasons (*i.e.,* claiming that they *have* a reason to believe *X*) which they could not possibly relate to from what they currently assume to be sound beliefs cannot be permissible. If, for moral reasons, we must consider coercive acts to be unacceptable if the reason supporting said acts cannot be made intelligible to those who are subject to these acts on their own terms and thus accessible to them, we must not attribute to them reasons and beliefs that clearly violate this condition in the way I described above. Attributed beliefs and reasons therefore must not ignore the set of beliefs and assumptions that currently structure a person's view of the world. In other words, we cannot attribute reasons and beliefs to others irrespective of what they are able to consider as an acceptable claim in light of what they

---

[45]Ibid., p. 19.

currently believe.

Is this to say that we cannot make claims of the kind "Alice has a reason to believe *X*" if we do not actually expect her to endorse *X* or the reasons for believing *X*, given that they seem to be incompatible with most of her other beliefs and assumptions? I believe this conclusion would be premature, for it implies that our belief systems are (1) wholly conscious and (2) static. This would imply that, in evaluating whether or not to endorse *X* or the reason she is presented with for endorsing *X*, Alice is (1) already aware of all possibly relevant considerations to draw on, and (2) no further, additional information could appear in and adapt her set of relevant beliefs. Both these assumptions provide a rather distorted perspective on what it means to have a sound reason for a belief.

To illustrate, I am going to revisit our previous example: I previously affirmed that we cannot sensibly expect Alice to acquire a belief which is wholly unsupported by the relevant set of beliefs she has come to accept in the past. Alice seems to have a good reason not to believe that the chairs in the library are blue, since it contradicts her long-held and, so far, valid belief that manufactured goods of blue colour do not exist. If this were the only relevant belief she could relate to, it does not appear to be outrageous to consider a claim of the sort "The chairs in the library are blue" to be unintelligible to her on her own terms. However, upon further inspection of Alice's set of beliefs, we might realise that, in fact, she can be said to have several potentially relevant beliefs that she did not take into account when making her initial judgement, some of which might nevertheless provide her with a reason not to reject the belief that the chairs in the library are blue. Let us assume that Alice recently met Colin, a visitor from abroad, who presented her with a bar of chocolate – another good which, so far, had not existed in Alice's country. Now, it seems to be fair to argue that, from the first-hand experience of having received this bar of chocolate, she may be said to be able to rationally infer to the following more general beliefs:

> *Y:* Although some goods may, so far, not have existed in this country, this
> does not mean they cannot exist in other countries.

and

> *Z:* Goods from other countries can be brought into this country.

These are perfectly sensible beliefs for Alice to accept, since they are supported by her first-hand experience of meeting Colin. In conjunction, however, these rather abstract beliefs also lend some support to other potential beliefs which Alice has been asked to evaluate and has so far rejected, such as my claim that the chairs in the library are blue.

When first considering whether to believe my claim about the colour of the chairs in the library, she did not refer to her beliefs *Y* and *Z,* but rather to drew on her long-held belief *W:*

> Manufactured goods of blue colour do not exist.

The fact that Alice did not reflect all beliefs which are relevant to the claim she intends to evaluate is neither surprising nor need it be a sign that she is not a competent reasoner. She may not be a perfect reasoner, but this is not a defect, since, as Christopher Cherniak argues, human belief systems are commonly segmented, with "relations between different 'compartments' [...] [to be] less likely to be recognised than relations among beliefs within one compartment".[46] Consequently, individuals may not usually be expected to bring to mind all their relevant beliefs, in particular, if they would need to be evoked in novel contexts.

However, since I am aware of the fact that, in addition to *W,* Alice accepts both *Y* and *Z,* I can point out to her that, with reference to these latter beliefs, my claim that the chairs in the library are blue is no longer inaccessible to her. This is because she is aware of – *i.e.,* has access to – reasons supporting said claim. Based on her awareness that imported goods may differ from those she has so far been surrounded with at home, as well as her experience that importing goods from abroad is possible, she is able to rationally infer that the presence of blue chairs in the library is, at least, a possible scenario, given that they might have been imported. So, I can demonstrate to her, that, in fact, she should be able to relate to my claim by drawing on some beliefs (*Y* and *Z*) which, initially, did not come to her mind when evaluating the potential belief I presented to her. My claim is thus not disjunct from what Alice is, in general, prepared to believe about the world, although she needs to restructure her belief system in such a way that the beliefs she acquired in relation to Colin's gift also consciously appear as relevant considerations in her evaluation of other claims. To do so would further confront her with a potential inconsistency within her belief set between the conclusion she can draw from *Y* and *Z* ("There may be goods that, so far, have not existed in this country.") and *W* ("Manufactured goods of blue colour do not exist."). Her acceptance of *Y* and *Z* does not in itself contradict *W,* but should, at any rate, be considered to diminish Alice's trust in the latter belief.

As this example was intended to demonstrate, a person may be said to have reason not to reject a belief as unintelligible, despite the fact that it seems to contradict other beliefs she deems herself to be committed to. This is true as long as said belief can be assumed to be made accessible to her by virtue of its relation to other information she may be said to be committed to recognising as part of her belief system. Such efforts to identify possible reasons a given person may have for endorsing or rejecting certain beliefs are part of a process which Gaus refers to as "open justification".[47] In taking this approach, according to Gaus,

> we treat *S* [a person's belief system] as open to new information and arguments and, from this external perspective, make judgements about what would then be justified in *S*. [...] Open justification, then, takes a person's current system of beliefs and asks, first, whether given this system that person is committed to

---

[46]Christopher Cherniak. *Minimal Rationality*. Cambridge, MA: MIT Press, 1986, p. 67.

[47]See Gaus, *Justificatory Liberalism: An Essay on Epistemology and Political Theory*, p. 31.

accepting some new piece of information, and second, whether that person is then committed to revising his or her system of beliefs in the light of that new information.[48]

In the previous example, I showed that Alice indeed seems to be committed to accepting some new information. In virtue of her acceptance of both *Y* and *Z,* it would be no longer rational for her to have unconditional faith in her belief *W* denying the existence of manufactured goods of blue colour. Consequently, she might need to revise her belief system such that she does no longer consider *W* a reason to outrightly reject all claims about the existence of manufactured goods of blue colour. This argument does not draw any conclusions Alice could not, upon reflection, possibly arrive at herself given her particular system of beliefs, despite the fact that she did not do so in this case.

If, as I aimed to demonstrate, her current system of beliefs offers more support for a proposed belief than she currently recognises, can it then be considered disrespectful to her to claim that she has indeed reason to accept belief *X*? In other words, would it be disrespectful to her as an autonomous and self-determining person to attribute to her a reason for believing *X*? Unlike a strongly externalist conception of reasons, weak externalism does not seem to dismiss her as the standard for what may be said to constitute a reason for her. It recognises that it is *her* judgement an attribution must strive to emulate in order to respect a person as an end in herself.[49] To attribute to Alice reasons which she may be shown to be committed to within a framework of open justification is to affirm that the validity of a reason *for her* is a function of all the factors she can be expected to be able to draw on in an evaluative process. An effort of open justification can be understood as emulating this reflective evaluative process. Thus, an external observer's warrant for attributing a reason to her is based on the assumption that, if presented with the respective argument, she would recognise the reasons which are attributed to her, because she can consider them as a rationally valid interpretation of her belief system.

Consequently, the attribution of reasons within a framework of open justification is first and foremost concerned with what *Alice* could recognise as a reason. Yet, it is still adopting an external perspective inasmuch as the authority to judge which potential reasons may be said to fulfil this criterion rests with the external observer. The attribution of reasons in open justification is, however, not strongly externalist to the extent that her belief system is considered to be irrelevant. On the contrary, as Gaus puts it, "that system is always the point of departure for the external criticism."[50] Yet, under open justification, the interpretation of said system is not deemed to be merely a private matter. Rather, the internal relations within a person's belief system, as well as the potential inferences that may be drawn between those beliefs, are supposed to be externally intelligible and open to scrutiny from an external

---

[48]Gaus, *Justificatory Liberalism: An Essay on Epistemology and Political Theory*, p. 32.
[49]See section 3.2.1.
[50]Gaus, *Justificatory Liberalism: An Essay on Epistemology and Political Theory*, p. 32.

perspective as well. Whether or not a person may be said to be justified in holding a particular belief, or whether such a belief is justifiable to her, is thus deemed to be a matter of external judgement. Adopting Gaus's terminology, I refer to the external attribution of reasons within a framework of open justification as *weakly externalist*.[51] Inasmuch as we conceive of the legitimacy of attributive acts of justification in terms of the conduct of the attributor,[52] *i.e.*, the person casting said external judgement, a weakly externalist approach to the sources of individuals' reasons provides us with a standard for assessing her attributions – a standard which reflects that any attribution must be able to be interpreted as an emulation of the individual person's judgement to whom reasons are attributed.

In asserting that weakly externalist attributions of reasons may be considered to express respect for those persons as ends in themselves, my argument so far tacitly assumed that the latter share their attributors' understanding of what constitutes a valid inference among a given set of beliefs. In other words, it presupposed that both parties are committed to the same epistemic norms. The viability of a weakly externalist conception of reasons rests upon the validity of this assumption. An individual could legitimately reject even weakly externalist claims, if she could rightfully argue that, as it appears *to her,* no valid inference can be drawn between a given belief and the reasons she is presented with for accepting said belief. Thus, if an individual could convincingly argue that the epistemic norms which apply to such inferences may be considered subjective rather than objective rules of reasoning – thus claiming to be committed to an idiosyncratic set of such norms – externally attributed reasons could no longer claim validity for her, as the process of open justification could no longer be said to emulate her internal process of reasoning. In the following section, I shall discuss whether such a subjectivist objection to weakly externalist attributions of reasons can be sustained.

3.3.3 Subjectivity of epistemic norms

In rejecting weakly externalist attributions of reasons to a person, those who claim epistemic norms to be subjective assert that what appears to be a valid interpretation of that person's set of reasons and beliefs does not need to be accepted by her as such, claiming that she might be unable to recognise the validity of the inferences drawn based on a selection of her beliefs. After all, according to epistemic subjectivists, she might be committed to an idiosyncratic set of epistemic norms which could differ fundamentally from that which other people are committed to. What might, from an external perspective, appear to be a valid inference within her belief system might thus fail to make any sense to her, rendering the proposed line of reasoning unintelligible to her.[53]

---

[51] Ibid., p. 32.

[52] See section 3.2.3.

[53] Consequently, she would also be rendered unable to access the reasons that are attributed to her in this process.

To illustrate, I return to my previous example, embarking upon another attempt to demonstrate to Alice that she has reason to believe that the chairs in the library are blue. This time, I show her a photo of the chairs in the library – accurately testifying to their blue colour. Let us assume that, previously, Alice has assured me of her belief *R:* "Photographic images constitute an apt representation of visual reality – a representation of what, as far as our visually perceptible environment is concerned, is in fact the case." When I show her the photograph, she further assures me of her belief *S* that the object I am presenting to her is a photograph in this very sense, and depicts the interior of the library in question, featuring only chairs of blue colour. These beliefs contain elements which can figure as the basis of a logical argument that fits the *modus ponens*. For this purpose *R* may be rephrased as a conditional claim $(P \rightarrow Q)$, forming the first premise of the argument:

> If (P) there is a photograph of an object, then (Q) there is a (visible) object of which the photograph is an representation.

We may refer to *S* as a second premise affirming (P), as it states that there is indeed a photograph of a set of objects.

> There is a photograph of blue chairs in the library .

Based on these premises, it is possible to conclude that *Q:*

> There are blue chairs in the library.

Inferring from Alice's acceptance of both *R* and *S,* and identifying them as premises of an argument in the form of *modus ponens,* I conclude that her possession and individual recognition of all these beliefs, by implication, provides her with a reason to believe that the chairs in the library are, in fact, blue. Being convinced of the validity of my external inference, I expect Alice to endorse my conclusion. However, having considered my argument, Alice still denies that her acceptance of *R* and *S* provides her with a reason to believe that the chairs in the library are blue. Instead, she claims that her epistemic norms do not commit her to accepting the validity of inferences drawn from an argument based on *modus ponens*. Her epistemic norms, she claims, do not include the *modus ponens* as a rule of inference. So, her affirmation of both *R* $(P \rightarrow Q)$ and S (P) does not strike her as inconsistent with the belief that the chairs in the library are not blue ($\neg Q$). Alice's subjectivist conception of epistemic norms thus resists even weakly externalist attributions of reasons. This is because it insists that inferences within her belief system must comply to the epistemic norms *she* deems to be applicable to her reasoning. Subjectivism of epistemic norms thus locates the authority to *interpret* a person's individual belief system solely within the person in question.

If individuals could provide a convincing argument in favour of their subjectivist conception of epistemic norms, this would set crucial limits to even weakly externalist attributions of

reasons. It would enable them to reject any such attribution on the grounds that no attempt to draw inferences within their belief system from an external perspective can constitute an adequate emulation of their process of reasoning, its outcome thus bearing no resemblance to inferences they would have drawn themselves. However, as I will argue, there are good reasons to affirm the universality of epistemic norms, as any successful attempt to argue against it would ultimately be self-refuting: in a world in which we did not share crucial epistemic norms, we would not be able to make ourselves intelligible to each other at all.

Wittgenstein's private language argument can be interpreted, as Christine Korsgaard does, as asserting that meaning is relational, *i.e.,* that it establishes a relation between an utterance and a phenomenon, *because* it is normative. That is, for us to be able to recognise an utterance (*A*) to have a particular meaning, to designate a particular phenomenon (*B*), we must recognise the norm that we ought to take *A* for *B*.[54] If this were not a normative relation, phenomena would not be denominated and linguistically linked in the same way by all. The existence of private languages is incompatible with this notion of normativity. Within their private language, reasoners simply reflect what they perceive to be the case. A private reasoner "would like to say: whatever is going to seem right to me is right."[55] The inferences a private language draws lack a "criterion of correctness",[56] because, by definition, their experience is idiosyncratic to the private reasoner. Yet, there can be no normativity in a denomination of a relation that cannot bind anyone to infer to a particular target but the private reasoner herself, because said target could not possibly be described outside the private language. This lack of normativity characterising the relations between a denomination and a perceived phenomenon that a private reasoners' language establishes renders said relations devoid of any meaning we could recognise when confronted with their utterances.

At the same time, we could not possibly interpret our interlocutors' reaction to our utterances, since we could not know what our statement means to them, in the sense that we could not know which inferences their private epistemic norms require them to draw based on what we just said. However, if our interlocutors react in a way that is intelligible to us – in a way we considered to be a rational way to react, given the epistemic norms we recognise – this indicates to us that their epistemic norms do not fundamentally deviate from ours. From their reaction, we can thus infer that they understood our statement in the way we intended it to be understood. Being in accordance with our expectations, their behaviour provides us with a kind of "bridgehead"[57] – an initial successful interpretation of their utterances, actions or reactions that indicate what might be their actual epistemic norms – which provides us with an initial insight into their way of relating to reality. In as far as we believe to be able to understand them – that is, our interpretation of their subsequent reasoning or behaviour

---

[54]Christine M. Korsgaard. "The origin of value and the scope of obligation". In: *The sources of normativity*. Ed. by Onora O'Neill. Cambridge: Cambridge University Press, 1996, p. 137.

[55]Ludwig Wittgenstein. *Philosophical Investigations*. New York: McMillan, 1953, p. 92.

[56]Ibid., p. 92.

[57]Martin Hollis. "The Limits of Irrationality". In: *Rationality*. Ed. by Bryan R. Wilson. Oxford: Blackwell, 1970, pp. 214-215.

does not contradict our initial interpretation of their reasoning (the bridgehead) – we must assume their epistemic norms to be similar to ours. If they were not, our interlocutors' reaction, that is, their behaviour and reasoning, would have to remain wholly unintelligible to us, since our interpretative application of our epistemic norms to their reasoning or course of action would not yield consistent outcomes. As Martin Hollis puts it, "what sentences mean depends on how the beliefs which they express are connected, and that to justify a claim to have identified a belief one must show the belief is connected to others."[58] If we are able to continuously make sense of these connections, they cannot be in contradiction to our epistemic norms. Their underlying "logic must either turn out to be a version of our own or remain untranslatable."[59]

Mutual intelligibility presupposes shared epistemic norms. Thus, where people appear to be intelligible to each other, epistemic subjectivism is prone to be refuted by the performance of communicative action that suggests that both parties to the conversation understand each other, that is, that they respond to their interlocutor's utterances and actions in mutually intelligible ways.

This, for instance, also applies to my previous example in which Alice rejected the *modus ponens* as a valid rule of inference. Insofar as we both understand the conditional of the first premise as the symbolic representation of a necessary inference[60] – which I may assume we do, since we appear to be perfectly intelligible to each other in all other respects, an observation which provides me with a bridgehead – her rejection of the necessary implication Q (Alice: $\neg Q$ – "it is not true that the chairs in the library are blue") does not leave me with any other option but to consider her to be inconsistent, due to her affirmation of both $P \rightarrow Q$ and $P \rightarrow \neg Q$. The fact that, otherwise, we seem to be perfectly able to understand each other suggests that the law of non-contradiction does indeed constitute an epistemic norm for her, casting doubt on the genuinity of her claim that, in all intellectual honesty, the *modus ponens* does not constitute a valid rule of inference to her.

As long as epistemic subjectivist claims of this kind remain isolated islands of alleged epistemic deviance without otherwise affecting individuals' mutual intelligibility, it is thus doubtful whether it would be disrespectful to attribute reasons to others despite their conscious rejection of these reasons on the grounds of their supposedly idiosyncratic epistemic norms. A general perception of mutual understanding provides sufficient reason to assume that fundamental epistemic norms and, among them, rules of inference are indeed shared

---

[58]Martin Hollis. "Reason and Ritual". In: *Rationality*. Ed. by Bryan R. Wilson. Oxford: Blackwell, 1970, p. 232.

[59]Ibid., p. 232.

[60]Hollis correctly observes that "we cannot first identify a native constant as 'if ... then' and then go on to show that *modus ponens* does not hold, since, if *modus ponens* does not hold, then the constant has been wrongly identified." (ibid., pp. 44, 232. Original emphasis.) It is certainly imaginable that Alice wrongly identified the conditional 'if ... then' as a statement that does not represent a necessary inference from P to Q. However, given that Alice and I share a language and do not otherwise differ in our linguistic interpretations, it is hardly conceivable that our understanding of the meaning (including the logical implications) of a statement of the form "if P then Q" should be expected to deviate substantially.

among those who debate the issue of whether one participant has reason for endorsing a certain belief. To successfully defend their subjectivist claims, those accused of falsely claiming to reason based on their idiosyncratic sets of epistemic norms would be required to sustain effective communication without recourse to the very norms they claim to be inapplicable to them. If they failed to do so, they would performatively contradict their very denial of the applicability of those norms to them. Yet they could only succeed if the meaning of their utterances remained unintelligible to their interlocutors. It thus seems to be impossible to even conceive of effectively sustainable claims to subjective epistemic norms. Even if such claims were genuine reflections of some subjects' mental constitution, they could not even hope to convince their interlocutors of their validity. Since the latter could not perceive vastly deviant reasoning as intelligible to them at all, they could not be expected to acknowledge subjectivist epistemic norms when confronted with an argument in their favour.

The cases which are relevant to my discussion of the external attribution of reasons are those in which communication can be sustained. As such, they allow for occasional claims of epistemic subjectivism to be treated as either erroneous assumptions about the inferences our epistemic norms require us to draw, or alternatively as intellectually dishonest attempts to deceive in order to evade epistemic commitments to undesired conclusions. In both these cases, a weakly externalist attribution of reasons which are inferred from a person's individual belief system, drawing on means of inference that may be deemed to form part of a mutually shared set of epistemic norms, does not infringe upon an individual's entitlement not to be coerced for purposes she cannot recognise to constitute ends to herself. This is because weak externalism operates on the reasonably well-grounded assumption that the person in question – if presented with a complete account of the inferential steps supposedly applicable to her – can reflect on and rationally assess any step in the external observers' reasoning.

3.3.4 Objections to rational inference

Throughout this section so far, I have defended a weakly internalist approach to identifying sources for reasons and proposed a rationalist standard of inference for emulating a person's reasoning based on these sources. In this final section, I will consider three objections to rationality as a standard for inferring from individuals' beliefs. None of these reject rationality as a shared standard of reasoning in general (like the subjectivists in the previous section), but they instead challenge its appropriateness as a standard for external inference due to imperfections of actual human reasoning. In other words, these objections claim that because human beings are not *fully* rational, that is, because they do not necessarily possess entirely coherent systems of beliefs integrated in a rationally consistent system across all their possible areas of belief, attributions of reasons which are the result of rational inferences among a person's existing beliefs are disrespectful since they cannot be assumed to resemble the judgements actual persons would arrive at. In other words, as the objection goes, individuals

cannot be assumed to be committed to a standard of perfect rationality because that is not how rationality as an epistemic norm affects the way in which individuals ordinarily relate to the world. I will proceed to discuss the following three varieties of this objection:

1. *bounded rationality;*

2. *irrational life; and*

3. *compartmentalized rationality.*

The first objection will frame the alleged unacceptability of rationalist inferences as a matter of natural constraints on our mere capacity to reason rationally, whereas the subsequent objections reject the validity of the *norm* which is implicit in the rationalist attribution of beliefs: that individuals should at least *strive* for a *fully integrated,* entirely rational system to govern their cognitions, including their beliefs and ends.

3.3.4.1   Bounded rationality

I will first consider *bounded rationality,* the least intricate of these objections. The assumption that human rationality is imperfect – or bounded – is fairly intuitive. Most people have difficulty to consciously keep track of all the underlying assumptions of their beliefs as well as their possible implications for the tenability of other beliefs they hold or some of the ends they value. This is not to say that they may not be able to recognise incoherences among some of their beliefs if they are pointed out to them, but merely emphasises that individuals do not usually – and may not have the capacity, to both in terms of time and physical ability – prophylactically connect and assess *all* the different areas of their cognitive landscape. Human reasoning appears to be subject to a variety of biases which, from a systemic perspective, appear to be irrational. Individuals' judgement has, for instance, been shown to be highly contingent upon the cognitive availability of information, its emotional impact, as well as the manner and context in which it is presented, rather than just its abstract content.[61]

Far from constantly crippling people in the practical conduct of their lives, some cognitive psychologists argue that these biases and cognitive short-cuts represent useful heuristics which enable individuals to make good decisions (*e.g.,* accurate predictions) most of the time, thus compensating for the deficiencies in their ability to reason comprehensively on the basis of all available information.[62]

---

[61]See, for example, Daniel Kahneman, Paul Slovic, and Amos Tversky. *Judgment under uncertainty: Heuristics and biases.* Cambridge: Cambridge University Press, 1982, or Richard E. Nisbett and Lee Ross. *Human Inference: Strategies and Shortcomings of Social Judgment.* Englewood Cliffs, NJ: Prentice-Hall, 1980.

[62]For empirical evidence on the efficiency of such heuristics, see, for example, Laura Martignon. In: *Bounded Rationality: The Adaptive Toolbox.* Ed. by Gerd Gigerenzer and Reinhard Selten. Cambridge, MA: MIT Press, 2001, pp. 147–172. See also Gerd Gigerenzer. *Adaptive Thinking: Rationality in the Real World.* Oxford: Oxford University Press, 2000.

However, the question remains, as Gaus phrases it, whether such biases "should be discounted in justification or whether what counts as good reasoning is determined by the actual ways of thinking that people employ".[63] Proponents of the *bounded rationality objection* would favour the second option. Their concern can be put as follows: we are not and probably cannot be perfect in our reasoning, so why should we and the claims we make – such as, for example, the claim that we do not have a reason to be reasonable – be open to criticism on the basis of imperfections or errors of reasoning?

Intuitively this claim appears to be blatantly false and it can sensibly be assumed that it would be rejected by most people in most contexts. Their reaction would probably be along the following lines: just because we recognise that we may not *be able* to achieve *perfection* in exercising a certain skill, this does not mean that another person may not legitimately criticise our mistakes. In other words, if we considered the unattainability of perfection to be a reason to reject any criticism as illegitimate, then almost no criticism would ever be legitimate, assuming that actual perfection is something that is, in fact, unattainable for most people with regard to most skills.

The following example illustrates this intuition: Alice and Bob are knitting socks. Their knitting skills are roughly the same, yet both are far from perfect in their exercise of their skill – if we take the standard for perfection to consist in an entirely symmetric pair of socks made up of perfectly uniform stitches – and are unlikely to ever achieve that standard. If the fact that Alice will never be able to knit the *perfect* sock allowed her to reject any criticism as inappropriate, Bob would not be entitled to point out a particular mistake of hers that consistently affects the uniformity of her stitches. If, in this case, Alice's proclaimed immunity against criticism strikes us as odd, it may be because it contradicts an underlying assumption we hold with regard to our aims when exercising our skills.

To clarify this idea, consider the following question: when would it make sense for Alice to reject Bob's constructive criticism as inappropriate? Assuming that it is presented to us kindly and respectfully, we would usually consider such criticism to be helpful, enabling us to improve our skill. Alice would have a reason to reject Bob's helpful advice, if she did not *want* to improve her sock, if she did not strive to knit – if not the perfect sock – the best sock she can possibly make, *i.e.,* to apply her skill to the best of her capacity. In addition, she would probably reject the criticism as inappropriate and disrespectful if she assumed that Bob was aware of her conscious choice not to knit the best sock she possibly could, as opposed to just *a* sock, interpreting his criticism as an attempt to assess her knit work by a standard that she does not share: a commitment to *improve* their skill.

It is this standard that we implicitly assume to apply to most individuals' exercise of most skills which may explain our puzzlement when confronted with the claim that criticisms that point out mistakes in a person's application of her skill are inappropriate. It also points to

---

[63]Gerald F. Gaus. "Ideology, Political Philosophy, and the Interpretive Enterprise: A View from the Other Side". In: *Liberalism as Ideology: Essays in Honour of Michael Freeden*. Ed. by Ben Jackson and Marc Stears. Oxford: Oxford University Press, 2012, pp. 178–198, p. 185.

the key fallacy of the bounded rationality objection. In claiming that errors in our reasoning cannot be subject to external criticism because perfection in reasoning is unattainable, and thus inappropriate as a standard that a person's reasoning should be expected to attain, the objection seems to suggest that the *striving* for perfection, or at least for improvement, is equally unsuitable as a standard by which a person's reasoning can legitimately be judged.

While an ought-implies-can type of argument can be successfully applied in order to rebut claims that only a perfectly rational set of reasons and beliefs entitles a person to legitimately reject a reason that has been attributed to her, it is without force against the alleged legitimacy of particular rationalist inferences. Such inferences do not criticise a person's reasoning hitherto for its failure to *achieve* perfect coherence among all her beliefs, but rather operate based on the assumption that said person would *want to improve* her system of cognitions by rendering it *more* consistent. The standard of criticism which external attributors' of reasons assume to be shared between them and their addressees is not perfect rationality, but rather the desire to render their reasoning and their sets of reasons and beliefs more rational.

The bounded rationality objection in its original form can thus be rejected as based on a mistaken interpretation of the standard to which attributors of reasons hold their interference with others' reasoning. To the extent that we can assume that people strive to improve their reasoning, we can assume that our efforts to point out arguments which aim to render their system of reasons and beliefs more consistent are aligned with their ends.

3.3.4.2   Irrational life

I have argued that we can rebut the bounded rationality objection with reference to the assumption that rationality as a standard for the validity of inferences is *shared* among individuals, given the assumption that they desire to render their set of beliefs and ends more consistent. My argument loses its validity once we give up the latter assumption. Can we imagine people who would not prefer an argument that integrates with their current beliefs and ends to a large degree over one that could not be tied – or only loosely – to a few parts of their system of cognitions? If we consider our everyday experience, this possibility is somewhat counter-intuitive. Does not everyone prefer the argument which appears to be aligned to the greatest extent with what they believe? It is, however, at least possible to imagine people who *choose* not to, or simply *do not* care about the rational merit of the beliefs or ends they decide to accept or act on in the same way as Alice in the previous example does not care about the degree of perfection (or imperfection) of her knit work.

Even though most people can probably be expected to strive for some coherence among their reasons and beliefs, this assumption alone does not give us warrant to treat those who declare no such desire as though they shared our desire. That would be to reject them not only as an empirical anomaly but as a moral one as well – one that does not even warrant our consideration when determining what it means to respect persons as ends. This will not

do. Just as we had to reject the subjectivist position on private epistemic norms by means of argument, we must consider the merits of subjectivist positions of this kind and, if need be, argue to reject them. It is important to note that the two objections of this kind I proceed to discuss – *irrational life* and *compartmentalized rationality* – do not, unlike the subjectivists in the previous section, claim that rationality does not constitute an epistemic norm for them. They do not deem rational inferences to be unintelligible to them. They recognise that, were they to infer between different beliefs, ends, and reasons, their means of inferring would be the same as other people's. They simply choose not to draw inferences between parts or the entirety of their system of cognitions in judging what to believe or how to act.

I will first discuss the more radical of these challenges to rationality as a legitimate epistemic norm for attributing reasons to others: *irrational life.* Imagine a person who claims that her decisions are not governed by any consideration about how they fit into her system of cognitions. Let us assume that Alice chooses her beliefs and actions not by deliberation about what she has reason to believe, or how she has reason to act, but by the toss of a coin. Unlike, for example, Parfit's consumer of the "irrationality pill" who chooses to be irrational in order to cease to be responsive to a threat by rendering himself unresponsive to its rational force,[64] she also claims not to have chosen this method for any particular reason, insisting that this is simply the way of choosing her beliefs and actions that comes naturally to her. Hence, she has not chosen irrationality for rational reasons. We cannot imagine such a life to be a very successful one in any ordinary sense: it is most likely that Alice's arbitrary behaviour will be met with bewilderment by others who are unable to engage with her in sustained communication or cooperation. Any meaningful interaction with a person of this kind seems to be impossible, given that there is no rationally discernible pattern to her beliefs and actions.

It is obvious that there is no way in which we could claim in good faith that such a person actually *shares* our desire for improving the rationality of our system of cognitions, thus rendering any rationalist inferences that we might draw among her beliefs from an external perspective both empirically inaccurate and morally disrespectful. We could not honestly claim that our inferences bear any resemblance to the manner in which we must assume she chooses her beliefs and ends. Does it follow that we *cannot* respectfully attribute reasons to a person by means of rational inference if said person exhibits a genuine comprehensive disconnection from rational life? Does this imply that we may ultimately be forced to coerce such a person based on reasons which have been derived from her beliefs in a way that she could not possibly be expected to consider applicable to her? Would we not, in doing so, treat her as a mere means to our justificatory ends if, given that in virtue of our knowledge about her manner of determining her beliefs and actions, we could not honestly assume that she could derive the reasons we intend to attribute to her from her system of cognitions?

This may very well be the case. Yet, it is questionable if it is at all possible for us not to act

---

[64]Derek Parfit. *Reasons and Persons*. Oxford: Clarendon Press, 1984, pp. 12–13.

towards any such person without threatening to treat her in a way that she considers to be unacceptable, irrespective of how we attempt to determine which reasons might be shared by her. Since we cannot assume that her actions follow any even loosely determinate pattern, there is no way to predict what might be acceptable to her. If confronted with such a person, we may have to conclude that we are unable to present her with appropriate justifications, since we have no means of determining what may be appropriate reasons for her.

In drawing this conclusion we have, in fact, given up on the assumption that said person is self-determining at all. Quite literally, such a person does not actively determine her beliefs or actions. She may still be the outlet of her beliefs or actions, in the sense that it is her who believes or acts. Yet, these beliefs are not the result of any conscious *act* of choice. Irrational life is hence akin to a person's death as an agent. If we have to conclude that such a person may not claim to be unjustly coerced when attempts of justification fail, this is not because her loss of agency renders her less deserving of justifications, but merely because we are literally *unable* to engage with her and potentially accommodate her ends.[65] This may lead us to conclude that, in persons who *are* rationally dead in the way I just discussed, we may well have encountered one of the boundaries of liberal justification.

However, is the same conclusion warranted in cases in which people have *committed* rational suicide,[66] as opposed to simply *being* dead as agents? I do not believe it is. A person who *chose* to reject rationality as the maxim guiding her choice of beliefs and actions committed at least one act of conscious self-determination. As Carol Rovane argues

> [i]t is certainly imaginable that a rational being might prefer to forgo engaging in rational activities. And this is imaginable even if the rational being regarded forgoing rational activity as a kind of suicide (because it would be the end of its life as a rational being). Now, if all this is imaginable, it is also imaginable that such a rational being might decide, upon reflection, not to commit this form of suicide [...].[67]

The fact that she had a choice and that she *chose* irrationality over other rational activities is indicative of the fact that irrationality is of some value to her. However, if this is the case – if her decision to forgo rationality is made because to do so is to act according to whichever value she recognises irrationality to have – it shows that rationality still guides her reasoning to some extent. Rovane argues that "the activity by which one reflects upon and comes to embrace the other substantive values in the light of which one can find reason to go on as a rational agent is already a rational process."[68] This is not only true if a person chooses to

---

[65]Note that this conclusion does not rely on the idea that the actual satisfaction of a person's ends is required for respecting her as an end (see section 3.2.3).

[66]This term has been borrowed from Carol Rovane's discussion of the tenability of different forms of irrationality in "Rationality and Persons", p. 340.

[67]Carol Rovane. "Rationality and Persons". In: *The Oxford Handbook of Rationality*. Ed. by Piers Rawling and Alfred R. Mele. Oxford: Oxford University Press, 2004, pp. 320–342, p. 340.

[68]Ibid., p. 340.

remain rational. The process by which she arrives at the decision to cease to be a rational agent is equally a rational process. Hence, Rovane's conclusion that "a commitment to being rational is already implicitly presupposed in the very evaluative process that is supposed to provide the conditional reasons for the commitment" applies to those who decide against irrationality, but also to those who consciously choose the irrational life.

Such a person cannot authentically reject the epistemic normativity of rationality in the same way as the irrational "native" which I previously discussed. She has at least once admitted that rationality is of relevance to her choices. In this respect, she remains an agent. Confronted with both the rational and the irrational dimension of her personality, we can conclude the following: any attempt to respect her chosen identity as a non-agent, on the one hand, renders us unable to determine which way of acting towards her would be in alignment with the way she would want to be treated – in the same way as in the case of the native irrational – thus depriving us of the *capacity* to treat her as an end in herself. If, on the other hand, we focus on the fact that rationality does possess some normative force over her, we can reasonably conclude that we share some commitment to the same standard of reasoning after all. Any attempt on her part to reject this assertion either plunges her back into her identity as a unpredictably irrational non-agent – if she simply denies our claim – or forces her to delineate her rational from the irrational dimensions of her identity. The latter, again, either requires a rational argument, thus reinforcing our claim that rationality applies to her, or, if she claims that such an argument cannot be made, renders her unpredictably irrational.

3.3.4.3 Compartmentalized rationality

A similar dilemma is encountered by proponents of the final objection I will discuss: those who insist that the attribution of reasons based on rational inferences across the whole of a person's system of cognitions is disrespectful, since different clusters of their ends and beliefs may not be, and may not need to be, fully integrated into an overall rationally consistent and coherent system of cognitions. Unlike proponents of the *irrational life objection,* they do not object to the idea that people are committed to striving for any consistency among their beliefs in general, but merely reject the claim that they must also be committed to striving for *overall* rational unity across all their different spheres of reasons, beliefs and ends.[69] They argue that people may well strive for rational unity within these spheres, while denying that a contradiction between beliefs from two different compartments commits them to revising one or the other.

Rovane asserts that

> [s]uch segregation is evident in degree in the lives of many human beings whom we find it possible to treat for the most part as roughly human-size persons. We may find, for example, that when we visit the corporation our friend "becomes"

---

[69]For an account of the emergence of such segregated spheres of rationality, see ibid., pp. 337–338.

a bureaucrat who cannot recognise the demands of friendship at all. What this means is that our friend's life actually takes up a bit less than the whole human being we are faced with, the rest of which literally belongs to the life of the corporation. We often describe this phenomenon as a kind of "role playing."[70]

William McGuire characterises this phenomenon as "a degree of 'logic-tight' compartmentalisation in the human thinking apparatus, by virtue of which certain sets of cognitions can be maintained isolated from one another, without regard for their logical interrelatedness",[71] while Herbert Simon notes that "human behaviour [...] viewed over a stretch of time [...] exhibits a mosaic character"[72] of this kind.

The *observation* that people segregate their cognitive systems into different compartments and do not seek to achieve overall rational unity appears to be plausible enough. McGuire suggests that compartmentalisation might be a coping mechanism (commonly exhibited by "authoritarian" personality types) allowing individuals to maintain conflicting, yet independently desired or required beliefs,[73] while Cherniak indicates that it may, on average, increase the overall efficiency of a person's "memory management".[74] Regardless of its empirical purpose and merit, our focus on what it means to treat individuals who exhibit compartmentalized cognitive systems with respect for their status as ends in themselves must incline us to ask whether a person's resistance to connecting these compartments is sufficient for her to legitimately reject conclusions drawn from rational inferences across segregated spheres of belief as unacceptable to her. In other words, may any person inclined to segregate her belief system in this way refer to the fact that she does not seek overall rational unity in order to ward off attempts to attribute beliefs to her?

As it turns out, proponents of the *compartmentalized rationality objection* encounter a challenge which is structurally similar to the one faced by those who choose to lead an entirely irrational life. Ultimately, they must either show that the segregation of their systems of cognitions is led by an overarching principle, thus conceding to some degree of rational integration across compartments, or concede that such a principle does not exist, *i.e.,* that the compartments have been chosen arbitrarily, or, more likely, have evolved over time in response to the demands and constraints of different situational contexts, such as family or working life.

The former case of *unprincipled* segregation generates the same challenge I discussed earlier with regard to the natively irrational person: if there is no pattern, or principle which delineates different compartments of a person's system of cognitions from each other and assigns

---

[70]Rovane, "Rationality and Persons", p. 338.

[71]William J. McGuire. "A Syllogistic Analysis of Cognitive Relationships". In: *Attitude Organization and Change: An Analysis of Consistency Among Attitude Components*. Ed. by Carl I. Hovland and Milton J. Rosenberg. New Haven and London: Yale University Press, 1960, pp. 65–111, p. 98.

[72]Herbert A. Simon. *Administrative Behaviour: A Study of Decision-Making Processes in Administrative Organization*. 3rd ed. New York: The Free Press, 1947, p. 80.

[73]See McGuire, "A Syllogistic Analysis of Cognitive Relationships", p. 98.

[74]Cherniak, *Minimal Rationality*, pp. 65–67.

specific cases and situations to specific compartments, external attributors of reasons face an impossible task in trying to predict which inferences would be considered appropriate by the person in question and which ones she would reject. Lacking any guiding principle, they cannot know which of a person's beliefs and ends cannot not be integrated with each other. This, however, renders them *unable* to act towards her in a way that they have reason to assume to be acceptable to her. In other words, it follows from the lack of a principle governing the segregation of her beliefs and ends into different compartments of rational unity that it cannot be ensured that her agency is respected in the attribution of beliefs. A person who insists on segregating her system of cognitions in an unprincipled way, hence, cannot be treated with the kind of respect for persons to whom, as I have argued, the ideal of justification is committed.

If, however, the person in question can provide us with one or several principles that explain her specific set-up of compartments of rational unity, she has indeed provided external attributors with the means to respect her compartmentalized system of cognitions, allowing them to distinguish why it would be wrong to expect her to integrate certain specific beliefs. Yet, in doing so, she performatively contradicts her claim that to infer across different compartments of her beliefs is inappropriate. A principle that explains why it would, for instance, be inappropriate to resolve an apparent contradiction between two beliefs in two different compartments, has, in fact, resolved their contradiction, by providing an account that offers reasons why they should not be compared to each other. Hence, it can be argued, that a principled segregator's different compartments are in fact integrated into a wider system which is characterised by at least the ambition of rational unity. The principled segregation of cognitive systems thus implicitly commits a person to recognising rationality as a standard which does not only apply within segregated spheres of reasons and beliefs, but also across them.

3.4 Conclusion

In this chapter, I aimed to identify standards that liberals should expect reasons which they attribute to other individuals to meet in order to warrant the claim that a given proposal can be justified to these individuals. Having highlighted the need for political liberalism to offer justifications to all due to its commitment to respecting persons as ends in themselves in the previous chapter, I have approached the task of identifying appropriate standards of justification by identifying the specific notion of respect for persons which is implicit in the process of justification itself. The ideal of addressing others with reasons, ultimately asking for their judgement on potentially coercive policies, reveals a conception of persons as ends in themselves that emphasises on individuals as *authors* of their choices. I have further argued that a conception of respect for persons as respect for their *self-authorship* imposes a minimum requirement of reflectiveness on the addressees' ends, protecting them against the potentially deceptive forces present within the process of justification that result

from the power wielded by the attributor of reasons as the agenda setter. The notion of respect for persons highlighted in this analysis defines respect in terms of the intentions of the attributor of reasons, rather than in terms of the success of such an attribution. I defended this perspective, arguing that an attributor-centric approach is appropriate in the specific context of justifying reasonableness to hypothetical citizens as part of an argument in political theory.

Considering possible sources of reasons that we may draw on in developing such an argument, I rejected strong externalism of reasons as incompatible with a conception of respect for persons that emphasises self-authorship, since it does not require people to be able to access the reasons which are supposed to be reasons *for* them. Weak externalism of reasons allows us to conceive of reasons attributed to a person as reasons the latter would herself have judged to apply to her, given that she can reflectively reach them from within her system of cognitions. Our reliance on rationality as an epistemic norm we draw on when inferring from and between different elements of said system cannot be said to diminish the authenticity of our emulation of a person's judgement. Their engagement in communicative interaction neither permits potential addressees of justifications to reject rationality as an unintelligible epistemic norm, nor can they consistently deny its irrelevance to their reasoning without rendering others *incapable* of relating to them in a manner deemed respectful to their agency.

A weakly externalist, rationalist approach to attributing reasons to individual persons can hence be argued to be compatible with the liberal commitment of respecting persons as ends in themselves within the scope of an argument which must draw on hypothetical citizens' judgement for support. In the following chapter, I will apply these conclusions to a substantive argument showing that reasonableness as a criterion for limiting access to the constituency of public justification can indeed be justified to individuals who are fundamentally unreasonable. In other words, I will look for sources of reasons which are rationally accessible from within individual citizens' sets of cognitions in order to show that they actually have reason to be reasonable when acting in the public political sphere. Unreasonable people, I will argue, can accept that reasonableness is a valid norm of political conduct.

Chapter 4

Factual commitments to agency

4.1 Introduction

Throughout the previous chapter, my concern about the compatibility between the external attribution of reasons for beliefs to others and the liberal commitment to treating people as ends has been rather abstract. I argued that it is not necessarily disrespectful to persons as ends to attribute to them reasons for endorsing a certain belief which they supposedly have. I emphasised that this can only be said to be true if the reason in question consists in, or can be inferred from, other elements of a person's belief system. To act on a thusly justified assumption that a particular individual should be committed to endorsing a certain belief does not violate the liberal commitment to avoiding coercive interactions with individuals if they cannot recognise the purpose of the treatment they are subjected to as an end for themselves. This is because weakly externalist attributions of reasons to others merely emulate the inferences a person herself would be rationally committed to drawing between an element of her internal set of beliefs and the purpose that is supposed to become an end for her as well.

At this point, I return to one crucial question raised in chapter 2: is it possible to show that unreasonable individuals have reasons to endorse reasonableness as a criterion for their exclusion from the constituency of public justification? The notion of reasonableness itself, in a Rawlsian sense, is taken to reflect crucial liberal procedural dispositions, such as a person's willingness to interact with others on fair terms of cooperation and to recognise the burdens of judgment.[1] These commitments, in turn, are considered to be rooted in the recognition of other persons as ends in themselves.[2] To ask whether unreasonable citizens have reason to endorse reasonableness as a valid criterion regulating access to the constituency of public justification is thus to ask whether they have reason to consider a commitment to recognising other persons as ends to be the only adequate attitude for citizens to adopt towards others as constituents of justification. As I explained in section 2.2.2, I am only concerned with

---

[1]Rawls, *Political Liberalism*, p. 49.

[2]As I have argued in section 2.2.3.

people who hold unreasonable views, but nevertheless aspire to enter and claim access to existing liberal societies' justificatory political processes for strategic reasons. Thus, in asking whether they have reason to recognise their unreasonable attitudes as a valid reason for their denial of access to these realms, we need to ask whether they themselves could actually deem these attitudes to be adequate within the institutions they aspire to enter. If we are to show that indeed they could not, we need to show that they would be committed to recognising others as ends if they were members of the constituency of public justification. Hence, this chapter aims to answer the following key question: would even unreasonable citizens have reason to recognise their fellow participants in the public political process as ends in themselves and thus as moral equals, deserving of being granted equal moral concern?

To be sure, to recognise the moral equality of persons is to acknowledge a *moral commitment,* not a factual belief of the kind I have so far been concerned with in defending a weakly externalist account of reasons. Ultimately, the idea of moral equality does not express a belief about what people *are,* but how they *ought* to be treated. It would, however, be premature to conclude that weakly externalist arguments are inapplicable to the issue at stake. After all, the liberal commitment to the moral equality of persons is not entirely devoid of a factual basis. Liberals consider all persons to be entitled to be treated as ends, and thus granted equal concern, because they recognise their principal equality in what they consider to be a morally relevant aspect: human beings' autonomy, reason, and agency, enabling them to conceive of themselves as ends, to determine and reflect upon their purposes and to act towards them.[3]

Without doubt, a commitment to the moral equality of all people does not depend on a person holding such a factual belief. A commitment to treating other persons as ends could also conceivably arise from mere intuitions and affections alone. A belief about the factual equality of persons as autonomous agents is thus not necessary for recognising their *moral* equality. (Neither is it sufficient, as I shall argue below.) However, the idea that people are equal in their autonomy and capacity for self-determination as agents is instrumental to what I consider to be a viable argument to show that unreasonable people may have reason to recognise other persons as ends. Drawing on the conclusions of the previous chapter on externally attributing reasons to others, in this chapter, I explore elements in unreasonable citizens' individual belief systems based on which a commitment to people's factual equality as autonomous agents may be attributed to them.

That said, a person's recognition of the supposed fact that people are equal in what liberals consider to be morally relevant aspects does not automatically need to entail a commitment on the part of that person to recognising the resulting *moral norm* to treat all people with equal moral consideration. These are separate issues. At this point, it still seems to be conceivable that a person may refuse to draw said normative conclusion which liberals take for granted. Therefore, a person's recognition of her commitment to the *factual* belief about the autonomy and capacity for self-determination of all people is not as such sufficient to

---

[3]See, for instance, Waldron, *Liberal rights: collected papers, 1981–1991*, pp. 36, 62; see also Alan Gewirth. *Reason and Morality*. Chicago, London: University of Chicago Press, 1978, chap. 2.

attribute to her the further-reaching moral commitment to acknowledging the equality of persons as ends. More argumentative work is necessary in order to establish whether or not such a conclusion may be drawn, and this is my main concern in the final section of this chapter. For the time being, I will defer this issue in order to consider a prerequisite step and present a line of argument which is capable of demonstrating that unreasonable people may be said to be committed to acknowledging a *factual* belief in other persons' agency.

4.2 Actions as sources for commitments

So far, I have framed a weakly externalist approach to the attribution of reasons as representing an individuals' commitment to a belief as a function of its relation to other beliefs which that person holds. For the purpose of exploring the permissibility of externally attributing reasons to others, this particular framing of the source of reasons for holding a belief was sufficient. It would, however, be somewhat simplistic to conceive of the reasons a person may have for acknowledging a certain belief only in terms of what may be inferred from other beliefs she holds – that is, from what has already been subject to cognitive affirmation on her part. This would only be true if nothing but her active and explicit cognitive affirmation of a given proposal (*i.e.,* the affirmative thought "I believe that *X*.") were to count as a legitimate indicator for a person's endorsement of said proposal, and thus for integrating it into her individual belief system. Such a restrictive conception of the origin of reasons excludes sources that reasoners may likely feel inclined to accept when deliberating about the beliefs they consider themselves to be committed to.

Consider the following example: Alice deliberates on whether she has reason to consider charity to be a valuable attitude. An overly intellectual person by nature, her deliberations tend to focus on various philosophical considerations about social justice, which initially lead her to the conclusion that she does not have any reason to consider charity to be of value to her. In response, Bob challenges Alice's denial of any such reason by pointing out that, for years, Alice has been donating money to the homeless people sitting in front of her local supermarket. Surely, Bob argues, Alice cannot claim that charity as a moral norm is of no importance to her, as she has regularly been *acting* in a way that structurally fits the description of a charitable action. Alice agrees that she acted out of charity, but nevertheless holds that her action does not give her reason to consider charity to be a valuable attitude.

Is Bob nevertheless justified in attributing to Alice a commitment to the value of charity by drawing on her charitable actions, despite her explicit refusal thereof? The reason why he – and indeed Alice – can do so lies in the purposiveness of human action. If our actions could not, at least putatively, be considered an expression of some commitment on our part, we would be compelled to conceive of the realm of individual human activity as wholly separate from the realm of individuals' purposes. In other words, we could not assume that what people do is related to what they feel committed to (*i.e.,* what they believe to be right

or good). This view, however, is incompatible with a conception of humans as agents, as self-determining beings deliberatively acting towards and based on their purposes. As Alan Gewirth argues, "there are no indifferent actions".[4] Actions always reflect a positive attitude towards the purposes they are intended to realise. If they did not, the agent could not account for her intention to act in a certain way and not another.[5] Alice could thus not claim her actions to be motivated by an attitude of charity – *i.e.,* assert that charity constituted the purpose of her action – while also denying that they give her reason to reflect positively on the value of said purpose.

Alice thus can only dismiss Bob's attribution of a commitment to the value of charity to her at the cost of (performative) self-contradiction. As I demonstrated in section 3.3.3, she cannot contend herself with a state of self-contradiction, at least not if she and her actions are to remain intelligible to others. With respect to her deliberations about the value of charity, this provides Alice with a reason to at least consider the belief (about the value of charity) which would render her performance of a charitable action intelligible to others (and herself), admitting that her actions constitute one source from which to draw relevant considerations in this context.

A person's actions thus permit some inference to be drawn to what she may have reason to believe. In our case, her actions provided a reason for Alice to believe – in contrast to the conclusion yielded by her theoretical considerations – that she might, nevertheless, consider charity to be a valuable attitude. As far as Alice is concerned, the commitment to charity is not as such necessary. She could, after all, stop acting charitably and embrace her theoretical considerations, thus rejecting the importance of charity as a moral norm. This does not damage the argument I presented so far. I argued that a person's actions may be understood as yielding commitments to beliefs whose endorsement is required to render said actions intelligible. However, this is not to say that the performance of these actions themselves is necessary in any case. Whether or not a person can refrain from performing a certain action, and thus whether the reasons it might provide may be eliminated in this way, is wholly dependent on the action in question as well as its respective circumstances. I discuss this point in more detail in section 4.5. At present, it is sufficient to conclude that a person's actions may figure as *one* source of reasons she might have for acknowledging a given belief.

In the example I just presented, Alice's action is only indicative of the origin of her commitment to charity, such as a moral intuition giving rise to an urge to act charitably. However, this is unproblematic for the argument that the intelligibility of an action is premised on the endorsement of a certain belief. Even though the action itself may not ultimately be the reason why a person may be committed to believing *X*, its continued performance is utterly unintelligible if the assumption that she actually believes *X* could be denied. There may exist an underlying source, but we do not need to draw on it, let alone identify it, in order to sustain our claim that the performance of a certain action alone gives a person reason to

---

[4]Gewirth, *Reason and Morality*, p. 40.

[5]See ibid., pp. 39–41. I will discuss this argument in more detail later in this chapter.

believe *X*. As I aimed to demonstrate in this section, the observation of the performance of an action is sufficient for an initial attribution of a reason to a given person for endorsing a belief which is crucial for rendering said action intelligible to herself and others. I refer to this property of human activity as its reliance on *implicit premises.*

4.3 Implicit premises about agency

The idea that the acknowledgement of a particular belief may be implicit in the performance of a particular action if the latter is to make sense is at the heart of Onora O'Neill's practical approach, exploring when and for what reason we should accord moral concern or, attribute "ethical standing",[6] as she phrases it, to others: trying to avoid "strenuous metaphysical claims, or blandly and groundlessly endorsing the actual views of scope and ethical standing of a particular time and place",[7] she suggests to treat questions of this kind "not as theoretical but as practical, that is as *questions that arise for and must be addressed by particular agents who need to determine to which other beings they must accord the standing either of agent or of subject (or both).*"[8]

Her motivation for drawing on the actions agents perform, that is, their practices when determining the moral concern they need to grant others, closely match the premises underlying my argument for the need to justify to unreasonable people the grounds for their exclusion from the constituency of public justification: people can only reasonably be expected to act or be treated based on terms they may be able to comprehend and accept. Neither contested metaphysical theories as to why people need to be committed to the moral equality of all people, nor plain assertions that people simply ought to be reasonable if they desire their claims to be accorded the same concern as those of others, constitute appropriate sources for people's attitudes towards their fellow co-citizens in a liberal but pluralistic society.[9] O'Neill is thus right in claiming that

> [i]f the elusive definitive analyses of personhood, agency, subjecthood and the like remain unavailable, it won't help to demand that agents base their views of the scope of ethical consideration on an objective account of ethical standing: this is precisely what they lack.[10]

Or rather, this is what they may not be expected to have or endorse. When deliberating about the moral concern people may or may not need to accord to others, we need to draw on

---

[6]Onora O'Neill. *Towards justice and virtue: a constructive account of practical reasoning.* Cambridge: Cambridge University Press, 1996, p. 93.

[7]Ibid., p. 93.

[8]Ibid., p. 93, original emphasis.

[9]See sections 2.2.3 and 2.2.4.

[10]O'Neill, *Towards justice and virtue: a constructive account of practical reasoning*, p. 99.

resources that are available and comprehensible to them, building a argumentative narrative they can freely endorse and feel themselves committed to, as I argued in chapter 3. This is in line with O'Neill's claim that people "need to construct rather than to presuppose an account of ethical standing,"[11] arguing that "the material they will have to hand to do this includes numerous interlocking assumptions about others on which they base their *activities*."[12]

O'Neill thus argues that a person's actions may not only be premised on beliefs that refer to herself, *i.e.,* what a person is committed to believing about herself in order for her actions to make sense, but also contain assumptions about what other people are like – or rather, must be like if her actions are to be intelligible to others and herself. In other words, what she is doing in relation to others may not only reveal what she is committed to believing about herself,[13] but also what she needs to believe about those she is interacting with.

Drawing mainly on extreme and violent interactions, O'Neill argues that by engaging in some such practices, the perpetrators performatively ascribe to their victims the very characteristics – "capacities, capabilities and vulnerabilities"[14] – they explicitly deny them to have when trying to provide justifications for their actions.[15] This is true, for instance in the case of adherents of Nazi ideology, justifying their practices of persecution, murder and genocide by depicting their victims as "subhuman – although their apparatus of torture and humiliation assumed victims vulnerable as only human beings are vulnerable."[16] In doing so, O'Neill claims, they performatively contradict their explicit assertions: those who treat other human beings in a way it only makes sense to treat human beings cannot at the same time consistently deny their very humanity. The intelligibility of the actions they perform is premised upon the assumption that their victims exhibit a particular set of characteristics. For an act of humiliation to make sense, it must be premised on its victim being "capable" of being humiliated, on it possessing a sense of self-worth, of dignity which may be degraded in an act of humiliation – attributes that, so far, only humans are deemed to share. O'Neill's argument thus aims to unmask the hidden (be it conscious or unconscious) elements of hypocrisy, of "pathological incoherence"[17] implicit in some kinds of (ethically more than questionable) activities by referring to the assumptions they are premised upon.

We cannot and do not chose these premises. They are not a matter of individual (moral) judgement, but of logical consistency. Apart from refraining from acting in a particular way, people can only change the narrative by which they explain or justify their actions. Yet, they cannot determine the premises underlying their actions, a denial of which would render said actions unintelligible. Thus, "commitments to others' ethical standing are taken on as soon

---

[11] O'Neill, *Towards justice and virtue: a constructive account of practical reasoning*, p. 99.

[12] Ibid., p. 99, original emphasis.

[13] As Alice, by virtue of her charitable actions, may be committed to believing that charity might, after all, constitute a value to her.

[14] O'Neill, *Towards justice and virtue: a constructive account of practical reasoning*, p. 100.

[15] Ibid., p. 100.

[16] Ibid., 106, n. 24.

[17] Ibid., p. 99, n. 15.

as activity is planned or begun".[18] If the commitments people incur by virtue of their actions may be said to form part of their internal set of reasons and beliefs, as I argued in the previous section, we are led to a seemingly paradoxical conclusion: what people do or intend to do might, with regard to some kinds of action, provide them with a reason to refrain from doing it.

One such reason which O'Neill recognises as a source of other individuals' ethical standing as human beings consists in their status as subjects and agents. Referring in particular to the mass atrocities happening in Nazi death and prison camps, she holds that

> [e]vidently many who organized and ran these camps combined strong assumptions that those whom they tormented and killed were *agents and subjects* (otherwise the whole hideous apparatus of torture, humiliation and secrecy, let alone the mythology of the International Zionist Conspiracy, makes no sense) with surface avowals that the treatment was appropriate since inflicted on beings who lacked ethical standing – "Untermenschen".[19]

In this context, the cognitive necessity to acknowledge the victims' status as subjects and agents does not arise from the violent acts themselves, but rather from the construction and affirmation of the narratives which are supposed to explain or justify – in short, rationalise – them. If the reasons based on which the imprisonment, torture, humiliation – that is, the infliction of pain (or death) on others – are being explained or justified consist in the need for precautionary measures against a conspiracy or punishment (as absurd as these claims may be in the particular context), the prospective victims' status as subjects and agents cannot be consistently denied by the perpetrators. This is because the very reasons by which they rationalise their behaviour with regard to their prospective victims presuppose some deliberate and purposeful prior or intended action – and thus the capacity for such action – on the part of the victims: defence against a conspiracy presupposes that others are *able* to conspire, punishment presupposes that others consciously acted the way they did and *could* have acted differently.

More generally, O'Neill argues, that

> [a]n attitude of resentment assumes that others acted knowingly in ways that are hurtful and that they could have done otherwise [...]. Wherever activity is based on the assumption of others who *can* act and react, the standing of those others cannot coherently be denied, whether or not those others in the event actually act or react.[20]

---

[18]Ibid., p. 100.

[19]Ibid., p. 102, n. 14. Emphasis added. Gaus argues in a similar vein in *Value and Justification: The Foundations of Liberal Theory*, pp. 292–293.

[20]Ibid., p. 103, original emphasis.

In doing so, she locates the justification for reacting with an attitude of resentment towards others in a specific quality of the will of the latter. As P. F. Strawson argues, we can imagine "occasions for resentment"[21] – "situations in which one person is offended or injured by the action of another"[22] – in which the "absence of special considerations [...] might be expected to modify or mollify this feeling or remove it altogether."[23] He frames these special considerations as instances in which the injury is accidental, *i.e.,* the person's action is not a manifestation of an intention to hurt, or as instances in which the action itself – whose outcome is hurtful to me – is not intended by the agent. These considerations are relevant to the justification of an attitude of resentment, because "they invite us to see the *injury* as one for which he was not fully, or at all, responsible".[24] As a reactive attitude, a feeling of resentment towards others, the focusing of my anger at them, however, presupposes that I assume them to be the source of my injury.

Yet in all these cases, they are not ultimately the source of my injury but rather an intermediary instance, as they did not intend the injure me. If they did not intend my injury to be the outcome of their actions – either because they willed something else, or their actions can not be seen as a result of them acting upon *their* self-determined will at all – the ultimate responsibility for my injury is to be located either in contingent factors (which prevented the actor from achieving his intended purpose) or in the will of those who forced him to act in a way that is hurtful to me. Hence, a feeling of resentment as a reactive attitude is not rationally justifiable if these considerations apply because it does not strike the actual cause responsible for my injury. In order for it to be justifiable, I must actually be able to attribute to the assumed initiator of my injury a certain quality of his will. As Strawson argues, reactive attitudes like resentment "are essentially reactions to the quality of others' wills towards us, as manifested in their behaviour: to their good or ill will or indifference or lack of concern."[25]

For other people's actions to be perceived as the realisations of their "ill will", or of a will of any quality, these people can only be consistently regarded as agents, that is, as individuals capable of purposive action, *i.e.,* of realising or acting upon their wills. However, it is not the perception of an *attitude* of resentment as such that compels a person to draw this conclusion.[26] Rather, it is the need to reflect upon her attitude, to rationalise it when required not only to explain, but to justify to a potential external observer her feeling of resentment. Only then will she be asked to render her subjective attitude intelligible to others. With respect to such instances which require the rationalisation of attitudes, O'Neill is right to emphasise that an attitude of resentment is premised upon an assumption of agency, rendering individ-

---

[21]P. F. Strawson. *Freedom and Resentment and Other Essays*. London: Methuen, 1974, p. 7.

[22]Ibid., p. 7.

[23]Ibid., p. 7.

[24]Ibid., p. 7, original emphasis.

[25]Ibid., p. 7.

[26]It is, for instance, certainly conceivable that I experience a feeling of resentment towards my cat after it scratched all of my furniture, although I could not honestly accuse it of having *intended* or willed to behave in a way which displeases me. Despite the fact that, in this case, I may *have* an attitude of resentment, I could not successfully rationalise it.

uals unable to rationally deny the agency of those they resent.[27] The adoption of a *rational* attitude of resentment thus yields a commitment to acknowledging the agency of those who are being resented.

O'Neill holds that such a commitment is in itself a moral one, thus not only requiring the factual recognition of another person's agency, but also an acknowledgement of her ethical standing – that is, of the moral concern she deserves to be granted by virtue of her quality as an agent. According to O'Neill,

> the assumptions on which activities are based [...] cannot be assumed for action or in taking up attitudes or for supporting policies and relying on practices, but then denied when ethical questions arise. In particular when agents *commit* themselves to the assumption that there are certain others, who are agents or subjects with these or those capacities, capabilities and vulnerabilities, they cannot coherently deny these assumptions in working out the scope of ethical consideration to which they are committed. Commitments to others' ethical standing are taken on as soon as activity is planned or begun.[28]

O'Neill's account contains a direct inference from the necessity of factually recognising others' agency to the moral implications of recognising another person as an agent, that is, as an autonomous being, capable of self-determined action. By linking the premises of activity to a commitment to the ethical standing of the objects of said activity, her account hence not only yields a conclusion about what others must be recognised as, but also about the treatment that is appropriate for them. O'Neill's argument thus presupposes that agents deserve to be accorded ethical standing, to be treated with moral concern *qua* their agency. Agency is thus assumed to be a reason for granting moral concern to an individual. This assumption reflects the liberal commitment to the moral equality of persons by virtue of their autonomy and capacity for self-determination.

Yet, it is worth asking whether O'Neill's inference from the factual recognition of agency to the necessity of recognising normative prescriptions with regard to the moral status of others proceeds too quickly. In particular, it might be vulnerable to criticisms which deny the derivability of norms from facts. It does not seem to be inconceivable that people could agree to the factual claim (about others' agency), but hold that the commitment to acknowledge those facts about others does not as such yield moral prescriptions. After all, they might hold that what they believe to be their moral commitments – including the question to whom they owe which degree of moral concern in a given context – is not a function of the facts they may be committed to believe about others, but stems from other sources, such as affections or

---

[27]This is not to say that they may not hold an irrational *attitude* of resentment towards others, but only that they cannot maintain this attitude if they are asked to rationalise their attitudes, as it may be the case in certain social contexts, such as the public political discourse, as I shall argue in the following subsection.

[28]O'Neill, *Towards justice and virtue: a constructive account of practical reasoning*, p. 100, original emphasis.

emotions in general. In order to be resilient against this kind of objection, O'Neill's practical account of individuals' factual commitment to another person's ethical standing needs to be supplemented by a further argument, explaining why the factual recognition of her agency, which is implicit in the performance of a particular activity, also yields a commitment to recognise the moral concern that is owed to her. Section 4.6 is dedicated to exploring whether there is a convincing argument linking factual and moral commitments with respect to the recognition to other persons' agency.

4.4   Rationality and intelligibility in public justification

I have so far been concerned with exploring the idea that actions may yield reasons if the intelligibility of their performance is premised upon assumptions the actor cannot reject. In particular, my discussion focused on forms of human activity which are premised upon the ascription of agency to others. I argued that the expression of a rational attitude of resentment may be considered one such activity. Thus, individuals who give expression to a rational attitude of resentment, or are in the course of rationalising their feeling of resentment, cannot consistently deny the agency of those they resent. As members of the constituency of public justification and participants in the public political discourse of a society, citizens can only draw on those attitudes which they can rationally sustain as reasons justifying their endorsement of the proposals they put forward in the public political process, as I am going to argue in the following.

Drawing on Jonathan Quong's Rawlsian definition, the constituency of public justification can be framed as the association of those people who, in the process of developing fair terms of cooperation, are committed to the public justification of political power.[29] In other words, they are willing to make proposals which they deem to be acceptable to all other members of said constituency, and, in turn, to consider such proposals put forward by others. Quong's and, to some extent, Rawls's conceptions of public justification[30] require citizens not only to assent to a given proposal, but to draw on a shared reason for giving their assent.[31] Public reasons are considered to be shared reasons. But even a less demanding conception of public justification, such as a convergence conception,[32] demands that participants in the public political process need to ensure that all other participants have reasons to consider any publicly binding decision to be acceptable to all. Under a convergence conception, it is sufficient for a proposal to be publicly justified if participants converge on their support for

---

[29]Quong, *Liberalism without Perfection*, pp. 290–291.

[30]Rawls's idea of public justification is indeterminate: his idea of the political conception of justice as freestanding emphasises its emergence from shared ideas of a democratic society modelled in the original position on the one hand, while also drawing on the support of an overlapping consensus of citizens' individual comprehensive doctrines. See Rawls, *Political Liberalism*, pp. 10, 25, 40.

[31]Quong, *Liberalism without Perfection*, chap. 9.

[32]See, for instance, Gerald F. Gaus. *The Order of Public Reason: A Theory of Freedom and Morality in a Diverse and Bounded World*. Cambridge: Cambridge University Press, 2011, pp. 283–287.

it, irrespective of whether they also converge on – that is, share – the reasons for supporting it. The individual reasons supporting a particular proposal thus only need to be accessible reasons for those who rely on them in justifying their support for a said proposal. In the following, I shall draw on such a modestly demanding conception of public justification in order to discuss the commitments which even a moderately rigorous framework of public justification might yield for unreasonable citizens.

Even a convergence conception of public justification requires the reasons citizens draw on in supporting a proposal to be of a certain character: they need to be intelligible to all other participants. Reasons which justify an individual's support for a proposal and draw on her own evaluative standards, such as her own set of reasons and beliefs, are intelligible to others if the latter can recognise as valid the inferences which link said individual's reasons to her endorsement of the proposal, or in Gaus's words, to proceed by "a sound deliberative route".[33] The requirement for reasons to be intelligible thus ensures that all members of the constituency of public justification are able to evaluate whether a supposedly publicly justifiable proposal can actually be considered to be justifiable to every last individual member on the latter's individual terms.[34] This requirement is not least an instrument which is supposed to facilitate dialogue among citizens, enabling people to defend or promote the merit of their proposals by criticising the reasons other citizens have so far deemed themselves to have for opposing these proposals.

Hence, when submitting a proposal to the public political process, an individual must be able to demonstrate to others that her reasons for endorsing it are indeed intelligible, *i.e.,* that her endorsement is rationally deducible from her personal system of reasons and beliefs. At the same time, she must also be able to demonstrate that others have individual, equally intelligible reasons to endorse her proposals. This imposes two procedural norms upon any participant in public justification, if their proposals are to be granted serious consideration within this process: *(1)* the reasons she cites in support of her proposal – both in defending her own and others' justification to endorse it – need to be rational in order for others to be able to evaluate their validity. *(2)* she needs to be willing to address others in order to promote or defend her proposal, demonstrating to them the reasons she deems them to have for endorsing her proposal.

Throughout my discussion of the challenge that the exclusion of the unreasonable from the constituency of public justification poses to liberal political theory, I have been concerned with those unreasonable citizens who wish to promote their unreasonable views on a political level, thus taking advantage of the public political procedures for strategic reasons.[35] It is thus not inadmissible to hold their unreasonable views to the test of whether they can be sustained if they *were* admitted to the constituency of public justification and had to be

---

[33]Ibid., p. 279.

[34]Ibid., p. 279.

[35]As opposed to those unreasonable people who wish to violently overturn liberal democratic procedures or otherwise pursue their aims by violent means.

subject to the procedural norms governing this realm. This provides a vantage point for engaging with the rational sustainability of fundamentally unreasonable views within the constituency of public justification.

4.5   Rational commitments to agency

What do the two procedural requirements which I identified in the previous section imply for people who aspire to introduce their unreasonable doctrines and proposals to the public political discourse? In particular, what constraints do they impose upon unreasonable citizens' ability to sustain unreasonable views within justificatory procedures among constituents of justification? I start discussion of this question by drawing on the first procedural norm: the rationality of justificatory reasons.

Throughout the preceding chapters, unreasonable people were understood to reject at least one of the following ideas: a conception of society as a fair and mutually beneficial system of cooperation, the freedom and equality of all citizens, or the burdens of judgement. In doing so, they may also be considered to reject the idea underlying these beliefs, *viz.,* that all human persons equally deserve to be treated as ends in themselves.[36] If they were allowed to enter the constituency of public justification, unreasonable people would ultimately be required to justify their position. This assumes that they intend to propose policies which draw on this position – or its derivatives – in the course of public justification. This assumption is not implausible, given that individuals are considered unreasonable precisely in virtue of their willingness to act in contravention to the key liberal principles mentioned above. If admitted to the constituency of public justification, we could imagine them to rely on one of the following reasons: first, they could refer to a diffuse feeling of resentment of the people whose moral equality they intend to deny. Second, they could deny the moral personhood of the latter, that is, their capacity to regard themselves as ends, and to desire to self-determine their purposes and actions. This would be to deny their agency, treating them merely as forces of nature which need to be controlled, but which are not considered to command our moral concern.

Imagine Alice to be a staunchly unreasonable person in the first sense, whose resentment of a particular group of her co-citizens – say those who do not belong the ethnically native population of her state S – instils her with a desire not to be required to live alongside them. Also, Alice is convinced that these experiences give her reason to demand that members of said ethnic group shall not be granted the same civic rights and protections (such as the right to free speech, voting rights, and the right to hold public office) as ethnically native citizens of S. Ideally she would like to evict them from society altogether, but deeming herself more likely to achieve her aims gradually by political rather than violent means, her first step is to rally and gain the public support of other citizens to deprive the objects of her resentment of

---

[36]See section 2.2.3.

any political influence and civic protections. In doing so, she expresses her denial of their equal moral standing, for her proposal intends to diminish their political status such that they will become subject to the coercive political power exerted by others.

Introducing her proposals to public political discourse, Alice – if she were a member of the constituency of public justification – would expect others to give due consideration to her proposal, taking for granted that they would assess her justification for recommending it, scrutinising her reason in terms of its intelligibility to themselves and its acceptability to her.[37] As I argued in the previous section, Alice's membership in the constituency of public justification would require her to show that her reasons for recommending her proposal can actually be considered to be rational reasons for her to endorse it. As an unreasonable person on a secondary level, Alice must ultimately refer to her resentment as a justification for endorsing a policy whose substance ultimately consists in a denial of the moral equality of some of her co-citizens.

However, in response, Alice's interlocutors in the process of public justification would be required to point out to her that her resentment cannot serve as a *rational* reason for recommending a proposal of this kind. As I argued in section 4.3, this is because a rational attitude of resentment itself is premised upon an assumption of agency. In other words, it can only rationally be sustained if those towards whom it is directed are recognised as agents who the speaker assumes to be capable of self-directed purposive action, *i.e.,* of acting upon their wills. Alice's success in drawing upon her resentment as a *rational* reason for recommending her proposal is contingent upon her implicitly acknowledging the agency of those whose rights she proposes to curtail.

It is this implicit commitment to others' agency that may be considered to set rational limits to her capacity to deny the moral equality of those others, imposing upon them constraints which are incompatible with their status as persons deserving equal moral concern. Those who consider agency to entail the acknowledgement of moral concern,[38] perceive the (necessary) affirmation of some people's agency on the one hand to be in sharp contradiction with the denial of their moral equality on the other.[39] More precisely, from this perspective, a person's explicit denial of others' moral equality in her proposal may be considered to be performatively contradicted by her commitment to acknowledging their agency as a result of her requirement to provide an externally intelligible, that is rational, justification for her proposal. Consequently, it is the very requirement to substantiate her proposal by rational reasons that could be seen as undermining the rational sustainability of her justification to recommend it. Within public justification, a possible reply to a person who, like Alice, out of resentment towards a certain group of people, proposes policies which express a denial of the latter group's moral equality could thus emphasise the following: that, in rationalising

---

[37]See also section 3.3.2.

[38]I explore this relation in more detail in the next section.

[39]As, for instance, Onora O'Neill. See *Towards justice and virtue: a constructive account of practical reasoning*, p. 103.

her attitude of resentment towards others, she implicitly ascribes to the latter the very characteristic (agency) that some consider to be the crucial reason for attributing moral concern to others. So far, this conclusion is conditional upon the validity of the inference from the premise of a factual commitment to a person's agency to the conclusion of a moral one to treating her as an end in herself. I consider this question in more detail in the following section. Up until now, my aim has been to show that the requirement to provide externally and thus rationally intelligible reasons for their proposals compels those people whose proposals are rooted in an attitude of resentment to admit to the agency of the very persons they resent.

I have so far been concerned with the requirement to offer *rational* reasons for proposals put forward in public justification, arguing that an attitude of resentment is inadmissible as a reason for the recommendation of a proposal which ultimately rejects the moral equality of persons. I will now consider the second kind of strategy unreasonable citizens might draw on in substantiating proposals which amount to a refusal of granting equal moral concern to some of their co-citizens: the very denial of their agency. In this case, an unreasonable person defending her proposal denying moral concern to others stresses that the reason why she deems herself to be justified to endorse the proposal in question does not consist in an attitude of resentment – which she could only rationally sustain by recognising their objects as agents. Rather, she might emphasise that it is the very lack of agency on the part of those to whom her proposal denies the same moral concern as all other citizens that renders said denial appropriate. However, is this path of justification any more sustainable than those which draw on attitudes of resentment? In order to discuss this issue, I turn to the second requirement which participation in public justification may be considered to impose upon citizens. Just as the procedural norms governing the realm of public justification require participants to justify their proposals by drawing on reasons which are rationally intelligible to all, they also presuppose a willingness to address said proposals to others, pointing out reasons the latter may have for endorsing them. It is this second procedural requirement which, as I am going to argue, sets another limit to the explicit framing of others as non-agents or non-humans within public justification.

If unreasonable people *were* granted access to the constituency of public justification, this requirement would equally apply to them, despite the fact that they might seek public recognition for proposals which deprive some of their fellow citizens of the same moral concern they are prepared to grant to others. Given that these citizens belong to the constituency of public justification as well, an unreasonable person would also be required to address her arguments to them. Consequently, she would be required to engage in a justificatory discourse with the "objects" of her proposal. This is despite the fact that it is the latter's very potential for meaningful interaction, their capacity for self-directed action as reflected in their agency, which she is intent upon denying as a justification for said proposal.[40]

---

[40] Of course, one might be inclined to interject, the denial of some citizens' intrinsically human characteristics such as their agency would never be acknowledged by all constituents of justification as a public reason, hence rendering it highly unlikely that the unreasonable proposals they recommend could actually gain a foothold.

Certainly, in our case, an unreasonable person who denies her interlocutors' agency would not interact with them because she expects herself to be able to *convince* them by arguments she could draw on in justifying a comprehensive curtailment of their rights to them. Instead, she would interact with them because, in public justification, she would be required to do so. Yet, from the moment she started addressing the people whose very agency she denies with a justificatory argument, even an unreasonable person would be caught in a conversation. Any unreasonable proposal would most likely be met with fierce protest and, importantly, a claim to be presented with justifications on the part of those whose moral equality it ultimately denies. Irrespective of the substance of her justifications, it is the fact that even an unreasonable person will be required to give such a reply at all that paves the way for the emergence of a contradiction between her behaviour and her explicit denial of her interlocutors' agency.

To engage in a justificatory discourse with a person can only be interpreted as a performative recognition of her agency. In presenting a person with reasons to endorse a given proposal, one performs a speech act. Such a performance, as J. L. Austin's theory of language affirms, cannot be reduced to its mere locutionary function,[41] *i.e.,* the conveyance of its ostensible meaning. (In this sense, a justificatory statement of the kind of "To believe X gives you reason to endorse *Y.*" could merely be said to convey its literal meaning as a factual assertive.) Yet, to justify a proposal to another person is not merely to express that she has a reason, but to request her to recognise said reason. Moreover, as I have emphasised throughout the previous chapters, processes of justification are ultimately aimed at gaining the consent of those who are being presented with reasons – reasons which are intended to convince the person in question to accept, rather than reject the proposal which it is deemed to support. In other words, to justify a proposal to a person is to convince her to recognise the reason in question as a valid purpose for her, and to act upon this purpose by giving her assent to the proposal the reason is deemed to support. Making a justificatory statement may be interpreted as a speech act "done with the design, intention, or purpose of producing"[42] a particular set of "consequential effects upon feelings, thoughts, or actions of the audience",[43] which Austin refers to as a *perlocutionary act.* To conceive of justification in the sense it has been understood throughout this work is to acknowledge that attempts to engage in a justificatory discourse cannot be understood without recognising the intentions which are embedded in the utterance of a justification. Or, as Donald Davidson holds "[w]hat we seek are intentions characterised in non-linguistic terms – ulterior purposes in uttering sentences."[44] Hence, to

---

However, this is besides the point. The question at stake is not whether or not unreasonable views could actually find their expression in public policies, but rather whether they can be refused articulation and serious consideration in the process of developing these policies in the first place. With regard to the latter issue, it is indeed a relevant feature of the public political discourse among constituents of justification that even those who seek public recognition for their fundamentally unreasonable positions as potential participants of said discourse are being forced to *interact* with the individuals whose humanity and agency they actively and explicitly deny.

[41] J. L. Austin. *How to Do Things with Words*. Oxford: Oxford University Press, 1962, pp. 100–101.

[42] Ibid., p. 101.

[43] Ibid., p. 101.

[44] Donald Davidson. *Inquiries into Truth and Interpretation*. 2nd edn. Oxford: Clarendon Press, 2001,

engage in a justificatory dialogue in the sense of performing a perlocutionary act is to intend to dispose one's interlocutor to act upon the reason she is presented with. To act with such an intention, however, can only be deemed to be rational if one makes some crucial assumptions about one's interlocutor's capacities. As Davidson affirms, in communication "the speaker must intend the hearer to interpret his words in the way the speaker intends, and he must have adequate reason to believe that the hearer will succeed in interpreting him as he intends"[45] Yet, there is a further premise implicit in the performance of a perlocutionary act of justification. To engage in a justificatory discourse with the intention of convincing one's interlocutor to recognise the reason she is offered as a valid purpose for her and to act upon this purpose is only a rational course of action if one assumes one's interlocutor to be able to act as one intends her to act. A justificatory argument can only be effective if the person it is directed towards is capable of controlling not only her actions (in this case the granting or withholding of her consent) by her volition, but her will itself. It is the latter that a justificatory argument intends to influence by disposing its owner to adapt her purposes in accordance with the reasons presented. It would be utterly senseless to appeal to a person with the intention of inducing her to adapt her purposes and to act upon them if one assumed her to be either incapable of self-determining her purposes, or of determining her actions based on those purposes. Both these capacities are representative of a person's character as an agent. Not to assume that one's speech acts could achieve the effect which is implicit in them as perlocutionary acts would be to admit to their inherent futility. To engage in a justificatory dialogue with a person whose agency one does not acknowledge would therefore be in contradiction with the implicit purposes of one's actions. A rational person hence cannot but consider her engagement in a justificatory discourse to be premised upon the assumption that her interlocutors are, in fact, agents.

Interaction within the constituency of public justification where members are required to seek each others' reasoned assent in justificatory dialogues thus contains an implicit commitment to the agency of their co-citizens. They treat each other as agents, or, at the very least, *as if* they were agents. Given this premise, an unreasonable citizen could no longer rationally deny the agency of any of her co-citizens she engages with in justificatory dialogues, as her denial would be performatively contradicted by her conduct towards the latter. In order to maintain consistency among the views she would be publicly committed to, she would either need to revoke her denial of others' agency, or refrain from justifying her proposal to them. It would appear to be rational for her to do the latter, as her personal justification for recommending a proposal whose substance essentially amounts to a denial of those others' moral equality crucially depends on her maintaining the former belief. However, if she decided to abort all justificatory discourses with those citizens whose agency she denies, she could no longer be said to honour even the procedural norms governing public justification. Her violation of said norms would place her outside of the only group of unreasonable people

---

p. 272.

[45]Davidson, *Inquiries into Truth and Interpretation*, p. 277.

whose exclusion from the constituency of public justification is at all controversial.[46] After all, unreasonable citizens who would not even be prepared to pretend to act as a proper member of the constituency of public justification by honouring its procedural norms if they *were* admitted cannot ask to be offered justifications for their exclusion, as their behaviour could not be taken to suggest that they even aspired to enter it.

Consequently, an unreasonable person who intended to resist her exclusion from public justification would need to continue adhering to its procedural norm of justifying her proposals to *all* other citizens. Yet, if she did so, she could no longer refer to some of her co-citizens as non-agents without undermining her personal justification for her respective unreasonable proposal. As she would no longer be able to rationally sustain the very belief she draws on in support of said proposal, her personal justification for recommending it would cease to be rationally intelligible to others. To refrain from violating the procedural norm of universal justification (to her co-citizens) would thus prevent her from meeting the procedural requirement I have discussed at the beginning of this section: the requirement for justifications to be rational.

Both of the arguments I presented in this section are intended to show that the procedural norms governing participants' conduct in the public political discourse ensure that all – including, and in particular, unreasonable – citizens remain publicly committed to all other citizens' agency. They limit the range of substantial beliefs and attitudes which may rationally be voiced by participants in the process of public justification and command the respect (if not necessarily assent) of all citizens. Consequently, those who aspire to enter the public political discourse cannot expect to be unconstrained in the choice of the convictions they intend to draw on in public justification. At this level, these constraints must only be perceived as procedural rather than moral ones, yet they might ultimately contribute to the maintenance and consistent protection of liberal values under non-ideal circumstances. In the preceding paragraphs, I have shown that unreasonable doctrines, specifically those based on attitudes of resentment or the explicit rejection of some citizens' agency, may rationally be rejected as untenable within public justification.

In the following, I intend to consider a moral implication that unreasonable citizens' factual commitment to the agency of their co-citizens may be considered to have in their capacity as aspiring members of the constituency of public justification, and which might ultimately provide them with a reason to endorse reasonableness as the only tenable attitude within this realm.

4.6   Agency and moral equality

The key question I shall discuss in this section is whether it is possible to infer from a factual commitment to another individual's agency to a moral commitment to considering

---

[46]We do not need to consider any other group of unreasonable people, as I argued in section 2.2.2.

their concerns on equal terms with one's own and those of all others in individual or collective decisions that affect them. More generally, the issue at stake in the following touches upon one of the fundamental problems in moral philosophy: the derivation of "ought" from "is". Implicit in this derivation is a transition from the theoretical to the practical level, *i.e.* from description to action. Thus, when asking whether or not we need to recognise certain others as our equals in moral terms, the question is no longer how to adequately describe them, but rather how this description can be said to affect our conduct towards them, and whether such a normative inference is possible at all. It is hence inevitable to confront questions about the very sources of normativity. However, my concern is not with morality as such – *i.e.,* what conduct with regard to others is to be considered appropriate in general – but with normative political theory – *i.e.,* what conduct towards others is to be considered appropriate in a political realm of a certain kind. Indeed, as will hopefully become clear by the end of this chapter, the norms structuring political procedures of public justification, may, in a nontrivial way, affect and govern inferences from factual to normative beliefs. Importantly, these norms require any such inference, including the refusal to draw one, to be comprehensible to all on a rational basis.

4.6.1   Emotivism and non-logical evaluative conclusions

I begin by drawing on this requirement in order to reject an emotivist argument, which asserts that evaluative conclusions based on factual beliefs are not only affective, but non-logical, leaving no room for a rational inference from factual to moral commitments.[47] Without touching upon the warrant of such an argument in moral philosophy more generally, I aim to demonstrate why it is inappropriate for a participant in public justification to base her refusal to infer from a person's agency to her moral equality on the claim that "[r]easons serve not to bring our attitudes into being but only to redirect them."[48] The emotivist non-cognitive view does recognise the possibility of drawing inferences between beliefs and attitudes, that is, it does not deny the possibility of a factual belief or statement being a reason for endorsing a particular norm.[49] Rather, it rejects the possibility of there being a logical reason that commits people to drawing a particular inference between a factual and a normative belief. As Stevenson asserts,

> reasons *for* approving [...] fall outside logic simply because they require inferences [...] from belief-expressing sentences to attitude-expressing sentences. [...] their bearing on the evaluative conclusion is neither logical nor illogical. It is simply nonlogical.[50]

---

[47]See, for instance, Charles Stevenson's discussion of moral attitudes in *Facts and Values: Studies in Ethical Analysis*, essay 5.

[48]Charles L. Stevenson. *Facts and Values: Studies in Ethical Analysis*. New Haven and London: Yale University Press, 1964, p. 90.

[49]Ibid., p. 84.

[50]Ibid., p. 85, emphasis added.

If the link between factual and moral belief was actually nonlogical, individuals could become encapsulated in their private ethical theory, capable of denoting fact *A* as the pivotal factor in bringing about attitude *B,* which would be immune to any external, rational criticism. They could thus not be restricted in making claims of the sort "agency may be a reason for *you* to consider others as morally equal, but I cannot recognise any link between this fact and its moral implication." As I have emphasised before, I do not intend to criticise the viability of this sort of ethical statement on a general, meta-ethical level.[51] From a political theorist's perspective, however, a refusal to subject one's moral reasoning – as far as it touches upon public matters – to rational scrutiny is unsupportable on the part of an (aspiring) member of the constituency of public justification. As stressed in the preceding chapters, the requirement of justifying one's public positions and proposals to all other members of said constituency entails a requirement for the rationalisation of these positions. Hence, one's reasons for adopting a particular moral stance cannot be merely subjective and private. They need, in principle, to be able to be made comprehensible to all of one's potential interlocutors, which can only be taken for granted if they are rationally deducible from their respective premises, given that rationality can be considered to constitute the shared standard of mutually intelligible communication. The public political discourse among members of the constituency of public justification thus cannot tolerate them asserting non-rational, non-logical links between factual bases and moral conclusions and rejecting rationalist ones, as far as they touch upon public matters. Within this political sphere, all "ought" statements need to be backed up by rational arguments. This is no less true for the claim that all people ought to be granted equal moral concern by virtue of their quality as agents.

4.6.2 Rationalist prescriptions

But what could such a rationalist path from the descriptive feature of agency to the prescriptive attribute of moral equality consist in? Alan Gewirth presents an argument deducing an individual agent's commitment to other agents' equal entitlement to her concern in terms of rights and freedoms. In doing so, he roots the relation between agency and moral concern in an agents' logical requirement to protect the basic requirements for her to be able to act as an *agent*, or as he phrases it "the generic features of his successful action".[52] The basic premise of Gewirth's argument is that individual agents can only deny the value of their own agency on pain of self-contradiction. Gewirth conceives of action as both voluntary (*i.e.,* free) and purposive (*i.e.,* intentional). This is to say that an action's performance is both "under the agent's control in that he unforcedly chooses to act as he does"[53] and "the agent acts for

---

[51]Framing the selection of factual reasons for endorsing a particular normative conclusion as an evaluative enquiry itself, the emotivist perspective claims to transform questions about the permissibility of concrete inferences from "is" to "ought", from meta-ethical problems into ethical ones. See, ibid., p. 87.

[52]Gewirth, *Reason and Morality*, p. 63.

[53]Ibid., p. 27.

some end or purpose that constitutes his reason for acting".[54]

The fact that purposiveness is one of the "features distinctively characteristic of the whole genus of action"[55] is crucial in furnishing a person's agency with an undeniable value to her. Given that it is irrational not to assign any value to the purposes of one's actions, *i.e.,* to the aims one is acting towards, a person can hardly deny the value of her *being able to act* towards these purposes. This twofold claim needs to be substantiated: one might, of course, conceive of a person who acts – or rather, behaves – in a certain way, while refusing to assign any value to the purpose of her actions. However, it is hard to consider this kind of activity, if genuine, as a form of action, that is, an individual's act of purposive self-determination, at all. If one's decision to perform a certain act did not imply some degree of endorsement of what one tried to achieve by it, one could hardly claim to have made a meaningful decision when choosing among several courses of action, or whether or not to act at all. (Which, itself, is a way of acting.) Assuming that a decision is meaningful if it selects the alternative which, considering all relevant circumstances, to one's best knowledge is most likely to achieve what one considers to be the best possible outcome, one could hardly claim to have decided at all if one could not discern between the desirability of the outcomes of these alternative courses of action. What is crucial is the act of having made a decision, as having decided presupposes an act of evaluation. If one could not discern between different purposes in terms of their value, one's selection of one of those values could hardly count as *choosing* a purpose. Such a choice would be no more than arbitrary and could thus hardly count as being self-determined,[56] for the outcome of such a selection process would be entirely independent of the individual in question. As long as a person claims to have acted, to have made a self-determined choice in selecting a particular course of action, she cannot eschew the conclusion that her choice is at least superior to all other alternatives, *i.e.,* that it has at least relative value. This argument provides some explanation to Gewirth's observation that "there are no indifferent actions, 'indifferent' meaning that the agent does not care at all whether he performs the action or not."[57] To do so would negate one of the basic premises of agency: self-determination.

If we need to assume that a person cannot but value the purposes of her actions, she can hardly reject the value of what Gewirth refers to as "the generic features that characterize all her actions."[58] In other words, a person must also value the goods which do not only (instrumentally) allow her to perform as an agent by providing her with the capacity of self-determinedly acting towards the purposes she considers to be good, but which are logical correlates of acting towards a valuable purpose. Gewirth considers these necessary, "generic

---

[54]Gewirth, *Reason and Morality*, p. 27.

[55]Ibid., p. 27.

[56]This scenario is not to be mistaken for the *decision* to choose randomly, that results from the conscious realisation that there *is* no best choice, that all alternatives are of equal value.

[57]Gewirth, *Reason and Morality*, p. 40.

[58]Ibid., p. 52.

goods"[59] to consist in the voluntariness or freedom of a person's actions, as well as in their purposiveness.[60] Lacking the former, she "would not be able to act for any purpose or good at all",[61] while the latter desire to increase the "level of purpose-fulfilment"[62] (or well-being) is a logical correlate of valuing one's purposes. Gewirth frames these generic goods as an agent's necessary objection to others' interference with her ability to control her conduct by her own choice and with the purposes she has already attained.[63] Yet, on a strictly more general level, these necessary desires may also be considered to implicitly contain a desire to be granted some concern by others, that is, for their desires to be attributed some weight in *others'* deliberations about how to act. She must want others to take due account of what is valuable to her, to the extent of preventing their interference with her generic goods. This is by itself not a moral claim. An individual agent's mere desire to secure the protection of the generic goods of her agency – despite the necessity of said desire – does not give rise to any obligations for others to act in accordance with this desire. When arguing that an agent needs to *conceive* of herself as having a right to the generic goods of freedom and well-being, Gewirth makes clear that he does "not directly argue that all persons or agents *have* rights [...] [but] rather that each agent must *claim* or *accept* that he has rights to freedom and well-being."[64]

The transition from an agent's individual recognition of her own need to demand – on pain of self-contradiction – the recognition and protection of her generic goods by others to the establishment of a general and universally binding principle requiring these goods to be protected proceeds via an individual agent's need to recognise the obligation which others' agency imposes on her. The latter need results from the epistemic force of what Gewirth refers to as "the criterion of relevant similarities".[65] This criterion issues the prescription to treat relevantly similar subjects or objects in the same way, given that the relevant similarity (as expressed by the attribution of relevance) is the decisive factor for considering the treatment in question to be appropriate for it. This principle thus requires agents – to whom the demand of their generic goods to be granted some weight in others' choices for action constitutes an epistemic necessity – to recognise others as having the same demand and to grant their demand the same weight as they cannot but claim for their own. Gewirth considers this moral application to be an exemplification of the logical principle of universalisability:

> if some predicate P belongs to some subject S because S has the property Q
> (where the "because" is that of sufficient reason or condition), then P must also

[59]Ibid., p. 52.

[60]Ibid., pp. 52-53.

[61]Ibid., p. 52.

[62]Ibid., p. 52.

[63]Ibid., pp. 52-53.

[64]Alan Gewirth. "Replies to My Critics". In: *Gewirth's Ethical Rationalism: Critical Essays with a Reply by Alan Gewirth*. Ed. by Edward Regis. Chicago, London: University of Chicago Press, 1984, p. 206, original emphasis.

[65]Gewirth, *Reason and Morality*, p. 104.

belong to all other subjects S_1, S_2, ..., S_n that have Q. If one denies this impli-
cation in the case of some subject, such as S_1, that has Q, then one contradicts
oneself. For in saying that P belongs to S because S has Q, one is saying that
having Q is a sufficient condition of having P.[66]

A person who recognises herself as an agent is thus committed to treating others in the same
way as she herself cannot but demand that others treat her. This commitment is not rooted
in a supererogatory concern for them as individuals. Rather, it emerges from the rational
necessity to extend to all other agents the attributes she needs to consider to be inextricably
tied to the status of a person who recognises herself as an agent. This commitment itself is
not a moral one, but a mere requirement of rationality. One person's necessary realisation
that another person cannot but desire others to take due account of her generic goods is still a
factual belief. The crucial factor in transforming this commitment from a merely descriptive
to a moral dimension consists in the perceived prescriptive character of an individual agents'
necessary desire to secure the generic goods of her actions. Agents cannot but want to
require others not to interfere with said goods. Again, the perceived prescriptive character of
these demands alone does not commit others to recognising these prescriptions. Yet, upon
universalisation, the prescriptiveness can hardly be deemed to vanish on the more general,
impersonal level, given that the person in question still considers her claim to the generic
goods of her action to be prescriptive. The evaluative-prescriptive tie she must believe to
exist between her own agency and the demands she needs to make upon others cannot, in
accordance with the principle of universalisability, change its (prescriptive) character only
because it is being abstracted from the individual herself.

Gewirth's account has been criticised for making too hasty a transition from a person's pru-
dential commitment (what she must want for herself) to a moral commitment she incurs by
virtue of rationality. R. M. Hare doubts that the prescriptivity of the claim to the generic
goods of action an agent is required to make is as universalisable as the fact that as an agent
she needs to make this claim.[67] To Hare, the crucial "question is, must he prescribe and want
the similar purposes of others to be fulfilled in similar circumstances? If not, his *prescrip-
tion* is not a universal one, and therefore not moral."[68] Gewirth responds to this criticism by
pointing out that, by acknowledging other agents' prudential requirement "to seek the neces-
sary conditions for achieving *their* purposes", an agent "is in the position of endorsing other
agents' fulfilment of their own agency needs – and this endorsement is a moral one because
the agent who says [...] [this] thereby takes favorable account of the interests of persons other
than or in addition to himself."[69] It is the notion of the process of universalisation which leads
to the assumption of a favourable attitude towards others' requirements that allows for the

---

[66]Gewirth, *Reason and Morality*, pp. 104-105.

[67]R. M. Hare. "Do Agents have to be Moralists?" In: *Gewirth's Ethical Rationalism: Critical Essays with a Reply by Alan Gewirth*. Ed. by Edward Regis. Chicago and London: University of Chicago Press, 1984, p. 56.

[68]Ibid., p. 56.

[69]Gewirth, "Replies to My Critics", p. 211.

transition from the recognition of the sharedness of a prudential commitment to achieving the protection of the generic goods of one's action among agents to the recognition of the prescriptive force of this commitment. In other words, the process of universalisation transfers an agent's own evaluation of her claims as prescriptive to the similar claims of similar agents, for what she accepts for herself as an agent in terms of prescriptiveness, she cannot deny to other agents without contradicting herself.

Gewirth's account has nevertheless been criticised for lacking this very element of a sense of mutual prescriptiveness. Striking a similar note as Hare, Christine Korsgaard points out that being compelled to "acknowledge that your desires have the status of reasons for you, in exactly the same way that mine do for me [...] does not force me to share in *your* reasons, or make your humanity normative to me."[70] Korsgaard does not deem an argument based on the universalisation of self-interested and essentially private reasons to be sufficient to furnish these reasons with an element of prescriptiveness for others. Rationalist arguments as Gewirth's, she holds, only point out to agents obligations they have to *themselves* to treat others in accordance with what they must recognise as adequate with regard to their agency. Genuinely moral reasons, according to Korsgaard, need to provide an individual with obligations owed to her fellow individuals, not herself.[71] They therefore cannot emerge from private ones which are ultimately rooted in an individual's self-interested regard. Moral reasons need to contain a regard for her interlocutor from the beginning, since "the gap from private reasons to public ones cannot be bridged by argument."[72]

Rather, Korsgaard argues that reasons are never private, but essentially public by nature.[73] As soon as reasons are being exchanged, they cannot remain subjective, but need to abstract from the individual's particular standpoint, in order to make the respective individual's reasoning accessible and intelligible to others. In other words, an individual needs to adhere to the norms of rationality when substantiating her claims in confrontation with others. Yet, does not all reasoning by which a person obligates *herself* to believing, claiming, or doing something take this form, irrespective of whether it is merely conducted in private or in public? Do we not always need to reason *as if* we were reasoning in public, *as if* we were required to be intelligible to others? Korsgaard rejects the idea that our consciousness can ever be private, since our standards of reasoning are socially determined.[74] In other words, we can only be intelligible to ourselves in the same way that we are intelligible to others. A reason we recognise as having ourselves thus can never only be considered a reason *for* ourselves. What we recognise as giving us a reason is the force of rationality. If we cannot say that rationality is private, we cannot claim that our reasons are. For, if the normative force of rationality is shared, "[t]o act on a reason is already, essentially, to act on a consideration

---

[70]Korsgaard, "The origin of value and the scope of obligation", p. 134, original emphasis. See also Bernard Williams' critique of Gewirth's account in *Ethics and the Limits of Philosophy*, chap. 4.

[71]Ibid., p. 134.

[72]Ibid., p. 134.

[73]Ibid., pp. 134-135.

[74]Ibid., pp. 136–138.

whose normative force may be shared with others."[75] The same conditions must give rise
to the same conclusions, irrespective of the subject to whom they apply. This also means
that when talking about the reasons we recognise ourselves to have, we are never merely
talking about ourselves in a first-personal sense, but always about the reasons that rationality
requires us to recognise as applying to *a* person with the very same attributes, and thus to
all such persons, under the very same circumstances. When talking about the reasons we
have, we thus always talk in abstraction – in abstraction from the person herself, considering
nothing but the relevant factors that commit *a* person to make a certain claim or endorse a
certain belief.

Thomas Nagel also stresses the requirement of abstraction, arguing that the reflective self
must reason on more universal terms, since only by withdrawing from the individual, first-
personal perspective it can achieve the "self-conscious awareness"[76] necessary to reflect on
what is being required of a person, given who and what she is.

> The reason we can no longer decide from the purely local perspective within
> which the original appearances or impulses are found, is that once we observe
> ourselves from outside, and achieve the distance of which Korsgaard speaks,
> our choice becomes not just what to believe or do, but what *this person* should
> believe or do. And that has to be a decision about what any person so situated
> should believe or do, since the external view does not give any consideration to
> the fact that the person is me – it describes me in terms which would be just as
> available to someone else sufficiently well informed about me.[77]

It is thus the abstract attributes of a person that guide our reasoning, irrespective of any
first-personal attachment.

In practice, this means that if a person recognises that it is the necessity arising from her
agency to value her purposes that gives her (an inevitable) reason to demand others to respect
the generic goods of her agency – which, due to its inevitableness, she cannot but want to
be binding – she needs to recognise the bindingness of the very same demand on the part of
all other agents. The publicity of reasons prevents the agent from willing her own demand
of respect to impose an obligation upon others – to be law to others – without also willing
all agents' similar demands to be law to others, including herself. As Korsgaard emphasises,
"if you are law to others in so far as you are just human, just *someone*, then the humanity of
others is also a law to you."[78] What is crucial is that as soon as a person is confronted with
both her first-personal and her abstracted self, the former can hardly reject the bindingness
of the latter's demand. That is, it cannot reject all other agents' demand on all other agents

---

[75]Korsgaard, "The origin of value and the scope of obligation", p. 136.
[76]Thomas Nagel. "Universality and the reflective self". In: *The sources of normativity*. Ed. by Onora
O'Neill. Cambridge: Cambridge University Press, 1996, p. 203.
[77]Ibid., p. 203, original emphasis.
[78]Korsgaard, "The origin of value and the scope of obligation", p. 143, original emphasis.

to respect the generic goods of their agency, since in doing so, the person would reject a demand she is committed to herself first-personally and would thus be contradicting herself.

At this point, it becomes apparent that Korsgaard's argument for the publicity of reasons and their intrinsic regard for those who issue demands based on public reasons is crucially similar to Gewirth's requirement of universalising one's claims. Contrary to Korsgaard's criticism of Gewirth's argument, the reasons an agent recognises herself to have for demanding others' respect for the generic goods of her action do not remain private as soon as the agent is forced to universalise her claim and to abstract from her first-personal self. In considering what she must deem an appropriate demand for an agent to make, she certainly asks herself what *her* agency requires her to demand, but the outcome of her deliberation depends upon the rational forces that the concept of agency (which is shared) exerts upon her reasoning, rather than upon the fact that *she* happens to be an agent herself. To universalise means to subject one's reasoning to publicly, *i.e.,* socially, recognised standards. In so far as all reasons we may consider to be meaningful must be universalisable (*i.e.,* lead us to draw similar conclusions under similar circumstances), all reasons are indeed public, as Korsgaard holds.

With regard to our initial question of what moral commitment can be said to be entailed by the factual recognition of another person's agency, these considerations on the publicity of reasons may lead us to conclude the following. A person who recognises herself as an agent is committed to assigning value to the generic goods of her actions, to demanding of others to respect said goods, and more generally speaking, to demanding of others to take her concerns into account when deliberating themselves on potential courses of action. As these requirements necessarily apply to a person by virtue of her quality as an agent, which becomes evident to her upon reflection, she cannot eschew the necessity to claim some weight in other individuals' deliberations on how to act towards herself. Hence, she cannot eschew demanding to be granted moral concern by others. Upon recognising this reason, however, due to its publicity, she cannot deny the legitimacy of the same claim by relevantly similar individuals, that is, other agents. In other words, she cannot deny to them the same moral concern she must claim for herself.

Now, it is of course possible for an agent to deny the publicity of reasons, to claim that what constitutes a reason in one's own case does not need to be recognised in relevantly similar cases of other people. However, such a denial can only be sustained at the cost of being unintelligible. Given that it is only the sharedness of reasons, their being subject to the normative force of rationality, which allows their meaningful exchange among individuals, a denial of the epistemic bindingness of these norms is equivalent to a surrender to obscurity – to unintelligibility. To do so also demonstrates an individual's incapacity or unwillingness to engage in a meaningful exchange of reasons with others. Irrespective of whether such an attitude is defensible on a more fundamental meta-ethical level, it is an intolerable position to take within the public political discourse among constituents of justification, for it is part of the very purpose of communication within this realm to make oneself and one's proposals

intelligible to others.

Whether or not one may permissibly remain unintelligible to others in determining of one's moral attitudes is a more fundamental question of moral epistemology which is beyond the scope of this work. This is not detrimental to its purpose, as I am not concerned with the question of whether people need to grant others equal moral concern in general, but solely whether they can be considered to have reason to do so if they *were* participants in public justification. It is their role as (aspiring) members of the constituency of public justification, as defined by the procedural norms which would apply to them *if* they were admitted to said constituency, that requires them to make themselves and their reasons intelligible to their co-citizens. This, in turn, demands adherence to shared standards of rational reasoning. For their part, these standards do not allow unreasonable citizens to sustain a denial of the moral equality of all other citizens to whose agency they can be said to be committed, while also recognising themselves as agents and thus as subjects who cannot but demand to be granted moral concern by other agents.

With these commitments and requirements in place, there is indeed a basis for concluding that in their capacity as potential actors in a public political discourse which stresses the importance of justification, unreasonable citizens may be said to *have* a reason to regard their fellow citizens as moral equals. This reason emerges from the implicit assumptions that their conduct in processes of public justification is premised upon. Only by admitting to these assumptions about the agency of their co-citizens could unreasonable citizens be said to be engaged in meaningful and intelligible justificatory interaction in accordance with the procedural norms governing public justification. Rationality requires them not to deny the agency and, as this final section aimed to show, the moral equality of their interlocutors in public justification as long as they are engaged in this process. Consequently, unreasonable people, by virtue of the satisfaction of their aspiration to become members of the constituency of public justification, could not but deem themselves to be publicly committed to considering those views which are fundamental to their unreasonableness (*i.e.,* their rejection of the moral equality of persons) to be rationally inadmissible within public justification. In other words, given that they are rational, we could consider unreasonable citizens to be committed to rejecting their quintessentially unreasonable attitudes if they were permitted to engage in processes of public justification. An unreasonable but rational person thus cannot sensibly deny that the constituency of public justification must exclude unreasonable doctrines from being pursued within said constituency due to their rational indefensibleness in meaningful, justificatory interaction.

4.7   Conclusion

The line of reasoning presented in this chapter allows us to attribute to unreasonable citizens a weakly externalist reason to recognise reasonableness as a valid criterion for regulating ac-

cess to the constituency of public justification. The liberal commitment to the moral equality of persons underlying reasonable attitudes[79] can be shown to be openly justifiable to an unreasonable person, given its rootedness in her own conduct as a hypothetical member in the constituency of public justification. The premises implicit in said conduct must be deemed to be both accessible and necessary in that they cannot be considered to be rationally deniable by the actor if *her* activity is to be recognised as a meaningful instance of those actions she is supposed to perform. It is for this reason that the argument I presented throughout this chapter may be assumed to achieve the assent of unreasonable, yet rational persons. Liberals may therefore consider unreasonable citizens to have reason not to reject reasonableness as an adequate criterion for admission to the constituency of public justification. That is, they may assume this criterion of reasonableness to be justifiable to unreasonable citizens as the grounds for their exclusion from said constituency.

---

[79]See section 2.2.3.

Chapter 5

Political liberalism and deep moral disagreement

5.1 Introduction

In the previous chapters, I have so far been concerned with unreasonableness of a very fundamental kind: with those citizens who reject values that are central to the liberal project, specifically the ideal of persons as free and equal. There is no question that their views have no place in the constituency of public justification of any liberal society. The question which I have so far addressed was how to square the exclusion of these fundamentally unreasonable people with the fact that liberalism is built on a foundation of conceiving of *all* citizens as free and equal and must, consequently, seek to justify coercion to all those affected by it. Since their exclusion from the constituency of public justification is ultimately a coercive act, liberalism must show that they have reason to accept the core dimension of reasonableness: a commitment to persons as free and equal.

In the final two chapters, I will examine a different kind of unreasonableness. I will consider the case of another set of people who find themselves outside the constituency of public justification, not for rejecting the ideal of persons as free and equal as such, but for more subtle infractions of the norms of reasonableness which derive from that core ideal. Merely showing that the freedom and moral equality of all persons is an ideal that is acceptable to all will be insufficient to justify exclusion from the constituency in these cases. This is because in these cases, the individuals in question *are* committed to that ideal. Yet, they face conflicts between demands that derive from that commitment and those of their other, private moral commitments. Compared to those who hold fundamentally unreasonable views, the infractions of the norm of reasonableness that are driven by those conflicts may be far less severe. Nevertheless, if treating persons as free and equal is the core standard of reasonableness, these infractions render those who commit them less than fully reasonable, as I will argue. By the standards of political liberalism, this disqualifies them from membership of

the constituency of public justification.

At this point, I deliberately leave open the question of the *extent* of these less than fully reasonable citizens' exclusion from the constituency of public justification, which should probably be proportionate to the severity of their infractions on the norm of reasonableness. At a minimum, they must be excluded from the constituency of public justification with respect to those matters which gave rise to the charge of unreasonableness in the first place. That said, the precise extent of their exclusion is of no consequence here, given that even partial exclusion is in tension with the liberal tenet which took centre stage when I made the case for offering justifications to fundamentally unreasonable people: that a commitment to treating all people as ends in themselves requires that state coercion must be justifiable to all.

The key question I will address in this chapter is whether there is a way in which political liberalism can restrict the constituency of public justification to those who are *fully* reasonable without running afoul of said commitment. In other words, can political liberalism justify the exclusion of those who fail to be fully reasonable in terms of reasons that those who are being excluded can accept? I will argue that political liberalism deprives itself of the philosophical resources to do so. It lacks the resources to adjudicate on a theoretical level between individuals' shared liberal commitments specifically and the moral commitments rooted in their private comprehensive moral doctrines more generally. I will then go on to argue that political liberalism cannot resolve that tension by referring to the idealised character of the citizens populating its theoretical constituency of public justification. Echoing an argument which I have first sketched in chapter 2, a highly idealised conception of the person of the kind political liberalism requires in this context is incompatible with the liberal project. Again, this leaves political liberalism with a justificatory gap, which I ultimately aim to close in chapter 6. I will begin this chapter by discussing what it means for citizens fall short of full reasonableness, as well as the circumstances under which this is likely to occur.

5.2 Reasonableness and deep moral disagreement

5.2.1 Failure to be fully reasonable

In which sense precisely can citizens fall short of being fully reasonable? Reasonable people, regarding each other as free and equal, realise that a society which embodies these values must be a cooperative one. Members of such a society, who are aware of their fellow citizens' likewise commitment to cooperation, will not see any moral disagreement that persists after debate has weeded out any factual errors or blatant mistakes of reasoning as an indication of disrespect. If all are committed to engaging and deliberating with each other out of respect for their fellow citizens' status as free and equal, persuading each other of the merits of their proposals and be persuaded whenever they can, any lack of success on that account

cannot be explained by an unwillingness to find common ground for political decisions. Among people who see each other as free and equal, and hence see society as a cooperative endeavour, failure to come to an agreement on a contested matter does not render people unreasonable in the eyes of their fellow citizens. If individuals cannot think of their co-citizens as merely mean-spirited or obviously incompetent reasoners, they must explain their persistent disagreements by other means. In political liberalism, reasonable people account for such disagreements with reference to what Rawls calls the *burdens of judgment:* these burdens are "the many hazards involved in the correct (and conscientious) exercise of our powers of reason and judgment in the ordinary course of political life".[1] They include, but are not limited to, the complexity of empirical and scientific evidence, the vagueness and indeterminacy of all concepts, and discrepancies of individuals' overall life experience which affect their assessment and balancing of moral and political values in a manner and to a degree which individuals themselves cannot account for.[2]

How are people who recognise these burdens supposed to act in the public political sphere? If we recognise that small differences in the information we have, the weight we assign to specific values, and other factors we cannot consciously and actively account for in our reasoning can lead to substantial differences in our considered moral judgements, it seems that we must not expect others to agree with our conclusions on what would be the right thing to do.[3] Therefore, acting our part as members of a cooperative society cannot simply boil down to ineffectively attempting to convince others of the merits of our own position when no common ground is in sight. As Rawls argues, "those who insist, when fundamental political questions are at stake, on what they take as true but others do not, seem to others simply to insist on their own beliefs when they have the political power to do so."[4] He insists that "reasonable persons see that the burdens of judgement set limits on what can be reasonably justified to others".[5] Instead, we must refer to *shared* values to justify our political decisions and refrain from drawing justifications from the content of individuals' private comprehensive moral doctrines which are certain to be contested. Rawls concludes that "[t]here is no reason [...] why any citizen, or association of citizens, should have the right to use the state's police power to decide constitutional essentials or basic questions of justice as that person's, or that association's comprehensive doctrine directs." Thus, in virtue of their acceptance of the burdens of judgement, reasonable people must never insist that their *non-public* reasons – that is reasons which are not shared by all members of society – are appropriate grounds for adopting a particular policy. Instead, they seek shared foundations to justify political decisions. For reasonable persons who are committed to the idea of society as a cooperative endeavour, the burdens of judgement specify the meaning of cooperation.

By this standard of reasonableness, a person can fall short of being fully reasonable without

---

[1]Rawls, *Political Liberalism*, p. 56.
[2]See ibid., pp. 56-57
[3]See ibid., p. 58.
[4]See ibid., p. 61.
[5]See ibid., p. 61.

rejecting the value of freedom and equality of their co-citizens as such, as was the case with fundamentally unreasonable individuals. As I am going to argue, a person's failure to offer only shared reasons is not necessarily indicative of a lack of commitment to that value, if we take into account the full complexity of her moral commitments more generally. I will explore the experience of individuals who encounter moral conflicts which are rooted in deep moral disagreement and which induce them to diverge from the path of full reasonableness despite their commitment to the ideal of persons as free and equal. However, since within political liberalism, the constituency of public justification is the domain of the reasonable, all persons who fail to be fully reasonable must face exclusion from said constituency. In the further course of this chapter, I will argue that political liberalism must account for the exclusion of those who fail to be fully reasonable by means of justification, and that the nature of their unreasonableness requires justifications which are distinct from those I have proposed in the previous chapter.

5.2.2   Deep moral disagreement

It is the key contention of this chapter that fundamentally reasonable persons, who regard their co-citizens as free and equal and are thus committed to offering them only shared, public reasons for proposed policies, may sometimes have reason to discount that latter commitment. Instead, they may sometimes be tempted insist on a particular nonpublic moral reason as the only valid basis for political action. As I will show in this section, this temptation arises when individuals encounter conflicts which involve *deep moral disagreement*. By that term I refer to conflicts among citizens' moral convictions which are characterised by *inaccessibility* of their respective sources of moral authority, and, in addition, possess a *foundational quality*. The moral authority that a particular conviction has for Alice's actions is *inaccessible* to Bob if Bob does not share her commitment to the source of said authority, such as a particular religious or philosophical worldview. He cannot come to recognise Alice's conviction as a valid reason for him, since the reasons that she presents him with do not hold any moral authority in his view.

Alice's and Bob's conflict has a *foundational quality* if any of their conflicting convictions constitutes, or is closely tied to a foundational tenet of their comprehensive doctrine, such as for example, the sanctity of life in Christian religious doctrine. Given that relinquishing or compromising such a core conviction may fundamentally affect an individual's conscience, as well as her conception of herself as a moral person, conflicts of such a foundational quality do not easily lend themselves to pragmatic or principled compromise. What is imperative to a person in cases where foundational matters are at stake is to do the right thing, and the right thing in their eyes may well be to defend and act on their nonpublic moral conviction. For example, Catholic Alice may be highly reluctant to act against her private convictions about the moral status of human fetuses, which is rooted in her religiously grounded belief in the sanctity of all life. Under these circumstances, she may well be unimpressed by other reasons

such as political expediency (*e.g.,* Bob offering to vote in her favour on another issue), or by principles which she considers to be subordinate to her foundational conviction in this case (*e.g.,* Bob appealing to her promise at an earlier occasion to support him in a future policy proposal, or a commitment to the bodily autonomy of human persons in general).

For citizens to promote a particular policy on the basis of their nonpublic values is only ever problematic in cases where political liberalism considers advocacy for policies on the basis of nonpublic values to be unacceptable. Only in political liberalism's realm of *public reason,* where policies must be justified on the basis of public values that are shared among all citizens, would a person face a conflict between her allegiance to the principles governing political decisions and her foundational moral convictions. Hence, it is important to ask whether matters which are likely to be subject to deep moral disagreement fall within the realm of public reason. For Rawls, the purpose of public reason is to determine constitutional essentials (procedures and basic rights) and matters of basic justice.[6] These are matters where the chosen norms affect citizens in the most fundamental of ways, which is why it is imperative that they be acceptable to all. Within the consensus conception of public reason, this is ensured by requiring them to be justified with reference to shared, public values. In turn, this implies that not all public policy backed by coercive force must meet the high standards of justification by public values. As Andrew March observes,[7] "having one's own share of the collective disposable income diverted to something we would not have chosen is not the same kind of assault on our dignity and autonomy as having the criminal law brought to bear on our personal choices or being excluded for arbitrary reasons from public goods or institutions."[8]

Yet, which political matters precisely are fundamental enough to be subsumed under the label basic rights and basic justice is not obvious. Rawls indicates that questions pertaining to citizens' participatory rights in politics, liberty of conscience, freedom of thought and of association,[9] as well as basic matters of distributive justice[10] are sufficiently fundamental to be subject to public reason. In one example, Rawls discusses the interaction of public values on the question of abortion,[11] indicating that questions respect for human life and bodily autonomy are also within the scope of public reason. It is not my intention to settle the question of what matters should count as matters of *basic* rights and justice. Neither do I intend to give a comprehensive account of all conceivable issues that could be subject to deep moral disagreement. That said, many issues that give rise to deep moral disagreement

---

[6]Rawls, *Political Liberalism*, 227ff.

[7]Note that March's proposed taxonomy of such matters is orthogonal to Rawls's. Nevertheless, his comment aptly captures what is implied by Rawls's distinction between basic rights and matters of justice on the one hand, and less fundamental political decisions on the other: namely, that not all coercion must necessarily trigger the same justificatory response.

[8]Andrew F. March. "Rethinking the Public Use of Religious Reasons". In: *Rawls and Religion*. Ed. by Tom Bailey and Valentina Gentile. New York: Columbia University Press, 2015, pp. 97–130, p. 121.

[9]Rawls, *Political Liberalism*, p. 227.

[10]Ibid., 228f.

[11]Ibid., 243, n. 32.

among citizens are likely to fall within the realm of basic rights and justice, if we take Rawls's examples as a rough indication of its scope. For example, this is certainly true of conflicts about bio-ethical matters which raise questions about what respect for (human) life and the person requires. These include, but are certainly not limited to, the regulation of genetic engineering, assisted suicide, and abortion. Other disagreements that draw on similarly fundamental values concern the morality of torture, the death penalty, and the ethics of war. The task of balancing the values of privacy, free speech, and national security is another basic matter that may well be subject to deep moral disagreements.

It should not come as a surprise that the sites of deep moral disagreement are likely to fall into the realm of basic rights and justice. Part of the depth of these disagreements – their persistence and immunity to political compromise – stems precisely from the fact that they concern decisions that some citizens see as *basic* matters of rights and justice. Decisions on these matters affect current and future members of society in some of their most basic functionings, such as their capacities to exercise their bodily autonomy, liberty of conscience, or freedom of expression. Therefore, in the eyes of some citizens, these are decisions that any adequate political system must get *right*. If they accepted any less, the resulting moral wrong would be too grave to make compromise a satisfactory option.

Why does such deep moral disagreement present a challenge to political liberalism? One could argue that reasonable persons share a commitment to political cooperation in virtue of their respect for persons as free and equal, and that a resolution which has been achieved by way of public reason or other cooperative procedures should therefore possess legitimacy in the eyes of those who are reasonable. If we want to know whether political liberalism can successfully address deep moral disagreements, we must ask precisely how this commitment can claim priority over individuals' foundational moral commitments which are rooted in their private comprehensive doctrines. More specifically, do citizens need to defer their commitment to those values which they merely hold privately – no matter how pressing their demands – to their allegiance to those values which are shared by all? This is the crucial question that this chapter aims to address.

Among the confrontations of shared public and foundational private reasons, we can distinguish between those that involve substantive public values that lend support to a particular policy, and those in which the public value invoked is the shared commitment to mutual cooperation in the public sphere. The latter situation arises when public reason proves to be incomplete, that is, when the set of reasons which are publicly shared does not provide any guidance with regard to some contentious issues of public policy. Under circumstances of incompleteness, procedural strategies, premised merely on a commitment to cooperation, present themselves as politically neutral methods to resolve the conflict.[12]

I will consider such cases of incompleteness in section 5.3.1, assessing whether reasonable

---

[12]See, for example, Micah Schwartzman. "The completeness of public reason". In: *Politics, Philosophy & Economics* 3.3 (2004), pp. 209-114, and Andrew Williams. "The Alleged Incompleteness of Public Reason". In: *Res Publica* 6.2 (2000), pp. 210-211.

citizens' commitment to cooperation in the political sphere does override their private commitments when they are in deep moral disagreement with their co-citizens' policy proposals. Examining procedural solutions for overcoming incompleteness, I will highlight their dependence on a higher-order norm of cooperation in section 5.3.2, before challenging the sustainability of such a norm within the framework of political liberalism in section 5.4. I will then proceed to argue (in section 5.5) that the structural features of political liberalism which render such a norm unsustainable do not only affect cases of incompleteness, but also political liberalism's capacity to address deep moral conflicts between *substantive* public and nonpublic values.

5.3 Deep moral disagreement and incompleteness

5.3.1 Indeterminacy, inconclusiveness, and gaps of legitimacy

As I argued in chapter 2, a commitment to the ideals of individual moral autonomy and self-determination lies at the very heart of liberal political theory. The ideal of public reason frames the conditions which the public political discourse must meet in order for potentially coercive decisions to be compatible with this commitment. What should be the bases and conditions of such discourse, if it is to ensure that individual citizens can be said to recognise political decisions as the object of their autonomous judgement? To this end, consensus conceptions of public reason emphasise the value of mutual justification based on shared grounds. Individuals are expected to be able to explain to their fellow citizens their reasons for supporting a given policy on terms which their interlocutors can *also* recognise as applicable to themselves. Hence, the need emerges for public reason to be conducted on the basis of a set of values that are shared by all. Rawls may be interpreted to endorse a consensus conception of public reason[13] when he argues that

> [...] our exercise of 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 them as reasonable and rational. [...] [T]he ideal of citizenship imposes a moral, not a legal duty – the duty of civility – to be able to explain to one another on those fundamental questions how the principles and policies they advocate can be supported by the political values of public reason.[14]

If, broadly speaking, the decisive reason for adopting a particular policy is shared by all citizens, each individual citizen can be said to have a reason to support the policy in question.

---

[13]Further proponents of a consensus conception include Jonathan Quong (*Liberalism without Perfection*, pp. 261-273) and Stephen Macedo ("Why Public Reason? Citizens' Reasons and the Constitution of the Public Sphere").

[14]Rawls, *Political Liberalism*, p. 217.

While it may well be that said reason would not have been the most salient one for some citizens had they alone been asked to decide on the matter, their recognition of the need for a collectively justifiable compromise on a contested matter and the fact that they attribute *some* value to said reason should allow them to regard that decision as worthy of their reasoned support.[15]

This source of legitimacy for political decision making is, however, lost if the shared values providing such public reasons turn out to be *incomplete.* Incompleteness in public reason occurs if shared values fail to guide citizens towards a reasoned consensus of judgements on a proposed policy. Considering more closely the roots of such failures to reach a reasoned consensus based on shared reasons, it is useful to differentiate between those cases in which public reasons are *inconclusive* and those in which they are *indeterminate*.[16] In cases of inconclusiveness, shared values do indeed yield support for a variety of different policies, but do not provide *decisive* reasons to choose one over the others. As Schwartzman explains "[t]his happens when citizens justify their political decisions on the basis of public reasons but disagree with one another about which of their positions is most reasonable."[17] In cases of indeterminacy, shared values do not provide any sufficient reason for choosing a particular policy,[18] either because they do not yield support for any position at all, or because they can be interpreted to advocate a number of mutually exclusive positions (nested indeterminacy).[19]

At first glance, failures to resolve instances of indeterminacy appear to be a more obvious threat to public reason's capacity to confer legitimacy on a given policy than persistent conflicts characterised by inconclusiveness. This is because the former allow for unmediated deep moral disagreement. Where public reason is silent because publicly shared values are indeterminate, disagreements on what policy to choose are necessarily rooted entirely in nonpublic reasons. That said, such disagreements do not have to be deep moral disagreements. Deep moral disagreements are, as previously defined, characterised by a resistance to compromise due to the foundational nature of the values at stake and the inaccessibility of

---

[15] See Quong, *Liberalism without Perfection*, pp. 208, 212.

[16] This distinction has been introduced by Gaus in *Justificatory Liberalism: An Essay on Epistemology and Political Theory*, 151–158. See also Schwartzman, "The completeness of public reason".

[17] Ibid., p. 194.

[18] Ibid., p. 196.

[19] See Gaus, *Justificatory Liberalism: An Essay on Epistemology and Political Theory*, p. 156 and Schwartzman, "The completeness of public reason", p. 197. Matters are further complicated by potential disagreement on whether a particular conflict in public reason is characterised by inconclusiveness or indeterminacy. For instance, individual persons may well disagree on whether or not some shared value can be interpreted in such a way that it allows for substantive conclusions to be drawn with regard to a particular contested issue. For example, liberal values are sometimes considered to be indeterminate with respect to the moral status of fetuses, given that they are said to merely demand equal respect for human *persons* as opposed to respect for human life as such. However, as Jeremy Williams argues, the apparent silence of liberal values with regard to human life which does not amount to personhood must not be interpreted as indeterminacy. Rather, it should, in isolation from other considerations, be assumed to signify a deliberate rejection of the status of non-persons (Jeremy Williams. "Public Reason and Prenatal Moral Status". In: *The Journal of Ethics* 19.1 (2015)). The same controversy arises with respect to the question of the indeterminacy of liberal values on the moral status of non-human animals.

their respective sources of normative authority. We may well imagine citizens to disagree on the basis of foundational values whose respective normative authority remains inaccessible to others, yet still discount the significance of said values with regard to a decision on the policy in question. Individuals may be deeply unconvinced by their interlocutor's substantive case for said policy, yet decide to endorse it for pragmatic reasons or in order to live up to a higher-order commitment to political cooperation.[20]

Consider the following example: Alice, a practising Catholic, and Bob, a humanist, disagree fundamentally on whether animals should be granted some protective rights, such as rights against the deliberate infliction of pain short of human self-defence. Alice, drawing on her commitment to the sanctity of life as such, sees value in animals' dignity and hence their freedom from pain and torture, and therefore privately supports legislation which would grant them these rights. Bob, on the other hand, whose private comprehensive moral doctrine does not extend its scope to animals, rejects such a value. Looking for shared reasons to support or reject a policy that would impose far-reaching restrictions on humans' treatment of animals, they conclude that publicly shared liberal values can merely be said to be concerned with the value and rights of *human* persons. Hence public reasons are indeterminate on this particular question, which is not concerned with human persons. Their private disagreement is – at least as far as Bob is concerned – characterised by an inability access their interlocutor's sources of normative authority. Still, their disagreement does not need to be foundational, as long as neither of them considers their respective nonpublic values to be of ultimate significance to the issue at hand.

Consider the case of a proposed ban of the use of animals in circuses. Alice's and Bob's shared commitments to the dignity of the human person are indeterminate in this case, while Alice's private commitment regarding the dignity of animals and Bob's belief in the cultural value of the use of animals in performances clash. It is, however, at least conceivable that Alice does not value her private belief in animal dignity more than the prospect of an alternative, cooperative solution to the conflict. She may thus find it acceptable to agree to decide the matter by means of a vote. That is, Alice may well be prepared to discount her substantive moral concerns when presented with the possibility of a solution which avoids prolonged conflict. This may, for example, be the case if she believes that it will be more promising to advocate a less radical policy which merely sets high standards for the welfare of animals in circuses, or, pragmatically, if she hopes that by accommodating Bob in this case, he will be more likely to support her on a different, more important matter at another time. That said, their disagreement *can* be characterised as deep moral disagreement if at least one of them insists that their respective commitments with respect to the moral status of animals are foundational to their conception of what morality requires of them and must therefore not be discounted.

---

[20]Whether or not political liberalism can be said to prescribe such a higher-order commitment will be central to my discussion on whether indeterminacy can legitimately be resolved by procedural means. At this point, however, I merely state that a person may herself believe to be bound by such a commitment.

In more general terms, when deep moral disagreement occurs under circumstances of indeterminacy, proponents of one particular policy are not only unable to recognise any genuine substantive value in their opponents' proposal, but also care too much about their own respective nonpublic commitments to discount them. In these cases, any political decision which does not pay heed to these latter commitments may crucially affect citizens' belief in the fundamental justice of the political institutions and the legitimacy of their exercise of power. As shared substantive reasons are unavailable to bridge deeply divergent private convictions about the appropriate course of action, indeterminacy ultimately leaves individuals' substantive moral judgements on those policies to be guided by their nonpublic private reasons alone. The policy which ultimately emerges from political procedures may not lay any *substantive* moral claim to all citizens' allegiance. In cases of indeterminacy, deep moral disagreement hence seems to reveal gaps in the capacity of public reason frameworks to ensure the legitimacy of public policies.

Such gaps may appear to be less likely to occur in cases of inconclusiveness in which all qualifying policies are supported by substantive moral values that are shared by the constituents of the realm of public reason. This means that individual citizens recognise the moral reasons supporting any of these policies as worthy of their allegiance and hence as generally valid grounds for action. As Schwartzman argues, if we further consider "the limitations of practical reason"[21] to which individual citizens are subject,[22] it seems "far too demanding to require that they act only on the best of all possible reasons".[23] In other words, inconclusive public reason yields policies which should be recognised as reasonable enough by all members of the constituency of public justification because they are rooted in shared values, and citizens understand that limitations of human reasoning diminish the chances for a collective consensus on a single, most reasonable policy.

But even in theses cases, disagreement may be fuelled by deep, foundational commitments and hence affect citizens' perceptions of a given policy's legitimacy. Even if a person generally endorses all the publicly shared reasons that lend support to different conflicting policies, she may not only have some preference for a particular reason and the course of action it recommends. Her assessment of the respective validity or relevance of any of these values with regard to the particular case at hand may well be rooted in more foundational considerations. These considerations may render a given public reason not only inferior to other conflicting public reasons, but discount or even deprive it entirely of its normative force in that particular instance.

To illustrate, imagine Alice, who is committed to the shared values of freedom of expression and the dignity of persons. On the inconclusive issue of restricting freedom of expression by

---

[21]Schwartzman, "The completeness of public reason", p. 199.

[22]Schwartzman mentions the "complexity of the evidence, the need to balance and sometimes trade off values that are vague or otherwise difficult to interpret, and the differences in life experience which inform the judgements that people make about such issues", as well as "limited time to process relevant information" (ibid., p. 199).

[23]Ibid., p. 199.

outlawing some kinds of speech which are thought to violate some individuals' or groups' dignity (such as racist or sexist speech, or other instances of what is often classified as "hate speech"), she attributes a higher value to freedom of expression and is hence inclined to reject any policy which proposes to restrict it. In fact, it is her particular private interpretation of the dignity of the person which sustains her opposition to restrictions of free speech: a conception of dignity which conceives of individuals' capacity to express their autonomy in active self-determination as the crucial trait of personhood and thus deserves of the highest protection. In this case, it is not obvious that Alice's general recognition of the value of human dignity will sufficiently console her to consider restrictions on freedom of expression to be worthy of *her* support. In fact, she may hold the certainly controversial view that permitting speech of any kind is far less dangerous to the dignity of persons than any infringement on the principle of freedom of expression. Under these circumstances, Alice's concern for the dignity of persons can hardly serve as a reason to justify restrictions on freedom of speech to her. It is therefore questionable whether inconclusiveness in public reason is, unlike indeterminacy, immune to gaps in public reason's capacity to confer legitimacy to political decisions. Both forms of incompleteness may lead to situations in which some citizens' actions may become subject to restrictions by political decisions which cannot faithfully be rendered as acts of morally autonomous self-determination. As with cases of indeterminacy, inconclusiveness may also conceivably lead to situations in which public reason fails to live up to the liberal ideal of protecting individual moral autonomy.

All this is not to say that we should disregard the distinction between indeterminacy and inconclusiveness. There are, however, reasons to believe that, at least in some cases, deep moral disagreement may affect both of these dimensions of incompleteness in a way that threatens public reason's capacity produce policies which can claim legitimacy on the grounds that each citizen can be said to have genuine, substantive reasons to accept them.

5.3.2 Procedural strategies

The lack of determinate or conclusive shared reasons cannot simply translate into legislative restraint. In some cases, the moral acceptability of the status quo is as controversial among citizens as any policy that would change it. In other words, the status quo is not necessarily morally neutral. Whenever legislative restraint translates into the permissibility of some morally charged actions, such restraint may be as unacceptable to some citizens in certain cases as any particular decision would be in others. Hence, there is a reason to look for alternative procedures for resolving the legislative deadlock of incompleteness which do not require the availability of shared substantive reasons. Schwartzman argues that "citizens should not abrogate their duty of civility even when it is difficult or impossible for them to decide how to resolve political issues". He proposes strategies which enable citizens "to respond to incompleteness without sacrificing the weighty moral and political values that

support their commitment to public justification."[24] He argues that "rather than impose their nonpublic reasons on others, citizens can choose to submit their disputes to various forms of procedural adjudication."[25] In this context, Schwartzman considers both democratic[26] as well as arbitrary decision procedures, such as lotteries.[27] Though they are not sensitive to citizens' substantive reasons for preferring one policy option over another, these second-order decision making procedures nevertheless reflect other shared values, such as, for example, fairness:[28] both of these kinds of procedure equally grant a platform to all citizens and attribute equal weight to each citizen's preferred policy.

There is no doubt that these virtues can generally be assumed to be of value to liberal citizens. However, whether citizens *must* consider the *outcomes* of such a cooperative solution to be acceptable, despite their conflicting substantive reasons to reject it, is not so obvious. Their cooperation can only be expected by liberal theory *in general* if citizens can be said to share *higher-order* commitment to mutual cooperation, which would give each individual a reason to discount the status of their private substantive concerns when public reason is deadlocked. Given liberalism's core commitment to respecting individual persons' moral autonomy, the status which *they* assign to their convictions cannot be dismissed lightly. Without a higher-order commitment to cooperation, procedural solutions are not *guaranteed* to prevail in their moral deliberations on how to act in light of their other substantive moral commitments.

In other words, without such a higher-order moral commitment to cooperation, procedural strategies and their outcomes are not guaranteed to be *morally* effective. That is, individual citizens may well fail to recognise them as a relevant input to their moral deliberations when assessing the acceptability of a given policy. This individual act of *judgement* is fundamental to a liberal conception of legitimacy which is committed to protecting and fostering individual autonomous self-determination. As I have argued elsewhere,[29] a distinguishing feature of liberal political theory consists in its commitment to respecting the individual person as an agent, respecting and protecting the capacities she is endowed with in virtue of her moral autonomy.[30] Hence, the moral effectiveness of any given strategy, its capacity to affect an individual's distinctly moral judgement as to whether a particular policy is compatible with her ends, is indeed of crucial importance.

One such potentially morally significant reason for attributing value to procedural solutions consists in a shared commitment to resolving conflicts in public reason in a fair and cooperative manner, rather than ultimately allowing the matter to be decided by the respective dis-

---

[24]Schwartzman, "The completeness of public reason", p. 208.

[25]Ibid., p. 211.

[26]Ibid., p. 211.

[27]On the use randomising devices for resolving indeterminacy, see also Williams, "The Alleged Incompleteness of Public Reason".

[28]See Schwartzman, "The completeness of public reason", p. 213

[29]See chapters 2 and 3.

[30]See, for example, Waldron, *Liberal rights: collected papers, 1981–1991*, pp. 36, 62. What liberalism recognises as crucially valuable to an individual is her capacity as a subject to determine her actions according to her ends (ibid., p. 41).

tribution of power and resources among the proponents of conflicting positions. Importantly, as I have argued, in order for the results of cooperation to take precedence over citizens' private moral convictions, they must attribute to their commitment to cooperation the status of a higher-order norm.

While this premise is not made explicit, there is nothing to suggest that the proponents of procedural strategies for resolving incompleteness would *deny* that these strategies are premised on such a commitment to cooperation being shared among citizens. Neither does said premise appear to be overly controversial, given that political liberalism specifically highlights that the willingness to cooperate with others on fair and mutually acceptable terms is a core characteristic of reasonable persons. My reason for highlighting this premise, however, is precisely to question whether reasonable citizens' general commitment to mutual cooperation can be used to substantiate a norm, which may require them to discount their private substantive moral convictions. In the next section, I will examine the position of the commitment to cooperation within the theoretical framework of political liberalism, arguing that political liberalism lacks the resources to assign to it the status of a higher-order norm.

5.4 Cooperation and the structure of political liberalism

There is no doubt that a shared commitment to cooperation is an integral element of political liberalism. A core characteristic of a reasonable citizen is her willingness to cooperate with others on terms acceptable to fair and reasonable citizens such as herself. According to Rawls,

> [p]ersons are reasonable in one basic 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. Those norms they view as reasonable for everyone to accept and therefore as justifiable to them; and they are ready to discuss the fair terms that others propose.[31]

In their political conduct, reasonable people are conceived as restricting themselves to draw upon what they have found to be mutually acceptable, striving for governance based on shared political principles, rather than insisting on the validity of their individual private convictions as suitable norms for regulating the public sphere. Individuals qualify as members of the constituency of public justification precisely because they are committed to the idea of society as a forum and project of mutual cooperation. Given that they are selected based on said commitment, it could be argued that they can be expected to discount the significance of their individual nonpublic substantive reasons for accepting or rejecting a given

---

[31]Rawls, *Political Liberalism*, p. 49.

policy in cases when insisting on their priority would be an obstacle to achieving a decision which is acceptable to others as well. In other words, according to this interpretation, citizens' commitment to a *higher-order* norm of cooperation constitutes an integral element of the basic theoretical framework of political liberalism. Following this line of argument, it seems that we can assume that procedural strategies aimed at resolving the legislative deadlock arising from incompleteness are indeed justified and rendered morally effective by an appeal to the higher-order status of one of the most fundamental of shared values.

This interpretation of reasonable persons' commitment to cooperation, however, fails to be backed by a closer analysis of the specific sources of normative authority that political liberalism relies on. Asking why individuals recognise such a commitment to be normatively authoritative to them, political liberalism refrains from attributing to it the status of an independent, foundational moral truth, or of a derivative of such a truth. In fact, this is the very claim which *political* liberalism seeks to avoid. One of the virtues of political liberalism – as opposed to a morally comprehensive liberal theory – consists in its compatibility with a wide range of comprehensive moral doctrines, which citizens individually recognise as sources of moral authority for themselves. Political liberalism requires these doctrines to be compatible with and supportive of a commitment to mutual cooperation. But said commitment does not necessarily have to be *foundational* to these doctrines. Within political liberalism, the value of cooperation may be rooted in other elements of a person's comprehensive doctrine.

Political liberalism draws on these comprehensive doctrines as sources of normative authority for its core principles. Liberal principles' claim to legitimacy rests on the idea that each citizen has an individual comprehensive reason to embrace them, which *she* considers to be morally authoritative. It is by steering clear of publicly rooting its core ideals in a particular comprehensive moral doctrine that modern liberal political theory can widen the scope of its appeal. It thus allows citizens to commit themselves to honouring liberal values in the political sphere for sets of reasons that are specific to each person, rooted in her particular understanding of the good life. To put it in Rawlsian terms as set forth in *Political Liberalism*, liberal values draw their legitimacy from being subject to an overlapping consensus, gathering support from a variety of comprehensive moral doctrines which are compatible with said values.[32] Liberalism thus conceived does not need to assume the same status as a comprehensive doctrine. Rather, liberal values and principles can be framed as free-standing, publicly detached from potentially divisive roots in comprehensive moral doctrines,[33] while remaining individually rooted in a variety of such doctrines, which furnish them with normative authority for each person.

For the individual, these liberal values may assume the status of moral truths or commands which they freely recognise as normatively authoritative for them. Fundamentally, the normative authority of liberal values and principles is drawn from the fact that citizens have reasons to accept them. Hence, political liberalism conceives of reasonable individuals'

---

[32]See Rawls, *Political Liberalism*, pp. 10-11.
[33]See ibid., p. 10.

public personality as rooted in their private comprehensive doctrines.[34] When applying their moral autonomy in giving or withholding their assent to a political principle, that decision is imbued with normative authority because, ultimately, it is supported by a reason they recognise as a source of normative authority over *them*. Political liberalism itself is neutral with regard to the diversity of such sources among reasonable citizens, as long as the particular principle they furnish with normative authority is the subject of an overlapping consensus. Collectively, liberal values must only acknowledge the political status of liberal values and principles as consensually chosen tools for regulating public affairs, which enjoy normative authority for each person in virtue of their support by her comprehensive doctrines.

This strategy relieves political liberalism of the necessity to account for the moral authority of its core principles by means of a specific, and potentially controversial, comprehensive moral theory, which would threaten its compatibility with a variety of comprehensive doctrines. At the same time, however, said strategy undermines the alleged higher-order status of the value of mutual cooperation, as I will argue in the following.

For the commitment to cooperation to gain the status of a higher-order norm, political liberalism's reliance on individual comprehensive doctrines as sources of normative authority for each individual requires the assignment of this status to be supported by said comprehensive doctrines. In other words, the assertion of a higher-order commitment to cooperation presupposes that, *within individuals' comprehensive moral doctrines,* said commitment enjoys precedence over other convictions they hold. While likely to be empirically accurate in many cases, this assertion is out of the theoretical reach of political liberalism. This is because its alleged normative authority to attribute a higher-order status to specific values depends on the very status that is actually attributed to those values by individual persons' comprehensive doctrines.

At first glance, political liberalism appears be able to prevent potential conflicts within individuals' comprehensive doctrines from surfacing on the theoretical level by refusing to attribute normative significance to the substance of such comprehensive doctrines themselves. Comprehensive doctrines are assumed to privately provide normative authority to the political principles that they support – such as a higher-order norm of cooperation – without themselves gaining substantive validation within the political sphere.

As Habermas observes with respect to Rawlsian political liberalism in *The Inclusion of the Other*,

> [o]nly the lucky convergence of the differently motivated non-public reasons can generate the public validity or "reasonableness" of the contents of this "overlapping consensus" that everyone accepts. Agreement in conclusions *results* from premises rooted in different outlooks. It is significant for the design of the theory as a whole that the participants can only register this convergence as a social

---

[34]See Quong, *Liberalism without Perfection*, p. 162 and Rawls, *Political Liberalism*, p. 12.

fact: "The express contents of these doctrines have no normative role in public justification."[35]

Political liberalism operates on the basis that the citizens have agreed to reason publicly without referring to their private comprehensive doctrines, and while they may individually acknowledge their individual substantive fundamental reasons for doing so, political liberalism as a whole cannot, given that its key appeal to modern pluralistic societies lies in its doctrinal autonomy.[36]

This is significant. If political liberalism publicly acknowledged the normative authority of individual comprehensive doctrines as such, it could, from a theoretical perspective, no longer simply assert the priority of publicly shared values over nonpublic ones which are all equally supported by the same individuals' comprehensive doctrines. However, in failing to attribute normative significance to the contents of individual comprehensive doctrines as such, political liberalism deprives itself of the very sources it relies on when claiming that any shared public value is recognised as normatively authoritative by citizens. Ultimately, it is the content of individual comprehensive doctrines which motivates individuals to recognise a particular value as normatively authoritative for them.

Hence, in order to assert the higher-order status of some values, political liberalism must either locate reasons to support said assertion within the relevant evaluative framework – that is, individuals' comprehensive doctrines – or, alternatively, appeal to a shared epistemic authority which is independent of the latter. As Habermas highlights, "a requirement of practical reason to which comprehensive doctrines must *submit* if an overlapping consensus is to be possible can only be justified by appeal to an epistemic authority that is itself independent of worldviews."[37] Both the appeal to individual comprehensive doctrines and the recognition of an independent source of epistemic authority would, however, force political liberalism to drop the assertion that its principles are doctrinally autonomous.

Instead, political liberalism denies any need to acknowledge any further source of normative authority at all. Rawls notes that "political values [...] are not simply presented as moral requirements externally imposed."[38] Rather, citizens are assumed to "understand those values as based on their practical reason in union with the political conceptions as free and equal and of society as a system of fair cooperation." In short, political values are supposed to gain normative authority from citizens' commitment to core liberal ideals.[39] However, the question *why* citizens should confer a higher-order status to these ideals in particular is precisely what is in want of an explanation.

---

[35]Jürgen Habermas. *The Inclusion of the Other: Studies in Political Theory*. Cambridge, MA: The MIT Press, 1998, 84, original emphasis. Habermas's quote of Rawls refers to Rawls, "Political liberalism: reply to Habermas", p. 144.

[36]Rawls, *Political Liberalism*, pp. 98–99.

[37]Habermas, *The Inclusion of the Other: Studies in Political Theory*, p. 93, original emphasis.

[38]Rawls, *Political Liberalism*, p. 98.

[39]For a similar argument along these lines, see Charles Larmore. "The Moral Basis of Political Liberalism". In: *The Journal of Philosophy* 96.12 (1999), p. 610.

As demonstrated above, political liberalism in its current form lacks precisely the resources to provide such an explanation and to independently single out some principles. To paraphrase the argument set out above, it cannot but refer to citizens' *decision* to commit themselves to said principles.[40] Hence, the reason why a higher-order norm of cooperation, as such a key principle, can claim some collective normative authority is that individual citizens have themselves ascribed to it such authority over each of them. If we assume that morally autonomous individuals act for reasons, this decision must be tied to some reason-giving source that individual citizens can access. Its legitimate authority as a regulating attitude in the constituency of public justification is therefore a result of its position as the final link in each of a number of different chains of normative authority, ultimately linking to a person's fundamental source of normative authority within her comprehensive moral doctrine.[41] These individual sources of normative authority may not matter to liberalism as a whole, but they do matter to each individual citizen. They provide citizens with their *reasons* to commit themselves to liberal principles.

For individual persons, political liberalism's core principles can thus merely be seen to be co-emergent with a variety of further moral commitments, all of which are rooted in said persons' individual comprehensive doctrines. A consequence of this specific relation between liberal ideals and private comprehensive doctrines that characterises political liberalism is that it cannot but attribute to individual citizens' nonpublic convictions the same status as to the shared commitment to cooperation in the political sphere. Any attempt to furnish some values with a higher-order status is thus hampered by political liberalism's reliance on individual private comprehensive doctrines as sources of normative authority.

Hence, even if, empirically, some citizens will accept a higher-order norm of mutual cooperation, political liberalism lacks the resources to assign to it the same higher-order status within the theory in general. It is unavailable as a theoretical resource to resolve conflicts of deep moral disagreement when public reasons have run out. As a result, procedural solutions designated to resolve such instances of incompleteness suffer from a lack of normative authority precisely under those circumstances in which they crucially require it.

---

[40] Rawls, *Political Liberalism*, p. 137.

[41] Rawls seems to acknowledge this link, stating "[t]he fact that those who affirm the political conception start from within their own comprehensive view, and hence begin from different premises and grounds, does not make their affirmation any less religious, philosophical, or moral, as the case may be" (John Rawls. "The Idea of an Overlapping Consensus". In: *John Rawls: Collected Papers*. Cambridge, MA: Harvard University Press, 1999, pp. 421–448, p. 432).

5.5 Deep moral disagreement and shared values

5.5.1 Asserting priority for shared values: idealisation

As I demonstrated in the previous section, political liberalism's reliance on individual comprehensive doctrines as sources of normative authority prevents the theory from claiming a higher-order status for particular elements of a reasonable person's set of convictions. I have arrived at this conclusion after scrutinising the moral effectiveness of procedural solutions to incompleteness in public reason in cases of deep moral disagreement. However, the effect of political liberalism's dependency on individual comprehensive doctrines as sources of normative authority reaches beyond that particular context. Said dependency does not only affect political liberalism's capacity claim priority specifically for a norm of mutual cooperation, but also for all other publicly shared *substantive* values. If comprehensive doctrines cannot be assumed to sustain a general norm of mutual cooperation because they could attribute overriding importance to a particular nonpublic value in a specific case, they may also fail to lend support to a particular publicly shared value in an analogous conflict where the respective private value is of similarly great significance. With no alternative, independent source of moral authority, whatever normative claims political liberalism can make about the general precedence of shared public values over nonpublic ones depends on whether the relevant comprehensive doctrines *do in fact prioritise* those values which, under political liberalism, are publicly shared.

This conclusion could have radically challenging implications for political liberalism: it may not only be incapable of defusing deep moral disagreement where shared reasons are silent, but also where shared reasons are indeed available to justify the choice of a particular policy. Where citizens encounter a genuine conflict between their shared and private values, the mere fact that a given society's public political sphere is made up of citizens whose reasonable comprehensive doctrines *generally* support a set of shared, public reasons becomes insignificant. It does not allow us to draw any conclusions as to which of these values they have reason to prioritise.

It could be objected that these consequences only materialise if we accept that a general commitment to shared values is sufficient for individuals and their comprehensive doctrines to be reasonable and hence to fall within the scope of political liberalism at all. We could defuse the challenge of deep conflicts between public and nonpublic values if we idealised the constituency of public justification such that only those comprehensive doctrines that prioritised shared values *at all times* would count as reasonable. Doctrines – and by implication citizens – that insist on prioritising nonpublic values, no matter how infrequently, would then be discredited as *unreasonable,* and would therefore fall outside the scope of concern of political liberalism.

A theoretical move of this kind is not as such dubious. Political liberalism, after all, does not claim to be compatible with *all* possible comprehensive doctrines, some of which contain

beliefs that are irreconcilable with the most basic of liberal commitments. Doctrines which fail to universally prioritise shared values could be thought to be vulnerable on that front: said failure might be interpreted as a sign that their allegiance to basic liberal commitments is not strong enough for a liberal theory to be concerned with them at all.

If we decide that these doctrines are, in fact, fundamentally unreasonable by that standard, this means in turn that for a comprehensive doctrine to be reasonable is to be structured such that those of its values that are public values are naturally more significant than its other nonpublic commitments. As a result, conflicts characterised by genuine deep moral disagreement simply do not occur among the comprehensive doctrines that remain within such an idealised constituency of public justification. Whenever a reasonable person is conflicted between a value she merely holds privately and one that belongs to the set of publicly shared values, she will consider the latter to be more weighty. This is not because it happens to be a public value, but because it is actually of a higher status within that person's reasonable comprehensive moral doctrine. A person who insists on prioritising her private commitments simply fails to recognise – or at least act on – the reasons she has. She can thus be said to be merely acting *incorrectly* by her own account. Under such idealised circumstances, her failure to prioritise shared values is merely a symptom of the fallibility of human reasoning.

At first glance, we might object to such a strategy of idealisation as a manner of cheating our way out of potentially fatal challenges in order to rescue the integrity of the theory, merely adapting the image of reality to fit our preferred conclusions. Idealisation may, however, avoid that charge if it can be justified not by such pragmatic considerations, but by reasons that reflect and are consistent with the aims of the theory.[42]

One such potentially admissible justification for idealising comprehensive doctrines such that precedence for public values is intrinsic to them could be the assumption that the pluralism allowed by a more narrow interpretation of reasonableness of this kind is actually consistent with political liberalism's ambition to identify political principles for free societies characterised by moral pluralism. In other words, idealisation could be justified, if it were true that, in such pluralistic societies, those comprehensive doctrines which support the idea of persons as free and equal and the burdens of judgement are also likely to grant precedence to publicly shared, key liberal values.

For political liberals, the empirical reality of citizens' comprehensive doctrines in these kinds of societies *should* have particular normative significance. As I argued in chapter 2, political liberalism is concerned with what it means to take seriously citizens' diverse bases for moral reasoning in modern pluralistic societies, given that such pluralism is likely to result from the kind of political institutions which reflect liberalism's foundational commitment to respect for individual persons' moral autonomy. The degree to which political liberalism can

---

[42]In David Enoch's words, "it has to be shown that the idealisation is consistent with [...] the considerations that led us down the path of tying the relevant normative phenomena to people and their responses." (David Enoch. "The Disorder of Public Reason". In: *Ethics* 124 (2013), p. 165).

accommodate comprehensive moral doctrines as they are in real pluralistic societies is thus a significant benchmark for the success of the theory.

If the structures of real and idealised comprehensive moral doctrines could be assumed to be approximately similar in a sufficiently large proportion of society, a strategy of idealisation might not be in contradiction with political liberalism's core commitment to respecting individual persons' moral autonomy. In that case, political liberalism could indeed draw on idealised comprehensive doctrines which are immune to deep moral conflicts between public reasons and nonpublic moral commitments. I am going to argue, however, that this assumption is unlikely to be accurate for two related reasons. If we assume that actual comprehensive moral doctrines that are considered reasonable in virtue of a general commitment to political cooperation are unlikely to encounter deep conflicts between public and nonpublic values, we both *(1)* underestimate their potential complexity, and *(2)* disregard their character as doctrines that are rooted in and composed of beliefs which do not as such need to be primarily liberal in nature.

Citizens' comprehensive doctrines may be *(1)* complex in that relations among the various beliefs and values that constitute a comprehensive doctrine are characterised by fine-grained priorities that may be sensitive to context and interactions among different values. Hence, what a person can genuinely be said to have reason to accept may be hard to assess if we consider her different commitments in isolation from each other and do not assess the entirety of her reasons and beliefs as they play out with regard to a particular issue. To assume that a citizen's general commitment to persons as free and equal and to mutual cooperation in the political sphere allows us to conclude that she is unlikely to digress from said commitment under any circumstances disregards the fact that her moral reasoning may be genuinely far more nuanced.

Part of the complexity that characterises reasonable comprehensive doctrines' interaction with core liberal commitments stems from *(2),* the diversity of their foundational commitments. Political liberalism allows for individuals' comprehensive doctrines to be rooted in foundational commitments other than core liberal values. In other words, the beliefs at the very core of a person's moral reasoning may be of a character entirely distinct from that of the liberal values they indirectly support. For example, assume that Catholic Alice ultimately locates her source of normative authority in God's will in the different forms in which she believes it to be revealed, such as scripture or personal prayer. She believes that divine love for each and every person grounds a norm to respect others as moral equals which requires her to engage with them in the spirit of mutual cooperation. Nevertheless, her derivative commitment to the latter norm may be qualified, if it is in conflict with another commitment rooted in what she takes to be divine revelation.[43] If one takes seriously the character of comprehensive doctrines as doctrines with (potentially) full and distinct moral and metaphysical backgrounds – the precise content of which may actually affect what individuals consider

---

[43]Hence Alice may, for instance, have reservations against laws that force healthcare providers to fund contraception, despite the fact that support for such laws may be derived from her shared public commitments.

to be morally authoritative for themselves – rather than as liberal theories in disguise, deep conflicts among public and private elements of said doctrines should not come as a surprise. To dismiss that potential, in other words, is to zoom in solely on the *reasonable*, *i.e.,* the substantively liberal elements of comprehensive doctrines at the expense of sincere attention to their particular distinct metaphysical and moral character that distinguishes them as *comprehensive* doctrines.

Idealisation of the kind described above is unlikely to match the constituency of views that political liberalism is committed to accommodating. Hence, we do not have reason to believe that idealisation merely corrects for individual citizens' erroneous reasoning when it claims that they are wrong in prioritising their nonpublic commitments. Their claim to the priority of said commitments may well reflect an accurate reading of their reasons, all things considered. If this is the case, idealisation presents political liberalism with the following dilemma between its core commitments to accommodating the pluralism that is generated by liberal institutions on the one hand, and to respecting individuals' moral autonomy on the other:

(a) Concern for liberal pluralism requires that the theory can actually account for the genuine moral commitments citizens would hold in those pluralistic societies that are shaped by liberal institutions. The constituency of those to whom the principles of political liberalism can be said to apply must strive to reflect these societies.

(b) Concern for individuals' moral autonomy requires that those within said constituency of public justification must be able to acknowledge that they have genuine reason to recognise the unconditional priority for shared liberal values as morally authoritative for themselves.

If we were simply to restrict the constituency of public justification to those whose comprehensive doctrines always prioritise shared values (reflecting *(b)*), said constituency would be small, given the complexity of actual comprehensive moral doctrines. That would weaken our claim that political liberalism is able to issue norms suitable for a large and morally diverse set of people, and would hence be in conflict with the commitment expressed in *(a):* a genuine concern for the kind of moral pluralism fostered by liberal institutions.

Enlarging said constituency to widen the scope of political liberalism to include a more diverse set of moral doctrines will, in turn, create tensions with *(b),* political liberalism's commitment to respecting individual moral autonomy. Such respect requires that we must, in good faith, be able to assume that the political judgements of the members of our idealised constituency match the conclusions that members of actual pluralistic societies draw from reasoning about political matters based on their respective comprehensive doctrines. But given the nature of pluralism in the societies I described above, this assumption is likely to be inaccurate. It is to be expected that some citizens face conflicts between their public

and nonpublic commitments which they are not necessarily able to resolve in favour of the former. In shaping the constituency of public justification based on the assumption that they are in fact able to do so, we risk dismissing what are genuine moral conflicts as mere errors of reasoning on their part. To knowingly disregard these moral conflicts is to express a lack of respect for their moral autonomy. If their individual morally autonomous judgements about what they see themselves as having reason to accept do not count as arbiters for the legitimacy of coercion unless they are in line with the preferred conclusions of political liberalism, the latter can hardly claim to truly respect their moral autonomy.

Idealisation, I have shown, is unsuitable as a means to resolve the tensions emerging from deep moral disagreement for a theory that is crucially committed to accommodating actual moral pluralism and respecting individuals' moral autonomy. Its operation is inconsistent with the basic motivations of political liberalism. It does not merely enable us to identify more clearly the justificatory links between reasons people can be said to have and principles they are expected to support by filtering out distortions of imperfect human reasoning. Instead, idealisation also eliminates theoretically significant variety in citizens' judgement or requires key liberal values to be abandoned. Rather than enhancing its understanding of how to deal with deep moral conflicts among reasonable people, idealisation undermines political liberalism's commitment to take seriously and accommodate anything close to the moral pluralism which is generated by liberal political institutions.

5.5.2　Can't we just bite the bullet? Accepting gaps of legitimacy

I have so far defended the claim that political liberalism lacks the internal resources to cast any judgement on what a reasonable person has reason to do when faced with conflicts among public and nonpublic values within her comprehensive doctrine. All it can declare is that reasonable citizens share *a* commitment to the core values that characterise reasonableness. Judgements on any particular conflicts of value are out of its reach. This means that there are gaps within political liberalism's capacity to ensure that potentially coercive policies are justifiable to all citizens, and hence in its capacity to confer legitimacy to political decisions on some controversial political questions. Closing these gaps by means of idealisation would come at the expense of contradicting some of the key motivations of political liberalism.

It seems like we have come full circle, and must conclude that public reason is in fact *incomplete,* given that the values political liberalism assumes to be shared do not provide guidance for each and every possible moral conflict that citizens encounter in the political sphere. My arguments so far have added to this conclusion in two way. First, they demonstrated that said incompleteness does not only occur when shared reasons relating to a particular issue are in fact absent – with public reasons being either indeterminate or inconclusive – but also when citizens hold genuine private moral commitments that conflict with them. Second,

they showed that legitimacy for decisions made under conditions of deep moral disagreement may sometimes be elusive, irrespective of whether we arrive at them by means of procedural strategies or by drawing on shared substantive values. In other words, liberal legitimacy – as it is understood within political liberalism as it currently stands – may well be bounded.

What is the upshot of that conclusion? It seems that we could just bite the bullet and accept these boundaries of liberal legitimacy as limitations to an otherwise still attractive theory. In other words, even though political liberalism is unable to resolve deep moral disagreements legitimately, its theoretical framework may, all things considered, still prove to be an adequate, or even the best, response to the question of how to ensure liberal legitimacy for modern pluralistic societies in general. Before we accept that conclusion, however, we should ask if are there any reasons to believe that these limitations are benign in the sense that accepting them does not threaten political liberalism's internal stability as a theory.

How might accepting political liberalism's gaps of legitimacy cast further ripples through the theory? As long as citizens can rely on the principles of political liberalism to yield legitimate policies most of the time, their overall trust in its capacity to sustain a well-ordered society may well remain intact, despite the fact that it falls short of its own standard of legitimacy on a handful of cases of deep but limited moral disagreement. However, if, in the long run, these cases become to numerous, they could threaten to erode the overall stability of the consensus on liberal principles, as I am going to argue.

In theory, however, any policy is potentially vulnerable to such a loss of legitimacy if there are any citizens that happen to raise genuine moral objections against it. If a large number of policies actually became subject to challenges of this kind, political liberalism's claim to provide a normative framework that ensures that individuals are governed by principles and policies that are acceptable to all would become increasingly less credible. In the long run, citizens' confidence in and allegiance to its principles might be diminished.

That said, in practice, only a limited number of issues are likely to involve genuine deep moral disagreement, leaving the effectiveness of shared principles intact in the majority of decisions. One worry remains, though, if we consider the reasons why deep moral disagreement proves to be so divisive in the first place: these are issues that citizens care deeply about, deeply enough for them to reject pragmatic compromise as an option that would allow them to live up to their commitment to cooperation. These are likely to be issues that crucially affect their perception of the overall moral decency of the political principles and institutions: questions about the contested moral status of unborn human and all non-human life, about the facts and value – or non-value – of suffering at all stages and in particular at the end of life (*e.g.,* medically assisted suicide), interference with the biological foundations of life (*e.g.,* genetic engineering), but also legal requirements to perform actions that one considers to be grave moral wrongs (*e.g.,* conscientious objection). These questions touch upon what some citizens regard as necessary moral tenets for a state to be free from fundamental moral corruption: one in which the basic dignity of life as *they* see it is preserved.

Political liberalism may deal with such disagreements by acknowledging gaps of legitimacy with regard to these deeply dividing issues. How many gaps on such issues citizens might tolerate before they affect their confidence in political liberalism as a theoretical framework which they can trust to produce a fundamentally morally decent society that is worthy of their allegiance, however, is an open question.

Now, one could simply object that people who are prone to such a loss of confidence are unreasonable. An excessive disposition to question the legitimacy of the system in response to its failure to deliver a particular outcome does surely give us reason to conclude that a person's commitment to principles of cooperation is too fragile for her to qualify as reasonable: such a degree of sensitivity may raise doubts as to whether her commitment to the moral equality of her fellow citizens is actually morally effective, if she is all too quickly prepared to call into question a system that is committed to giving effect to their moral status.

On the opposite side, it is hard to defend the claim that no amount of frustration can warrant legitimate doubts about the adequacy of the system as a whole. There is no point denying that societies led by liberal principles *are* capable of producing (and have produced) policies that are now recognised to warrant (and have warranted) doubts about the overall legitimacy, not of a political system comprised by liberal principles as such, but of their particular arrangement. To deny that citizens may ever justifiably challenge the adequacy of the system based on the substantive injustice it is perceived to generate is to claim that its principles will never fail to transfer sufficient moral value to the policies they produce, no matter what they are. That claim, again, should strike us as putting excessive strain on political liberalism's avowed respect for individuals' moral autonomy. If it is to avoid that fallacy, political liberalism cannot simply reject as unreasonable *all* doubts that a morally autonomous person could develop with respect to the overall moral decency of a political system shaped by liberal principles, no matter the character or degree of frustration that triggered them.

And while we can be certain that some such doubts may well be symptoms of unreasonableness, we lack a criterion to specify precisely the amount of frustration that individuals can be expected to tolerate before they may justifiably lose faith in the effectiveness of political liberalism without casting fundamental doubts on their commitment to others' moral equality. In trying to identify such a criterion, we re-encounter the familiar challenges that, in the previous section, I showed to afflict any attempt at specifying an appropriate degree of idealisation. Any such criterion cannot but draw its moral authority from the same roots as the very doubts it is destined to assess: the comprehensive moral doctrines that the respective individuals recognise as valid sources of normative authority for themselves. Again, that does not mean that no such criterion could be universally supported by all citizens. It does mean, however, that we may not simply allow that criterion to be shaped by whatever is required to vindicate political liberalism's claim to deliver a stable political framework for pluralistic societies. Hence, we must accept that under some degree of strain from too many gaps of legitimacy, political liberalism may well prove to be unstable.

I have so far argued that we have reason to at least acknowledge the possibility that gaps of legitimacy might threaten to de-stabilise a society governed by principles of political liberalism and public reason. The force of that threat does not least depend on how likely we are to encounter such potentially de-stabilising gaps in modern pluralistic societies. I do not want to overstate that likelihood. It is hard, if impossible, to predict if and when a single gap might be sufficient to trigger such a loss of trust in these principles overall. But if we consider the kind of political polarisation that characterises many modern democratic societies, there is some evidence of tendencies to question the overall legitimacy of the institutions of government if they fail to decide in one's favour on specific issues. From such a perspective, a political opponents' position is not only seen as threatening one's chances of realising one's own vision, but also the overall moral decency of a system which allowed that position to prevail.

A recent study of political polarisation in U.S. politics revealed not only high levels of partisan antipathy, but found that 27% and 36% of Democratic and Republican party supporters respectively agreed that the other party's policies "are so misguided that they threaten the nation's well-being."[44] Some kinds of political rhetoric also translate such a heightened sense alarm about one's political opponents' policies into mistrust of political institutions in general, as, for example, it was the case in the wake of the United States Supreme Court's decision in favour of same-sex marriage.[45]

While none of these observations directly concerns political or public reason liberalism (which is not as such implemented in actual political societies), they illustrate that pluralism in modern democratic societies does in fact harbour positions that, to their holders, are of a moral gravity sufficient to raise such fundamental doubts. For political liberalism, this characteristic of modern pluralism is an important variable to consider. It gives us some reason to believe that its gaps of legitimacy might actually affect its capacity to ensure stability in societies ordered by its principles.

How should political liberals react to that threat of instability? A precise answer is beyond the scope of this work. My focus here has been on identifying and examining political liberalism's weaknesses when confronted with deep moral disagreement. However, in the course

---

[44]Pew Research Center. *Political Polarization in the American Public.* `http://www.people-press.org/files/2014/06/6-12-2014-Political-Polari-zation-Release.pdf`. June 2014, p. 35.

[45]See, for instance, former Arkansas Governor Mike Huckabee's vow that he will "not acquiesce to an imperial court any more than our Founders acquiesced to an imperial British monarch. [...] We must resist and reject judicial tyranny, not retreat." (*Presidential Candidates React: From 'Historic Victory' to 'I Will Not Acquiesce'.* `http://www.bloomberg.com/politics/articles/2015-06-26/presidential-candidates-react-from-historic-victory-to-i-will-not-acquiesce-`. Accessed November 2016) U.S. Senator Ted Cruz denounced the legitimacy of the ruling as a symptom of the wider political system's failure to respond to the will of the people. "[T]he biggest divide we have politically is not between Republicans and Democrats. It's between career politicians in both parties and the American people." (*Ted Cruz: States should ignore gay-marriage ruling.* `http://www.politico.com/story/2015/06/ted-cruz-gay-marriage-ruling-reaction-npr-interview-119559`. Accessed November 2016).

of assessing and rejecting a set of theoretical moves that the theory could take to ward off the challenges of deep moral disagreement – idealisation or acceptance of gaps of legitimacy – I also highlighted the obstacles these strategies face. And this may help to specify the conditions that a successful attempt to preserve political liberalism's integrity must meet: in order to ensure stability under conditions of moral pluralism that motivate political liberalism, it needs to find a way to address these deep moral conflicts, either by justifying a higher-order norm of cooperation or by providing alternative reasons for private restraint. However, political liberalism is limited in terms of the sources it can draw on to identify such reasons. Detaching the moral authority of its principles from individuals' comprehensive moral doctrines would turn liberalism itself into such a comprehensive doctrine. If it continues to rely on individuals' comprehensive moral doctrines as sources of moral authority, we must locate the solution to deep moral disagreements in said doctrines. At the same time, political liberalism's aim to provide principles for modern pluralistic societies and its commitment to citizens' moral autonomy set limits to the degree to which it can idealise the content of their comprehensive doctrines. This suggests that, rather than treating comprehensive doctrines as black boxes that political liberalism does not substantively engage with, it may be worth examining them and look for similarities in their character and structure that might yield internal reasons for citizens to exercise restraint. This is valuable information if we consider the project of a *political* liberalism to be worth preserving.

On the other hand, the results of my assessment can also be taken as evidence that other variants of liberal theories are preferable. I neither want to affirm nor reject that possibility. It was not my aim to defend a consensus conception of public reason liberalism as the superior variant of political liberalism, or liberalism in general. I merely intended to highlight theoretical tensions which liberals need to address if they care about the values embodied specifically by a consensus conception of political liberalism.

5.6   Conclusion

In this chapter, I argued that citizens in modern liberal societies are likely to encounter moral conflicts between the consequences of their commitment to cooperation in their political conduct and other commitments which are part of their comprehensive moral doctrines. Instances of deep moral disagreement, in which their moral reasons point them towards prioritising the latter over the former, render them less than fully reasonable by the standards of political liberalism, and hence place them and the reasons they deem to be valid outside of the constituency of public justification. Yet, as I have argued in previous chapters, such exclusion must be justified to those who are being excluded, if political liberalism is to live up to its commitment to treat individuals as ends in themselves.

In political liberalism, justification for political principles must draw on the values and principles individual citizens can be said to accept. It does not seek to assert the validity of

its principles by tracing their roots to a particular coherent and comprehensive set of philosophical first principles. Instead, political liberalism leaves the grounding for its principles as a blank to be filled by individual citizens and their particular conceptions of the world, including their comprehensive moral doctrines. It relies on the fact of support by individual citizens' reasonable comprehensive doctrines for its normative validation.

From the perspective of the theory, it is this social fact which, given political liberalism's metaphysical restraint, provides shared values with their normative authority. Political liberalism can claim that these values should govern individuals' conduct in the public political sphere because individuals *themselves* have reasons that render said political values normatively authoritative to them. For the individual citizen, however, liberal values and principles are not vindicated by the *fact* of support by their and others' comprehensive doctrines, but by the *support* of the latter. From the perspective of the individual, said doctrines, as the foundations of their moral consciousness, are the substantive sources of values they have come to accept as morally authoritative for them.

This is crucial to the central claim I defended in this chapter: instances of deep moral disagreement reveal gaps in the capacity of political liberalism to confer legitimacy to public policies by drawing on a consensus of shared reasons. Said gap can only be bridged if political liberalism can assert priority for shared public values over nonpublic commitments. Though essential to the normative framework of political liberalism, said values cannot claim independence from their roots in the individual foundational beliefs and values which are considered citizens' ultimate sources of normative authority. This lack of independent normative authority crucially affects political liberalism's internal coherence and capacity to confer legitimacy to a political system precisely when it is confronted with deep moral disagreement. This defect cannot be remedied by means of idealisation. Idealising those individuals populating the imagined constituency of public justification in such a way that they will always subordinate their nonpublic reasons to their liberal commitment to cooperation is in tension with a truly liberal conception of the person, and with political liberalism's stated aim to devise principles which can accommodate the pluralism of liberal societies. In short, political liberalism can neither provide justifications for excluding those citizens who fail to be fully reasonable, nor can it coherently place them outside the scope of its concern.

The next chapter will keep them right within that scope, proposing an argument which enables political liberalism to justify restricting the constituency of public justification to those who refrain from drawing on nonpublic reasons in public justification, *i.e.,* to those who are fully reasonable. As with my previous argument justifying the freedom and equality of all people to those who are fundamentally unreasonable in chapter 4, this argument will explore the reasons individuals can be assumed to have, given their actions, and the reasons they profess to have. This method, as I have argued in chapter 3, is itself compatible with the core liberal commitment to respecting individuals as ends in themselves, and is therefore suited to resolving the tensions identified in this chapter without creating new ones.

Chapter 6

Non-rationality of foundational beliefs

6.1 Introduction

In this chapter, I will explore the reasons citizens have to refrain from offering nonpublic reasons in the process of public reason. Those who draw on nonpublic reasons fail to be fully reasonable by the standards of political liberalism and therefore face exclusion from the constituency of justification. Yet, their exclusion must be justified to them, given political liberalism's commitment not to coerce individuals for reasons that they do not accept. In the previous chapter, I argued that political liberalism has so far failed to provide such justifications. Here, I will show that reasons to be fully reasonable can indeed be offered to those citizens who currently fail to live up to that standard if we look closely at the internal structure of their nonpublic reasons.

In order to do so, we must be prepared to consider nonpublic reasons in a fresh light. In the previous chapter, I briefly alluded to the question at which point a person who does not refrain from proposing nonpublic reasons qualifies as *fundamentally* unreasonable. In other words, when is a person's claim to be fundamentally committed to respecting their fellow citizens as free and equal persons clearly and justifiably undermined by her failure to propose reasons that others can accept?

The idea underlying these questions is that a person, under these circumstances, clearly has reason to recognise that her actions defy her belief in her commitment to the freedom and equality of all people. Such a person, one could argue, will either henceforth tread the path of full reasonableness or must consider her exclusion from the constituency of public justification as justified on the grounds of fundamental unreasonableness. But it is not easy to determine where precisely on the slope towards full unreasonableness an individual must be in order for that justification to succeed, and I do not pursue that question any further because it does not matter for the issue at hand. Unless we believed that every failure to be fully reasonable is indicative of fundamental unreasonableness, we would still have to contend with the cases of those individuals who merely fall short of full reasonableness, and

whose exclusion from the constituency of justification would still warrant an independent justification.

Nevertheless, it is instructive to examine the assumption underlying the image of the slope towards fundamental unreasonableness: namely, that individuals would consider there to be a conflict between their assertion of nonpublic reasons in the sphere of public reason and their commitment to respecting others as free and equal. That assumption relies on an inaccurate view of how individuals relate to their nonpublic reasons. As I will argue in this chapter, it is quite conceivable that for many individuals there is no obvious conflict between offering reasons which are, in the eyes of external observers, nonpublic reasons, and their commitment to others as free and equal. I will argue in section 6.3 that this is because they ultimately believe the foundations of these reasons to be universally acceptable. More precisely, such a belief is crucial to the character of a vast majority of foundational beliefs present in modern pluralistic societies and the comprehensive moral doctrines of which these beliefs are a part.

Approaching individuals' reasons and beliefs in the way in which they themselves see them is beneficial to the project of political liberalism. If supported by comprehensive moral doctrines as they appear not merely to the external observer, but to those who subscribe to their respective sets of beliefs, political liberalism draws on a more authentic representation of the moral pluralism present within the kind of liberal societies it strives to accommodate. In addition, paying attention to the way in which individuals view their comprehensive moral doctrines and the character of the beliefs they contain ensures that political liberalism takes seriously individuals' moral autonomy and pays respect to those beliefs which constitute their sources of normative authority. Adopting such an internal perspective helps to realise these key ideals, which arise from the fundamental liberal commitment to persons as ends in themselves.

Importantly, for the purpose of this chapter, this approach opens a door to justificatory resources which are internal to citizens themselves. If, according to *their* perspective, their comprehensive moral doctrines provide reasons not to draw on nonpublic reasons in the process of public reason, political liberalism can justify excluding from the constituency of justification those who fail to exercise such restraint. It can do so while steering clear of the defects which have plagued other attempts to address the tension within political liberalism arising from the presence of those who are not fully reasonable,[1] strengthening key liberal principles instead.

In section 6.2, I will first look at how political liberalism can approach nonpublic reasons from an internal perspective, and consider how drawing on nonpublic reasons in public reason can appear to be compatible with the requirements of reasonableness from that perspective. I will ultimately argue that there are good reasons not to equate the proposal of

---

[1]*I.e.,* the failure to accommodate the kind of moral pluralism generated by liberal societies, and an impoverished, illiberal conception of the person. See sections 2.3 and 5.5.

nonpublic reasons with an attitude of disrespect for the freedom and equality of others. It seems that we have reason to doubt whether those who draw on nonpublic reasons in the sphere of public reason are rightly seen as unreasonable at all. But digging further into those individuals' internal perspective on their nonpublic reasons proves these doubts to be unfounded. In fact, understanding their reasoning guides us towards a better understanding of how and why a person's perspective on what it means to respect others in public reason diverges from the standard – that is, external – liberal view. This kind of approach provides us with a genuinely *shared* understanding of individuals' comprehensive moral doctrines, and thus with a route to justifying the prohibition of nonpublic reasons in public reason. Section 6.3 lays the groundwork for such an understanding by analysing the character of nonpublic reasons and the foundational beliefs they are rooted in. Finally, in section 6.4, I argue that the character of these beliefs ultimately gives individuals internal reasons to accept the exclusion of nonpublic reasons from public reason and, consequently, their own exclusion from the constituency of public justification should they fail to respect it.

6.2 An internal conception of liberal respect

Liberal respect in public reason requires that individuals must only be coerced for reasons acceptable to all members of the constituency of public justification. Nonpublic reasons do not seem to live up to that standard whenever they are approached from the perspective of an external observer. Such an observer relies on an *external* assessment of both the reasons acceptable to another person – which is likely based on her *articulation* of the reasons she considers herself to have – and the nonpublic reasons that are proposed to her by her interlocutor. From an external perspective, all judgement on what can or will be acceptable to a person is ultimately derived from whatever reasons she presents as currently acceptable to her. This means that any reason that others offer to her is likely to be dismissed as a nonpublic reason unless there is some overlap between it and the reasons she *presents* as acceptable to her, insofar as they can be identified by an external observer. In other words, the question of justifiability is settled by a somewhat static comparison of individuals' outward expression of their reasons and beliefs.

Is that the only plausible interpretation of the ideal of liberal respect? Can individuals only be said to be respectful of the moral autonomy of their co-citizens if they adopt the external perspective and determine what is acceptable to others based on the information which the latter provide to them about their reasons? This is the key question that I will explore in the following.

The notion of respect that I just described reflects, in particular, a mindfulness of the moral personalities of individuals and their choices *as they are.* The reasons that individuals *present* to others are recognised as the sole authoritative enabler of coercive power. Respect in this sense translates into trust that individuals themselves must *actually* be the ultimate judges of

what they may and what they must not (be coerced to) do. Hence, we should take a person's willingness to offer to others only reasons which match the reasons her interlocutors present as acceptable to them as an important marker of respect for their actual exercise of their moral autonomy. This dimension of respect should not be discounted lightly. If we are committed to paying respect to individuals' expression of their moral autonomy, yet do not pay attention to whether we actually respond to and act upon their real choices, said commitment could be seen as failing to grasp the very essence of the idea of autonomy.

However, this conception of respect fails to account for the significance that is often intuitively attributed to a person's intentions and attitude towards others which accompany her actions when judging whether or not she has acted respectfully. In assessing whether a person's conduct towards another is respectful, we may often be inclined to credit her for her intentions, even if the intended action is ultimately unsuccessful. In this case, we may well consider it to be to her credit if she *sincerely aims* to do whatever respect requires under the circumstances in question. (In the context of public reason, it requires her to present reasons acceptable to her interlocutors.) For example, if a person acts in a way that is considered respectful in her own culture, but disrespectful in another, her intention to act respectfully is not voided by her failure to make her interlocutors feel respected within the latter cultural context. Exculpatory reasons help account for the mismatch between her intentions and the reactions her actions elicited. If someone is unaware of what respect requires in a particular context, her actions are no longer blamed on a lack of good intentions, but on a lack of knowledge. But even in the absence of such exculpatory reasons – "The manner in which she acted was just plain stupid!" – our reaction to someone whose intention it was to express respect is different from our reaction to someone who acted in the same way without regard – or blatant disregard – for the requirements of respect. A person's attitude is *a factor* in assessing the moral character of a person's actions.

To further substantiate that point, consider the opposite example of a person merely feigning respect. Imagine a person who outwardly acts in line with what respect requires in a given situation, but does so not motivated by an intention to show respect, but by other reasons, such as a fear of repercussions or a desire merely to be seen to be showing respect. If we discovered this discrepancy between her attitude and her actions, we would struggle to continue to think of her actions as respectful, precisely because we would judge her attitude to be wanting. Attitude is a significant dimension of the expression of respect.

Within liberalism, this attitudinal dimension of respect is particularly important. Liberal values are crucially concerned with the attitude that citizens are supposed to adopt towards each other. They are supposed to regard each other as free and equal persons – as equal sources of valid claims. It is not surprising that there are sometimes discrepancies between individuals' actions and their fellow citizens' expectations as to what actions befit an attitude of liberal respect. But people who intend to relate to each other as free and equal citizens can debate and learn from each other what respect requires in terms of actions. That is what

public reason is about. From a liberal perspective, the successful expression of respect cannot be more important than the attitude which renders such acts of respect morally significant in the first place. There can be no liberal respect among citizens who do not see and intend to treat each other as moral equals.

Nevertheless, there is also good reason not to focus on professed intentions alone and disregard entirely the way in which a person's professedly respectful actions are perceived. Such disregard would ultimately provide cover for individuals who cast a blind eye to easily identifiable reasons, either out of insincerity or negligence. Negligence is disrespectful – assuming appropriate standards for the effort to obtain information about others' actual reasons – because it expresses a disposition to shirk reasonable amounts of effort, giving us reason to doubt the sincerity of individuals' pursuit to identify acceptable reasons. Insincerity in their professed intentions to present reasons acceptable to others is clearly disrespectful because it is incompatible with a proper *attitude* of concern for the moral autonomy of others. For a conception of respect that puts weight on its attitudinal dimension, it is insufficient for individuals to merely pay lip service to a commitment to others' moral autonomy: it must be backed by a sincere concern for offering to others reasons that are acceptable to them, and by the willingness to expend a reasonable amount of effort on identifying them.

It seems to follow that, under circumstances where information on citizens' beliefs and reasons with respect to a particular matter is readily available, those looking to justify a policy on the subject cannot respectfully disregard that information. Not only can there be no doubt that their justificatory efforts will be successful when they use reasons they *know* to be acceptable to their interlocutors. Failure to do so, it seems, can only spring from negligence or insincerity. After all, one would not forgo ensured justificatory success, and the successful expression of respect that goes with it, unless one either neglected to consider what respect required, or intended to disregard it. While we can dismiss those who merely feign respect as clearly beyond the scope of reasonableness since, by definition, they lack the required attitude, those who do not draw on reasons known to be acceptable to others may simply fail to be sincere enough in their commitment.

Still, it would be too rash to subsume all appeals to reasons that a person does not currently accept as an indication of a lack of a sufficient concern for others as morally autonomous, free and equal persons, and thus as a deficient attitude of respect. I will go on to paint a more complex picture of the relation between reasons and respect. At this point, it is worth considering what the information we have about others' beliefs and reasons actually reveals to us: it reveals the reasons they *currently* consider themselves to be able to accept. However, given the fact that human rationality is bounded, the reasons we currently accept are never reflective of our full system of reasons and beliefs. We are not fully transparent to ourselves. Hence, our current reasons can only ever paint a partial picture of our set of reasons and beliefs. In other words, what we deem to be acceptable to us right now may tell us only little about reasons that could be acceptable to us as well, given the entirety of beliefs we hold.

Relatedly, what we deem acceptable now also tells us little about the strength of said beliefs, the exact nature and structure of their foundations, as well as their respective strength.

Hence, the set of reasons and beliefs that a person currently accepts does not necessarily provide much information – and certainly no determinate information – about the way in which her position with regard to a particular subject *might* be transformed. Yet, such transformation is a common occurrence. When we deliberate on complex normative matters, we are prepared to – and frequently do – change our minds after re-examining the reasons we already have and considering potential reasons we have come across. In fact, this process is precisely what we would expect the responsible use of our moral autonomy to amount to, as I will explain in the following.

Respect for autonomy requires that we recognise individuals as the ultimate judges of any potentially coercive measure. But to take this recognition to mean that we cannot expect individuals to reflect on their judgements would be to discount the normative dimension of moral autonomy. To be autonomous does not merely protect individuals' prerogative to cast the ultimate judgement on what is acceptable to them, but also to examine which judgement is *right* for them to cast. In chapter 3, I argued that this view of moral autonomy is compatible with – and is in fact implicit in – the ideal of justification. This is not to say that respect for these judgements should depend on their quality in that latter respect, but merely that the normative dimension of moral autonomy discourages resistance to reflection. Careful consideration of normative positions and the reasons that support them is what is required if an individual is serious in her endeavour to decide what is right for her to do, what reasons for action she should accept, and what reasons she is right to reject.

Hence, approaching others with the expectation that they might accept a reason presented to them only after reflection is compatible with respect for their moral autonomy. In other words, there is nothing disrespectful in acting on the assumption that a person's current set of reasons and beliefs *could* be transformed by confrontation with the nonpublic reasons that their interlocutor believes to be applicable to them.[2] Addressing others with reasons that they do not currently accept, but which might induce a change of mind, should not be deemed disrespectful of their moral autonomy just because it looks past their current allegiances. As an appeal to the very trait that enables them to *change* their position on the matter at hand, we can also take it to reflect an emphatic embrace of their moral autonomy, rather than a dismissive attitude towards it. In summary, to present others with reasons in the hope of transforming the reasons they currently accept does not necessarily undercut the sincerity of an individual's commitment to only coerce others based on reasons that are acceptable to them. While such behaviour may also mask negligence or insincerity, it does not *necessarily* reflect either of these.

---

[2]For now, my intention is merely to highlight that acting on the perception of such a potential is not as such disrespectful. In section 6.3, I will explore the conditions that affect the transformation of beliefs and provide a more qualified account of when it is appropriate to assume that another person's reasons can be transformed.

6.3 What kind of reasons?

In the previous section, I argued that a conception of liberal respect for moral autonomy which takes seriously the perspective and attitude of the individual proposing nonpublic reasons must not dismiss these reasons merely for their lack of ties to any reasons and beliefs that their addressees currently accept. An attitude of liberal respect is compatible with the proposal of reasons which are claimed to be acceptable to others merely on the basis of a person's *internal* assessment.

At first, this conclusion might give us pause: are we wrong in brandishing those who propose nonpublic reasons as unreasonable? How much weight should we give to their internal perspective on the reasons they propose in public reason – according to which these reasons are acceptable to others – compared to the external perception – according to which they are not? Fortunately, there is no need to tackle this question head-on. My main reason for embarking on an analysis of individuals' internal perspective on nonpublic reasons was to identify reasons they have to refrain from proposing these nonpublic reasons in the process of public reason. Taking seriously that perspective, I argued, would provide us with a robust justification for their exclusion from the constituency of justification. As will become apparent in the further course of this chapter, the basic assumption integral to the internal perspective which has so far held up – namely, that nonpublic reasons can be assumed to be acceptable to those who currently do not accept them – ultimately proves to be untenable. Consequently, the apparent conflict between the internal and external view collapses. There is, after all, a conflict between an attitude of respect and the proposal of nonpublic reasons, and it is one that individuals can grasp entirely from within their internal perspective on their reasons. This perspective thus provides them with reasons for justificatory restraint and, in turn, furnishes political liberalism with justifications for excluding them from the constituency of justification should they fail to exercise that restraint. But in order to arrive at that conclusion, we must first dig further into individuals' internal perspective on their nonpublic reasons. Specifically, we must ask what precisely gives them reason to believe that the latter are acceptable to others.

6.3.1 Truth and acceptability

A person's individual internal experience of the nonpublic reasons she is prepared to propose may differ significantly from the way in which they are perceived from an external perspective. What appears to her to be acceptable to others may actually seem quite alien to them. But that is not decisive from the perspective we have so far adopted in this chapter – namely, a perspective which is concerned with the question of whether an individual can *sincerely claim to believe* that the reasons she proposes are acceptable to others. All that is required for her to be able to make such a claim is that she sincerely believes that she has valid *second-order reasons* to believe that her nonpublic reasons are acceptable to others. Some such

valid reasons may draw on information that a person has about her interlocutor's overall set of reasons and beliefs. To be able draw a consistent and coherent argument in favour of a currently nonpublic reason from higher-level or even foundational beliefs contained in said former set of reasons and beliefs certainly constitutes a good reason to believe that it may be acceptable to the owner of that set.

Yet, from an internal perspective, a valid reason to believe that a given reason may be acceptable to others does not necessarily require any such connection to a person's current beliefs. Another valid reason to believe that others may come to share a particular reason is a sincere belief that it is *true,* and therefore *universally* acceptable. This requires further explanation. I do not mean to claim that truth actually enables acceptability. All I mean to say is that the insight one believes to have gained into the world *as it is* must also be *believed* to be acceptable to others insofar as they also ultimately strive for truth with respect to their beliefs. For others to be able to recognise said truth, a person may also believe it to be necessary for them to accept a potentially quite expansive network of other beliefs. They may have to buy into a new view of the world, so to speak! But the potentially high cost that may be associated with enabling others to recognise the truth – and hence acceptability – of a given belief does not need to affect a person's perception of its universal acceptability. From a person's *internal* perspective, being convinced of the truth of a particular belief may still constitute a good reason for her to believe that it is, in principle, acceptable to others. In fact, many of the most controversial beliefs that people have are backed by a sincere belief in their truth: this includes religious beliefs and non-relativist moral beliefs that contain a truth-claim.[3] According to the preceding argument, such beliefs may be perceived, by those convinced of their truth, to be acceptable to all.

This warrants a more detailed explanation. Consider the following example of *cars:*

1. I have the perception of a red car.

2. I trust that my relevant perceptive faculties (my eyesight) are truth-sensitive.

3. I have – to the best of my capacities – assured myself that my faculties operate without distortion.

4. Therefore, my perception of the red car warrants my belief that it is true that the car is red.

5. I believe that others also aim to believe what is true.

6. I trust that others' relevant faculties are also truth-sensitive.

---

[3]As an aside, this is precisely how Rawls characterises the beliefs which individuals may be tempted to introduce as nonpublic reasons: "those who insist, when fundamental political questions are at stake, on what they take as true but others do not, seem to others simply to insist on their own beliefs when they have the political power to do so. Of course, those who do insist on their beliefs also insist that their beliefs alone are true: they impose their beliefs because, they say, *their beliefs are true and not because they are their beliefs.*" (Rawls, *Political Liberalism*, p. 61, my emphasis).

7. Therefore, others' faculties – if operating without distortion – should generate the same belief as mine.

8. I have reason to believe that the belief that the car is red is acceptable to others if I believe that it is true.

Step 2 and the inference from step 6 to 7 are the crucial elements in this line of argument. I will discuss them in turn.

Step 2 assumes that I may trust that my perceptive faculties generate true conclusions. This assumption invites a sceptical challenge. May I trust my faculties if I have no means to ensure that they do not deceive me and actually accurately generate the kind of insight I believe them to generate? To comprehensively address the sceptical challenge is beyond the scope of this chapter. However, I do not believe it is necessary to assuage the sceptic. The sceptic might challenge my trust that my respective faculties generate true beliefs about the real world because they are my only source of such insight, and hence cannot be tested against any information obtained independently of them. But this is not to say that my trust may not be epistemically warranted. To assume that only external validation of my perceptive faculties' capacity to yield true beliefs may vindicate such trust would be to apply an inappropriate standard of epistemic warrant, which is profoundly at odds with the way in which we relate to world. We cannot escape the reality that, in our day-to-day conduct, we do trust these faculties and consider them to be indicative of what is indeed true. The belief that our perceptive faculties are truth-sensitive – though fallible – is normative for us. To seek further justification for the validity of our normative beliefs may simply not be feasible, as Alan Millar argues:

> The key question to consider is what aiming to have only true beliefs requires of us. Certainly it requires us to take such steps as are feasible to ensure that we believe only what is true. But what steps are feasible? Any steps we take proceed from a starting point which we have not chosen and could not reject wholesale even if we tried. The starting-point is our perspective on the world, which comprises the concepts we have acquired and the propositions which are normative for us [...] Having this perspective commits us to managing our beliefs and evaluating beliefs generally in certain ways. Indeed, if we manage and evaluate beliefs competently it would seem that we do as much as could feasibly be done to serve the aim of believing only what is true. [...] Doing what is feasible to serve the aim of believing only what is true does not preclude forming beliefs on grounds of a sort which do not reliably yield true beliefs.[4]

To seek further justification for our belief that our faculties are truth-sensitive is not a feasible step because said belief qualifies as precisely the kind of starting point which, according to

---

[4]Alan Millar. *Reasons and Experience*. Oxford: Clarendon Press, 1991, p. 213.

Millar, we "could not reject wholesale even if we tried." It is implicit in most, if not all, of our conduct that we do not choose to subject it to epistemological scrutiny. In a similar vein, Crispin Wright argues that our day-to-day functioning sets limits to what is feasible in ensuring that we believe only what is true. Awareness of these limits entitles us to trust that our faculties are indeed truth-sensitive:

> [A person] is so entitled because the need to take decisions will, time and again, trump whatever may be the limited possibilities – especially in the light of scep- tical argument – for gathering positive evidence that the general presuppositions hold good in the particular context, and because – as a rational agent – her deci- sions have to be informed by *reasoned* beliefs about what is for the best. Since such beliefs will be possible for her only in a context in which she has trust in what she knows to be necessary conditions for their being soundly arrived at are met, only a thinker who has such trust can be a rational agent.[5]

In other words, as an agent, I must not seek further evidence for the truth-sensitivity of my faculties, since to do so would undermine my capacity to arrive at reasoned decisions. An approach like this, which draws on our conception of ourselves as functioning agents, is congruent with the liberal perspective and therefore highly attractive.

I now turn to a discussion of steps 6 and 7, explaining why my assumption that others' per- ceptive faculties – if operating without distortion – are also truth-sensitive requires me to infer that they will generate the same conclusions as mine. I cannot reject this conclusion, because if I believed that my faculties were truth-sensitive while also believing that they might not generate the same conclusions as others' faculties of the same kind – absent any distortions – I would need to entertain serious doubts as to whether or not we share the same reality. Talking about truth – about what is the case – is only intelligible if we assume that what is real for me is also real for others. The relativism implicit in doubting that we share the same reality is fundamentally at odds with the project of public reason. Not only could we no longer talk about truth, but the whole endeavour of seeking political principles to govern interaction among individuals would be futile. This is because it is premised on the idea that we can come to recognise some principles as mutually acceptable – a convergence which is not purely incidental but principled: it arises from a shared appreciation of what morality demands under certain circumstances and why. Within public reason (and liberal political theory more generally) disagreements are not explained in terms of moral relativism, but in terms of distortions and the general boundedness of human reason. Moral relativism is incompatible with the project and those committed to such relativism cannot be genuine members of the constituency of public justification, since they are incapable of committing fully to the idea that public reason discovers and expresses what *are* the appropriate princi- ples governing our political interaction. Hence, as a member of the relevant constituency, I

---

[5]Crispin Wright and Martin Davies. "On Epistemic Entitlement". In: *Proceedings of the Aristotelian Society, Supplementary Volumes* 78 (2004), p. 198.

cannot reject the belief that, if operating undistortedly, distinct truth-sensitive faculties will generate the same conclusions. Hence, if I believe others' faculties to be truth-sensitive, I have reason to believe that what I believe to be true is also acceptable to others.

The same structure of argument may also be applied to moral beliefs. Consider *animals:*

1. My moral judgement leads me to conclude that animals should be granted the same moral status as humans.

2. I trust that my relevant faculties (my moral sense and moral reasoning) to be truth-sensitive.

3. I have – to the best of my capacities – assured myself that my faculties operate without distortion: I have reflected on whether my conclusion has been influenced by factors which I do not judge to be morally relevant and, if appropriate, whether my reasoning contains any (logical) errors.

4. Therefore, my moral deliberation warrants my belief that (it is true that) animals should be granted the same moral status as humans.

5. I believe that others also aim to believe what is true.

6. I trust that others' relevant faculties are also truth-sensitive.

7. Therefore, others' faculties – if operating without distortion – should generate the same belief as mine.

8. I have reason to believe that the belief that animals should be granted the same equal moral status as humans is acceptable to others, because I believe that it is true.

While this line of argument may appear persuasive with regard to beliefs about the material world like in *cars*, it may seem to loose some of its appeal when applied to normative beliefs as in *animals*. This is because we are generally inclined to trust that we share the same material reality. Intuitively, we seem to be more reluctant to accept relativism about the material world than we are in relation to normative judgements. However, none of this matters for the present argument. As I argued above, we can easily ward off doubts about the existence of a shared, objective moral reality, not by means of any substantive philosophical argument, but with reference to our shared commitment to such a reality as expressed by our commitment to the project of political liberalism and public reason. Therefore, the structure of the argument supporting my inference from my belief in the truth of a particular belief to its acceptability to others may be transferred to moral beliefs. Hence, whenever I engage in normative discourse that I believe to be meaningful and to which I believe others to be committed in the same way, my belief in the truth of a normative proposition warrants a belief that said proposition should, in principle, be acceptable to others, assuming that they also aim to believe what is true.

6.3.2   Acceptability and transformative potential

This conclusion brings us to the core of the story. According to the manner in which I have depicted the internal conception of liberal respect so far, it tolerates the inclusion of a wide range of nonpublic beliefs, including highly controversial moral and religious convictions. The fact that individuals present their interlocutors with reasons that may be entirely alien or even contradictory to anything the latter currently believe is compatible with a sincere *attitude* of respect towards others' moral autonomy. In other words, what is wrong with nonpublic reasons is *not* that their proposal in the context of public reason betrays a person's lack of a commitment – or insincerity thereof – to coerce others only for reasons that are acceptable to them.

So what *is* wrong with nonpublic reasons? I have so far argued that an individual's perception of the reasons she believes to be acceptable to others should matter. But their perception of their own beliefs is not *all* that citizens should care about if they care about respect for others' moral autonomy. Respect for others' moral autonomy requires that the latter cannot be expected to simply take their interlocutors' judgement of acceptability at face value. They must be allowed and *enabled* to arrive at that judgement themselves. This echoes my argument in the previous section that counting on the mere transformative potential of some reasons is not only compatible with liberal respect, but is, moreover, also reflective of the following crucial dimension of individuals' moral autonomy: their capacity to *change* their minds upon evaluating relevant reasons available to them. Therefore, a person's perception of the path towards the transformation of her interlocutor's set of reasons and beliefs is also relevant to the question of liberal respect.[6]

In that context, the epistemic conditions mediating said transformation matter: a person who proposes reasons that she believes to be acceptable to others must distinguish between her perception of said reasons themselves and her perception of the conditions that affect her capacity to induce that perception in others.

More specifically, she must acknowledge the distinction between:

(1) the abstract belief that some reasons *are* universally acceptable, because they are true;

(2) the belief that others *can come to recognise* them as acceptable in general; and

(3) the belief that they can do so under certain conditions.

It is possible to conceive of reasons which fall under *(1)* but not under *(2):* reasons that are sincerely believed to be true and hence believed to *be* universally acceptable, yet unrecognisable as such by others. Implicit in the concept of an externally unrecognisable universal truth is the assertion of privileged access, which disqualifies reasons that draw on such a truth

---

[6]This concern is more pressing with regard to reasons which are deemed acceptable not because of their roots in others' current set of reasons and beliefs, but because of a more abstract belief in their universality.

claim as reasons whose proposal is respectful of others' moral autonomy. Their acceptability cannot be ascertained by means of assessing their substantive content. If others did accept such reasons after being presented with them, they would not – and could not – do so because they actually judged them to be valid, but only because they would have deferred to the epistemic authority of the proponent. But for the proponent of reasons to rely on their authority, rather than on a belief that the reasons in question would withstand a person's substantive evaluative judgement, is fundamentally incompatible with a commitment to their fellow citizens' moral autonomy. It counts on the latter's willingness to be directed by others, rather than exercising their capacity for self-directed judgement. This is not to say that a person cannot *decide* to defer to someone else's epistemic authority. She may have good reasons to do so (*e.g.,* trusting others' expertise on a particular subject matter), but the judgement that these second-order reasons are good, again, implies that they have been, or can be judged on their substantive merit (*e.g.,* with reference to evidence that such trust has proved to be warranted in the past). If the proponent of an externally inaccessible universal truth were to propose such second-order reasons to substantiate her unique competence in accessing said truth, others would be enabled to exercise their moral autonomy. But that is precisely what the proposal of externally inaccessible truths alone fails to do, for it relies solely on external authority for its transformation of others' reasons and beliefs.

Returning to my earlier threefold distinction regarding the way a person may perceive the reasons she deems to be universally acceptable, the same conclusion does not seem to be warranted with respect to beliefs of the second kind. Beliefs that are claimed to be universally acceptable and recognisable as such by each and every individual *in general* do acknowledge the need for a *substantive* transformation of the beliefs of others to take place in order to satisfy liberal respect for moral autonomy. However, a person's general belief in the ultimate possibility of such a transformation alone does not account for her capacity to *induce* such transformation in others. This is important because, without an argumentative path towards the belief she wants and believes others to be able to accept, the sincere claim that such transformation is possible would, again, demand acceptance on the basis of the proponent's epistemic authority alone, rather than enabling and relying on the addressees' substantive judgement. Hence, to satisfy the internal conception of liberal respect, a belief in the universal acceptability of a given reason must not discount the process of reasoning that is thought to lead towards the acceptance of the proposed reasons. To care about this process requires us to ask what it takes for individuals to follow the proposed path. Returning to the previous distinction I made between different ways in which a person may interpret her conviction of the universal acceptability of some reasons, we must conclude that we have reason to take seriously the *conditions* under which others can come to accept the proposed reasons. Importantly, we must ask whether these conditions can be satisfied in the context in which the proposal of reasons is taking place, which, in our case, is the process of public reason.

We may distinguish between internal and external forms of such conditions: as far as *internal*

conditions are concerned, we must ask what about a person's internal epistemic constitution has to change in order for her to accept the currently nonpublic belief that is presented to her. For example, does she have to accept one particular (relatively isolated) belief, or does her whole set of interrelated reasons and beliefs have to undergo major transformations? In terms of *external conditions,* we must ask what constraints affect our ability to persuade her of the validity of our claim that the nonpublic belief in question is really acceptable to her. For example, we may genuinely believe that others can come to accept a particular set of beliefs, but that they will only be persuaded of their validity after they have been exposed to a particular set of experiences (*e.g.,* experiences while living, for a while, according to a particular set of rules or principles, or having experienced the loss of people close to them).

The more comprehensive the required change in a person's epistemic constitution (internal conditions), the less likely it is for said change to occur by comparatively non-invasive means (external conditions), such as deliberation and debate in the sphere of public reason. But as long as it is genuinely thought to be possible to transform others' set of beliefs, attempts to transform others' epistemic constitution by means of proposing nonpublic reasons to them are not disrespectful. They are merely less likely to be successful under some circumstances. However, if we do indeed have reason to believe that the epistemic constitution required for a person to accept a particular belief could not *possibly* be induced by another person given the external constraints, we would fail to live up to our commitment to respect others' moral autonomy if we insisted that our nonpublic belief could constitute justifiable grounds for coercion to her.

Colin Bird offers an argument to the effect that we must not hope to bring about the internal conditions, *i.e.,* transformation of others' epistemic constitution, required for them to accept the reasons we propose. In his view, we have reason to believe that it is impossible in principle to convince others to adopt the required epistemic state, *precisely because* we are unable to convince them in actual debates on controversial issues. Therefore, he believes this approach to be inappropriate. According to Bird, "individuals ought to acknowledge and respect each other's authority to interpret their opaque experiences."[7] Experiences or beliefs are referred to as opaque if they are based "upon the interpretation of experiences which are not available for public political scrutiny", while those which are "based on the interpretation of experiences available to everyone for critical scrutiny" are referred to as transparent.[8] He argues that "we can know whether the grounds of dispute are transparent or opaque in a given instance. This is itself a question about which the grounds of dispute are normally transparent."[9] In other words, Bird claims that whether one of our beliefs is opaque or transparent is itself a question which may be discussed on the basis of shared experiences. According to Bird, in the case of deciding whether or not the grounds of a given dispute are opaque, the relevant shared experience is that of failures to reach agreements on the

---

[7]Colin Bird. "Mutual Respect and Neutral Justification". In: *Ethics* 107.1 (1996), p. 76.

[8]Ibid., p. 71.

[9]Ibid., p. 76.

matter at hand. From that perspective, what does count is our sincere conviction that others would come to share our view if they had only gone through similar experiences and if, as a consequence, their set of reasons and beliefs resembled our own.

> But although I am convinced that anyone who went through a similar experience would be compelled (as I was) to reject abstract dogmas about the sanctity of life, I am unable to convince others that this is the right place to start when reflecting on the morality of abortion. [...] The fact that I am unable to make any headway in such arguments, despite my continuing conviction, ought to convince me and observers that the grounds of dispute in this case are opaque. Failure to reach agreement on issues of this sort is surely an interpersonal demonstration of the fact that in such instances the grounds of dispute are opaque. This conclusion is not simply a matter of personal conviction: it is a conclusion based on a demonstration whose force can be equally appreciated by frustrated disputants and nonparticipating observers.[10]

However, I believe that Bird's inference from actual disagreement to opaqueness is flawed. The fact that we recognise that we are unable to convince others to expose themselves to the very experiences that we claim would make them appreciate the validity of our proposal does not commit us to accepting that our proposal rests on beliefs which Bird classifies as opaque: namely, beliefs which are based on the interpretation of experiences that are not available for public political scrutiny. We may still sincerely believe that *if* others were to expose themselves to the relevant experiences, they would come to share our conclusions. The fact that they are unlikely to undergo such exposure does not need to affect our sincere convictions about what *would* happen if they did. Hence, we may not have to accept that our experiences of persistent disagreement when debating with others are in fact transparent grounds for deciding whether or not the grounds which we introduce into the debate are opaque. We may well be in a position to consider the experience of *actual* disagreement to be irrelevant to the question at hand. Consequently, since we do not have a reason to believe the grounds underlying our and others' positions to be opaque, it would be unfair to accuse us of disrespecting others' "authority to interpret their opaque experiences",[11] merely for holding on to *our* sincere belief that our proposal can be acceptable to them. In doing so, we do not dispute their authority to interpret their experiences, but merely *suggest* an interpretation which we sincerely believe everyone can come to share. To return to my original point: even in the face of actual disagreement, we may hold on to sincere convictions that the reasons we propose, including their structure of supporting reasons and beliefs, may be acceptable to others. Actual disagreement does not commit us to believing that the internal conditions for others to share our reasons cannot be brought about.

---

[10]Ibid., p. 77.
[11]Ibid., p. 76.

As I discussed above, our capacity to bring about these internal conditions is bounded by external conditions, that is, by the means for confronting and addressing others that are available to us in the sphere in which our debate takes place. In the sphere of public reason, the external conditions are fixed: reasons must be communicable by means of *rational* argument. Public reason does not allow for reasons to be backed by invasive means of persuasion that require more of the addressees of reasons than to allow themselves to be confronted with and to consider verbal arguments that others offer. It does not envisage the addressees to modify their *behaviour* in order to access, and eventually come to share, the insights underlying the reasons that others claim to be acceptable to them. This is not surprising: after all, liberalism is concerned with protecting individuals' moral autonomy – *i.e.,* their capacity to act on their reasoned judgement on what is the right thing to do. It cannot ask them to *act* in order to recognise what is right according to another person. At the point of being asked to do so, they would still be required to act without a reason that they can currently accept, and hence to relax their claim to all normative authority over themselves. Within liberalism, judgement must precede action, and therefore all reasons must be communicated and substantiated by means of verbal, rational argument alone.

This is important if we consider the foundational convictions underlying individuals' other, more specific nonpublic beliefs. These foundational beliefs, as I am going to argue in the next section, are not communicable by means of rational argument. The reason for this is that even the person holding the foundational beliefs cannot herself account for them by rational means.

6.4   Foundational beliefs cannot be rationally accounted for

As human persons, we develop systems of moral reasons and beliefs which, over time, constitute the bases of our moral judgements. Between persons, these systems vary greatly in terms of shape, complexity, and coherence, let alone in terms of their substantive content. But they share some structural features. They are systems of support and dependency, and the currency *within* these systems is rational inference. For instance, we accept a particular moral belief as valid if it follows from a higher-order belief that we hold, or because it is a consistent and coherent addition to a set of related and mutually supportive beliefs. This is not to say that a workable system of moral reasons and beliefs must be – or could even be – without gaps, inconsistencies, and incoherences. That standard would be impossible to live up to, given that human rationality is bounded. We lack the intellectual capacities to assess all of our explicit beliefs – let alone identify all implicit ones – in relation to each other in each conceivable context. Nevertheless, rational inference as a shared standard gives us hope that moral beliefs may be rationally *communicable*. We can indeed lead others to recognise a given moral proposition as valid if we can only identify in their set of moral reasons and beliefs one or several beliefs that allow us to rationally infer from them to said proposition.

The question naturally arising from this claim is the following: are all moral beliefs communicable in this way? If we just persistently work our way through others' sets of moral reasons and beliefs, could we eventually succeed in re-shaping any such set in such a way that it supports the moral belief in question?

The boundedness of human reason is certainly likely to impose some limits on our capacity to do so. Others' moral beliefs and their interactions may quite often be simply too intricate for us to genuinely understand them in all their nuances. But while this is a reason to refrain from drawing any quick conclusions from the fact that a person holds certain beliefs, it is not a principled reason to refrain from cautious moral discourse in the *hope* that, in this specific case, we may be able to work through the intricacies of the relevant moral beliefs. This process may be lengthy and may prove to be irreplicable on a large scale, but this aspect of the boundedness of human rationality is no reason to reject the idea that a specific moral belief that we hold may be rationally communicable in principle.[12]

Still, our capacity to rationally communicate the most foundational of our beliefs may be limited in another way: by our inability to rationally *account* for them ourselves. Our foundational beliefs are our ultimate reasons – reasons that we do not account for by means of further argument. Even if we strive to render our systems of reasons and beliefs consistent and coherent, the boundedness of our reason – that is, the limits of our intellectual capacities, our inability to handle too many complex variables, paying attention to all possible interactions between them – means that we are unlikely to ever achieve rational closure. Even if we arrive at a reasonably well grounded normative argument, we and others are likely to be able to identify flaws, or at least able to point to alternative interpretations of the premises that support it.[13]

Yet, for each and every individual, such deliberation must at some point come to an end – at least temporarily. If there were no such end, individuals would be incapable of intentional action: we must always consider some reason to be good enough if we are to act. Intentional action comes with a commitment to certain beliefs, as Alan Millar observes. "We cannot act intentionally in familiar ways unless we have beliefs about the world around us. You can intentionally go to buy food in some location only if you have beliefs to the effect that food can be bought at that location, that you can get there by such-and-such a route using such-and-such means, and so forth."[14] Just like we cannot intelligibly account for our intention to act in a certain manner unless we are committed to (implicit) beliefs about the material circumstances involved, we cannot intelligibly account for said intention unless we are committed to (implicit) beliefs about our motivation to act. In other words, we cannot

---

[12]If we realise that we do fail at this hurdle of communicability, we surely have a reason to give up on our claim that others can come to share that specific moral belief. But that possibility does not mean that we have reason to assume that we cannot in principle succeed.

[13]This argument rests merely on the common experience of moral deliberation. It does not comment on the nature of morality, and hence on the possibility that consistency and coherence may be inappropriate or elusive standards.

[14]Millar, *Reasons and Experience*, p. 199.

act intentionally unless we have beliefs about why we deem acting in that particular manner to be appropriate in the given situation. We cannot intelligibly talk about our decision to go buy food, if we cannot, when asked, cite at least one reason that prompted us to direct ourselves towards that aim. At the point of action, there is a reason which, to the person in question, constitutes sufficient grounds for action. Because this is the case, any intentional action necessarily reflects the fact that for the person acting, deliberation on reasons for that action has – at least temporarily – come to an end. Individuals may have more specific, conscious reasons to put an end to deliberation and settle on convictions that they deem to be sufficiently well established. But the fact that we cannot escape acting on some reason is ultimately why it is possible to claim that each *agent* has to settle – however briefly or permanently – on a configuration of normative beliefs without seeking to establish them any further.

The normative beliefs which drive our actions in a given situation are necessarily accepted as *foundational* at the point at which we decide to act. They are foundational in the sense that, at that point, we do not accept them because of any further *substantive* reasons: we do not rationally infer to them from any further substantive beliefs. Any doubt or deliberation on alternative reasons that would require them to be supported by means of inference from other beliefs would only continue the process of deliberation, which must come to an end for action to be possible. Hence, the end of deliberation implies that whatever reasons we accept at that point, we accept as *groundless* reasons. This argument can be summed up as follows: insofar as we do have foundational beliefs – and we do, because, eventually, we need to accept some chain of reasons to act on – we are not able to account for them rationally by means of further substantive reasons.

This is not a defect of rationality. In fact, as Millar notes, "the very idea that we could conduct a global check on our groundless beliefs is incoherent since any checking would require us to take some things for granted which we have not checked."[15] Any judgement as to whether a particular belief passes muster is premised on our prior *acceptance* of the standard to which it is held. A failure to accept some standards as foundational would lead to an infinite regress of accounting for these standards. To believe that simply accepting a groundless belief is irrational would ultimately paralyse our judgement, and thus render intentional action impossible. Insofar as we think of ourselves as acting intentionally, putting an end to rational argument in justifying one's beliefs is not irrational.

Although our systems of reasons and beliefs necessarily have such rationally unsupported boundaries, these boundaries may well differ in the way in which individuals account for them. Some of these boundaries are:

(1) deliberate: we – often quite literally – put faith in a particular conviction. Such faith may be rooted in its great explanatory and supportive power for other lower-order beliefs, or in some other non-rational experience, but not in any further substantive beliefs; or

---

[15]Millar, *Reasons and Experience*, p. 214.

(2) incidental: our foundational convictions are the convictions that we happen to accept because we recognise that there are limits to our capacity to make sense of our normative (and material) experience of the world. These are not convictions that we affirm because we are convinced of their ultimate truth, but because we consider them in good faith to be the closest approximation to such truth that we can attain, given the conceptual boundedness of human reason and given that the there are limits to the time and mental effort we can spend on assessing our convictions.

This is not necessarily a binary distinction. Some individuals' reasons to accept some of their convictions as foundational may draw on both *(1)* and *(2)*. Others may reject *(1)* and could best be described by *(2)*, while nevertheless falling short of *explicitly acknowledging* the premises of *(2)*.

What is significant is that irrespective of the way in which we relate to our foundational reasons – be they (religious or secular) faith, be they acceptance of the fact that we cannot argue any further – these reasons cannot be communicated by means of rational argument. We cannot relate them to others via any rational argument, because we cannot account for them in that way ourselves.[16]

If individuals who are engaged in public reason recognise this limitation, they have reason to realise that they cannot consider the nonpublic reasons which they sincerely believe to be acceptable to others in general to be acceptable to them under the relevant conditions, that is, in the sphere of public reason. As long as a given reason is rooted in a foundational belief which individuals cannot assume others to share already, they must, in order to uphold their claim of acceptability, assume that others can come to share said foundation and, if necessary, the entire epistemic structure in which it is embedded. If others were only sufficiently like their interlocutors in terms of their epistemic structure as far as it pertains to the subject matter at hand, they would share their interlocutors' belief in the acceptability of the reasons that are proposed to them. Yet, those who propose nonpublic reasons cannot rationally account for the reason why their addressees should be like them in that respect – why they should share their respective foundational beliefs – precisely because these beliefs are foundational. And therefore, by extension, they cannot rationally communicate – that is, argue for – the pillars of the relevant parts of their system of reasons and beliefs, thus preventing them from inducing the required epistemic state in others within the process of public reason.

This conclusion does not call into question the integrity of individuals' comprehensive moral doctrines. Their inability to argue for the most foundational of their convictions on the basis

---

[16]This conclusion is related to what G. A. Cohen refers to as the paradox of conviction. Talking about the "beliefs that are central to our lives", he argues that "[i]t is an accident of birth and upbringing that we have *them,* rather than beliefs sharply rival to them, [...] [and] *we consequently do not believe as we do because our grounds for our beliefs are superior to those which others have for their rival beliefs.*" (G. A. Cohen. *If You're an Egalitarian, How Come You're So Rich?* Cambridge, MA: Harvard University Press, 2000, p. 9, original emphasis.) The focus of my argument differs slightly from Cohen's: while Cohen explores the question whether persistence in such a belief must necessarily be judged to be irrational – which he affirms – I argue that we cannot, in fact, escape such irrationality, since it is a condition for a belief to qualify as foundational.

of rational reasons neither needs to threaten individuals' firm belief in the validity of these moral doctrines and that of their derivative moral convictions, nor their genuine belief that they are, in fact, universally acceptable. After all, outside the sphere public reason, individuals are not required to ground their foundational beliefs in rational argument. Individuals are, however, required to recognise that the requirements for granting others access to these foundations are beyond the means afforded to them by public reason.

The implications of this conclusion are profound. Within the framework political liberalism, people are reasonable when they express their commitment to the ideal of persons as free and equal by treating society as a cooperative endeavour in which political decisions are backed by reasons that are acceptable to all. Considering individuals' internal perspective on what it means for their reasons to meet that standard, it seemed at first that it might leave some room for nonpublic reasons.[17] But as it now turns out – quite to the contrary – individuals' internal perspective yields reasons *not* to draw on nonpublic reasons when justifying a proposed policy to others. These reasons for restraint are implicit in the very foundations of those beliefs that are in conflict with individuals' political commitments as reasonable persons. This means that the reason not to argue for a particular policy on the basis of these beliefs is necessarily available to them at the very time when the normative authority of said beliefs appears to be hard – or even impossible – to discount. Whenever individuals feel the urge to draw on nonpublic reasons, they have a reason not to act on it if they *rationally* assess the character of their most foundational beliefs. As I have argued in chapter 3, rationality is integral to human experience and interaction, and the stakes are high for an individual to dismiss it as a standard for reasoning. As a result, it is difficult for individuals to deny that they have a reason to accept a norm that prohibits the proposal of nonpublic reasons in public reason. This means that restricting the constituency of public justification to those who draw on shared reasons alone is justifiable to all.

This resolves the key tensions within political liberalism that I have been concerned with throughout this dissertation. Drawing on the character of individuals' comprehensive moral doctrines, which are central to the texture of liberal, morally pluralistic societies, political liberalism can exclude those people from the constituency of public justification whose actions in the political sphere do not measure up to key liberal standards. It can do so in accordance with its commitment that justifications for coercive measures must be acceptable to all. In fact, my approach affirms a value central to liberalism: by engaging both with the bases of individuals' moral deliberations – their comprehensive doctrines and the beliefs they contain – and their internal perspective on these bases, it affirms that moral judgements must be valid to them within *their* frame of reference. Thus, the approach I proposed takes seriously political liberalism's commitment to individuals' moral autonomy.

---

[17]See section 6.2.

6.5 Conclusion

At the end of the previous chapter, we were confronted with political liberalism's difficulty of justifying or otherwise accounting for its prohibition on the use of nonpublic reasons in public justification. In this chapter, I proposed an analysis of individuals' internal perspective on nonpublic reasons as a promising source of such justifications. At first glance, a look at individuals' internal perspective suggests that their insistence on proposing nonpublic reasons does not necessarily indicate a deficiency in their commitment to respecting their fellow citizens as free and equal persons. From their internal perspective, the proposal of nonpublic reasons is compatible with their commitment to propose to others only reasons that the latter can accept.

It was my aim in this chapter to present the most charitable reading of individual citizens' internal perspective in order to furnish my interpretation of public reason liberalism with a conception of the individual person which is attuned to the diversity and complexity of individuals' sets of reasons and beliefs. Justifications for the exclusion of nonpublic reasons from the realm of public reason are all the more robust as a result of being rooted in such an account. While individuals' claim that they can genuinely believe their nonpublic reasons to be acceptable to all holds true, further examination of the character of their foundational beliefs shows that they cannot rationally believe to be able to persuade others of the validity of their reasons in public reason. They are unable to communicate their reasons for believing their nonpublic reasons to be valid by means of rational argument, because they themselves cannot rationally account for the reasons why the foundational beliefs underlying their nonpublic reasons possess for them the moral authority they clearly perceive.

It is for this reason that we may expect citizens to exercise self-restraint when they are inclined to introduce nonpublic reasons in public reason. We may do so because citizens *themselves* – in virtue of the very character of their nonpublic reasons – have reason to recognise that these reasons cannot be communicated within public reason.

Chapter 7

Conclusion

Moral pluralism is a fact of life in modern societies governed by free institutions. Freedom of conscience and freedom of expression enable a variety of comprehensive moral doctrines – religious and nonreligious – to develop and thrive among groups and individuals who seek answers to questions both of what constitutes a good life in general, and what they owe to each other as fellow human persons and citizens specifically. Political liberalism is mindful of liberal pluralism: it does not root its principles in a comprehensive philosophical argument which may be at odds with the metaphysical and moral beliefs of many comprehensive doctrines that are found in modern pluralistic societies. Instead, it argues that support for liberal political principles can be found in many of these doctrines if they meet the standard of reasonableness, providing support for the core liberal ideal of persons as free and equal and for accepting the consequences of the burdens of judgement. However, the pluralism produced by liberal institutions may also include views which are diametrically opposed to the liberal project thus defined, as well as views which are broadly compatible with the standard of reasonableness, but whose demands are not always aligned with what it requires. Hence, there is a gap between the pluralism produced by liberal institutions and the pluralism that political liberalism addresses.

It has been the key assertion of my dissertation that this gap is problematic for the following reason: respect for persons as ends in themselves endowed with moral autonomy is a fundamental tenet of any liberal political theory, and thus for political liberalism. What liberalism recognises as crucially valuable to an individual is her capacity as a moral agent to determine her actions according to the ends she recognises as valid. The idea of public reason pays heed to this commitment, stating that coercion is only legitimate for reasons that are acceptable to all: a person's actions must not be driven by reasons that elude her, however compelling they may appear to others.

There are two upshots to that liberal commitment to respect for the person: *(1)* justifications for coercive acts must be offered to *all* persons, irrespective of their worldviews; and *(2)* the pluralism represented in our theorising must reflect the complexity of individuals' moral

beliefs that result from their use of their moral autonomy. From *(1)* follows that the unreasonable – both those who are fundamentally unreasonable and those who merely fail to be fully reasonable – must not be excluded from the constituency of public justification without offering them justifications for their exclusion that they can accept. From *(2)* follows that we must not, from the outset, restrict the ambition of political liberalism to identifying which political principles would gain the support of *ideally* reasonable people and thus frame the constituency of public justification such that it only includes such individuals.

The liberal commitment to offer justifications to *all* renders the classic Rawlsian response to the challenge of the fundamentally reasonable – to contain them "like war and disease"[1] – untenable, while respect for the liberal conception of the person rules out Quong's internal conception of political liberalism as a solution to that challenge, as I argued in chapter 2. Similarly, with regard to the challenge of deep moral disagreement, it will not do to refuse to offer justifications to those who do not qualify as fully reasonable or dismiss their presence as irrelevant to the liberal project when it is interpreted as merely being concerned with the reasons ideally reasonable citizens would offer each other. Such idealisation is no solution: it is in conflict with a truly liberal conception of the person. Neither can we assert in this case that individuals' commitment to the ideals of reasonableness necessarily commits them to discount the significance of their comprehensive moral commitments whenever they clash with what reasonableness requires. Political liberalism cannot account for the requirement to refrain from introducing such nonpublic reasons into public justification. As I argued in chapter 5, political liberalism lacks the philosophical resources to make that kind of argument, precisely because its normative authority for each person is rooted in that person's individual comprehensive moral doctrine.

Yet, in order to live up to the two aforementioned core liberal ideals, political liberalism must not disregard the status of both the fundamentally unreasonable and those who fail to be fully reasonable within its theoretical framework. I argued that political liberalism must therefore strive to justify the demands of reasonableness to these two groups on terms which are acceptable to them as the morally complex individuals that liberalism is committed to taking them to be. In virtue of its theoretical nature, such a justificatory argument can only engage with the reasons we might assume these individuals to have, thus requiring us to attribute reasons to them. In chapter 3, I discussed the character that such externally attributed reasons need to possess if their attribution is to be compatible with the liberal commitment to respecting persons as ends: it requires that we address them with reasons that can be shown to be available to them from within their individual sets of reasons and beliefs. Rejecting a strongly externalist conception of reasons on the grounds that it is does not pay due respect to the above commitment to persons as ends, since it affirms the existence of reasons irrespective of their accessibility to the individuals to whom they are supposed to apply, I embraced a weakly externalist conception of reasons based on the idea of open justification. I argued that in order for reasons to be acceptable to an individual, they need to be accessible to them

---

[1]Rawls, *Political Liberalism*, p. 64.

in the sense that they are part of – or can be inferred from within – a person's internal belief system. I further argued that such a weakly externalist attribution of reasons is not vulnerable to objections to the validity of rationality as a standard for such inferences within a person's belief system, since people's mutual intelligibility in communicative situations is indicative of their commitment to the shared epistemic norm of rationality. I argued specifically that the claim that individuals do not strive for the elimination of inconsistencies within their set of reasons and beliefs is unsustainable.

Precisely this universal commitment to rationality provides us with a key resource of justifications for reasonableness both in the case of fundamentally unreasonable people, and in the case of those who fall short of being fully reasonable, having encountered deep moral disagreement. It is the fact that they cannot but seek rational consistency within their overall set of reasons and beliefs which allows us to draw on the beliefs implicit in individuals' actions, as well as in the character of their foundational beliefs as sources for justificatory reasons. I pursued these two argumentative strands in chapters 4 and 6, respectively.

Drawing on the idea of open justification in conjunction with the shared nature of epistemic norms, I argued in chapter 4 that a person's actions may serve as a source of beliefs and thus of reasons for her. This is due to the fact that some kinds of human activity may be said to be premised upon assumptions whose rejection would render their performance irrational and thus unintelligible to others. This is crucial for participants in a justificatory process, as one of the procedural norms governing public justification requires them to render the reasons they introduce into this procedure intelligible and accessible to all other participants.

I relied on these technical constraints in considering which reasons unreasonable people may be said to have for accepting their exclusion from the constituency of justification on the grounds of their unreasonableness. Drawing on a hypothetical scenario – asking whether unreasonable citizens could rationally sustain their unreasonable attitudes if they *were* allowed to participate in public justification – I explored two possible reasons upon which they could rest their justification for recommending fundamentally unreasonable proposals: attitudes of resentment and the denial of human personhood. These, as I argued, ultimately amount to a denial of the moral equality of some of their co-citizens. None of these reasons could intelligibly be sustained in public justification, with their introduction into a justificatory argument yielding a commitment to the agency of those whose moral equality the argument ultimately intends to deny. Finally, I assessed whether a commitment to recognising other persons as agents may also be deemed to yield a commitment to acknowledging their moral equality. Again, it is due to the procedural requirement of mutual intelligibility that a person who conceives of herself as an agent cannot consistently refuse to universalise others' demands to be attributed moral concern which she herself must also demand in virtue of her agency. As such, she is required to affirm the value which is foundational to an attitude of reasonableness – equal moral concern for all – thus giving her reason to assent to its use as a criterion for membership in the constituency of public justification.

These justifications, which we can offer to fundamentally unreasonable people for their exclusion from the constituency of public justification are insufficient to address the challenge of those people who merely fail to be fully reasonable as a result of encountering deep moral disagreement. They do not need to be given reasons for regarding their fellow citizens as free and equal, since they already relate to them in that way. What renders them less than fully reasonable is their failure to refrain from drawing on nonpublic reasons in public justification. Justifying their exclusion from the constituency of public justification to those who fail to be fully reasonable means to offer them reasons why they should base the justifications they propose to their fellow citizens on values which are shared by all, and I identified such reasons in chapter 6. First, I argued first that liberals have reason to consider how nonpublic reasons are perceived by the individuals who propose them and concluded that, from such an internal perspective, many nonpublic reasons actually appear to be acceptable to others. I concluded that we should therefore refrain from branding the proposal of nonpublic reasons as inherently indicative of an attitude of disrespect for the moral equality of their fellow citizens as ends in themselves, and thus indicative of a more fundamental kind of unreasonableness. I sketched an alternative *internal* conception of liberal respect, which zooms in on individuals' genuine beliefs about their reasons, their attitude towards and beliefs about their co-citizens, as well as the reasons the latter might find acceptable.

At first glance, this internal conception of liberal respect seemed to permit a wide range of nonpublic reasons to enter the sphere of public reason. Most beliefs which a person genuinely holds to be true could, from her perspective, lay claim to universal acceptability. And while this conclusion appears to open the floodgates to the use of nonpublic reasons in public justification, it actually proves to be a source of justifications for the very requirement to refrain from drawing on these beliefs as reasons in public justification. This is because we may question citizens' capacity to rationally substantiate that genuine belief in the universal acceptability of their nonpublic reasons within the specific constraints of public reason. They have, I argue, reason to believe that they cannot expect to render their foundational convictions – which ground their nonpublic reasons – accessible to others by the means available to them in the process of public reason: that is, by means of rational argument. What is wrong with nonpublic reasons from the perspective of an internal conception of liberal respect is not that those holding them are wrong to consider them to be acceptable to others, but that they cannot expect to *communicate* that genuine conviction within public reason. We can thus successfully argue that they have reason to accept the requirement of reasonableness to draw only on shared reasons in public justification. We can do so precisely by taking seriously the character and complexity of their comprehensive moral doctrines: the reasons individuals have to refrain from advocating policies based on what, to them, is fundamentally and obviously true are rooted in the foundational character of the very reasons that rendered individuals' relationship with their liberal commitments so complex in the first place. This strategy for justifying reasonableness hence enables political liberalism to address all people on terms acceptable to them, while taking seriously the character of the comprehensive doc-

trines affirmed by the kind of morally autonomous citizen who is championed within liberal theory.

Political liberalism can thus justify reasonableness. Having shown that fundamentally unreasonable people and those experiencing deep moral disagreement have reason to endorse the norms of reasonableness as a requirement for participation in public justification enables the scope of liberal political theory to be broadened without surrendering necessary restrictions to the membership of the constituency of justification that ultimately selects political principles for a society of citizens who regard each other as free and equal. Liberals may insist that unreasonable views of all kinds must be dismissed from procedures of public justification, while still living up to their commitment to treating all persons as ends in themselves.

Bibliography

Audi, Robert. *The Architecture of Reason: The Substance and Structure of Rationality*. Oxford: Oxford University Press, 2001.

Austin, J. L. *How to Do Things with Words*. Oxford: Oxford University Press, 1962.

Benn, Stanley I. *A Theory of Freedom*. Cambridge: Cambridge University Press, 1988.

Bird, Colin. "Mutual Respect and Neutral Justification". In: *Ethics* 107.1 (1996), pp. 62–96.

Cherniak, Christopher. *Minimal Rationality*. Cambridge, MA: MIT Press, 1986.

Cohen, G. A. *If You're an Egalitarian, How Come You're So Rich?* Cambridge, MA: Harvard University Press, 2000.

Connolly, William E. *The Terms of Political Discourse*. Princeton: Princeton University Press, 1983.

Davidson, Donald. *Inquiries into Truth and Interpretation*. 2nd edn. Oxford: Clarendon Press, 2001.

Dworkin, Ronald. *Justice for Hedgehogs*. Cambridge, MA: Harvard University Press, 2011.

Enoch, David. "The Disorder of Public Reason". In: *Ethics* 124 (2013), pp. 141–176.

Estlund, David. "The Insularity of the Reasonable: Why Political Liberalism Must Admit the Truth". In: *Ethics* 108.2 (1998), pp. 252–275.

Forst, Rainer. *The Right to Justification: Elements of a Constructivist Theory of Justice*. New York: Columbia University Press, 2012.

Friedman, Marilyn. *Autonomy, Gender, Politics*. Oxford: Oxford University Press, 2003.

Galston, William A. *Liberal Purposes: Goods, Virtues, and Diversity in the Liberal State*. Cambridge: Cambridge University Press, 1991.

Gaus, Gerald F. "Ideology, Political Philosophy, and the Interpretive Enterprise: A View from the Other Side". In: *Liberalism as Ideology: Essays in Honour of Michael Freeden*. Ed. by Ben Jackson and Marc Stears. Oxford: Oxford University Press, 2012, pp. 178–198.

— *Justificatory Liberalism: An Essay on Epistemology and Political Theory*. Oxford, New York: Oxford University Press, 1996.

— *The Order of Public Reason: A Theory of Freedom and Morality in a Diverse and Bounded World*. Cambridge: Cambridge University Press, 2011.

— "The Rational, the Reasonable and Justification". In: *Journal of Political Philosophy* 3.3 (2001), pp. 234–258.

Gaus, Gerald F. *Value and Justification: The Foundations of Liberal Theory*. Cambridge: Cambridge University Press, 1990.

Gewirth, Alan. *Reason and Morality*. Chicago, London: University of Chicago Press, 1978.

— "Replies to My Critics". In: *Gewirth's Ethical Rationalism: Critical Essays with a Reply by Alan Gewirth*. Ed. by Edward Regis. Chicago, London: University of Chicago Press, 1984.

Gigerenzer, Gerd. *Adaptive Thinking: Rationality in the Real World*. Oxford: Oxford University Press, 2000.

Habermas, Jürgen. *The Inclusion of the Other: Studies in Political Theory*. Cambridge, MA: The MIT Press, 1998.

Hare, R. M. "Do Agents have to be Moralists?" In: *Gewirth's Ethical Rationalism: Critical Essays with a Reply by Alan Gewirth*. Ed. by Edward Regis. Chicago and London: University of Chicago Press, 1984.

Haworth, Lawrence. *Autonomy: An Essay in Philosophical Psychology and Ethics*. New Haven: Yale University Press, 1986.

Herman, Barbara. "Moral pluralism and political consensus". In: *The Idea of Democracy*. Ed. by David Copp, Jean Hampton, and John E. Roemer. Cambridge: Cambridge University Press, 1995, pp. 270–291.

— "Pluralism and the Community of Moral Judgment". In: *Toleration. An Elusive Virtue*. Ed. by David Heyd. Princeton: Princeton University Press, 1996, pp. 60–80.

Hollis, Martin. "Reason and Ritual". In: *Rationality*. Ed. by Bryan R. Wilson. Oxford: Blackwell, 1970.

— "The Limits of Irrationality". In: *Rationality*. Ed. by Bryan R. Wilson. Oxford: Blackwell, 1970.

Iyengar, Shanto, Mark Peters, and Donald R. Kinder. "Experimental Demonstration of the Not-So-Minimal Consequences of Television News Programs". In: *American Political Science Review* 76 (1982), pp. 848–858.

Jacobs, Lawrence and Robert Shapiro. *Politicians Don't Pander: Political Manipulation and the Loss of Democratic Responsiveness*. Chicago: University of Chicago Press, 2000.

Kahneman, Daniel, Paul Slovic, and Amos Tversky. *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press, 1982.

Kant, Immanuel. *Groundwork of the Metaphysics of Morals*. Ed. by Mary Gregor and Jens Timmermann. Cambridge Texts in the History of Philosophy. Cambridge: Cambridge University Press, 2012.

Kelly, Erin and Lionel McPherson. "On Tolerating the Unreasonable". In: *Journal of Political Philosophy* 9.1 (2001), pp. 38–55.

Kinder, Donald R. "Communication and Opinion". In: *Annual Review of Political Science* 1 (1998), pp. 167–197.

Korsgaard, Christine M. *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press, 1996.

— "The origin of value and the scope of obligation". In: *The sources of normativity*. Ed. by Onora O'Neill. Cambridge: Cambridge University Press, 1996.

Kramer, Matthew H. *Liberalism with Excellence*. Oxford: Oxford University Press, 2017.

Kymlicka, Will. *Liberalism, Community, and Culture*. Oxford: Oxford University Press, 1989.

Larmore, Charles. "Political Liberalism". In: *Political Theory* 18.3 (1990), pp. 229–360.

— "The Moral Basis of Political Liberalism". In: *The Journal of Philosophy* 96.12 (1999), pp. 599–625.

— *The Morals of Modernity*. Cambridge: Cambridge University Press, 1996.

Macedo, Stephen. "Why Public Reason? Citizens' Reasons and the Constitution of the Public Sphere". Available at SSRN: `http://ssrn.com/abstract=1664085`. Accessed July 2015. Aug. 2010.

March, Andrew F. "Rethinking the Public Use of Religious Reasons". In: *Rawls and Religion*. Ed. by Tom Bailey and Valentina Gentile. New York: Columbia University Press, 2015, pp. 97–130.

Martignon, Laura. In: *Bounded Rationality: The Adaptive Toolbox*. Ed. by Gerd Gigerenzer and Reinhard Selten. Cambridge, MA: MIT Press, 2001, pp. 147–172.

McGuire, William J. "A Syllogistic Analysis of Cognitive Relationships". In: *Attitude Organization and Change: An Analysis of Consistency Among Attitude Components*. Ed. by Carl I. Hovland and Milton J. Rosenberg. New Haven and London: Yale University Press, 1960, pp. 65–111.

Millar, Alan. *Reasons and Experience*. Oxford: Clarendon Press, 1991.

Nagel, Thomas. "Moral Conflict and Political Legitimacy". In: *Philosophy & Public Affairs* 16.3 (1987), pp. 215–240.

— "Universality and the reflective self". In: *The sources of normativity*. Ed. by Onora O'Neill. Cambridge: Cambridge University Press, 1996.

Nisbett, Richard E. and Lee Ross. *Human Inference: Strategies and Shortcomings of Social Judgment*. Englewood Cliffs, NJ: Prentice-Hall, 1980.

Nussbaum, Martha C. "Perfectionist Liberalism and Political Liberalism". In: *Philosophy & Public Affairs* 39.1 (2011), pp. 3–45.

O'Neill, Onora. *Constructions of Reason: Explorations of Kant's Practical Philosophy*. Cambridge: Cambridge University Press, 1989.

— "Kant: Rationality as Practical Reason". In: *The Oxford Handbook of Rationality*. Ed. by Piers Rawling and Alfred R. Mele. Oxford: Oxford University Press, 2004, pp. 93–109.

— *Towards justice and virtue: a constructive account of practical reasoning*. Cambridge: Cambridge University Press, 1996.

Parfit, Derek. *On What Matters*. Vol. 1. Oxford: Oxford University Press, 2011.

— *Reasons and Persons*. Oxford: Clarendon Press, 1984.

Pew Research Center. *Political Polarization in the American Public*. `http://www.people-press.org/files/2014/06/6-12-2014-Political-Polarization-Release.pdf`. June 2014.

Presidential Candidates React: From 'Historic Victory' to 'I Will Not Acquiesce'. `http://www.bloomberg.com/politics/articles/2015-06-26/presidential-candidates-react-from-historic-victory-to-i-will-not-acquiesce-`. Accessed November 2016.

Quong, Jonathan. *Liberalism without Perfection*. Oxford: Oxford University Press, 2011.

— "The Rights of Unreasonable Citizens". In: *Journal of Political Philosophy* 12.3 (2004), pp. 314–335.

Rawls, John. *A Theory of Justice*. Cambridge, MA: Belknap Press of Harvard University Press, 1972.

— *Political Liberalism*. 2nd ed. New York: Columbia University Press, 2005.

— "Political liberalism: reply to Habermas". In: *The Journal of Philosophy* 92.3 (1995), pp. 132–180.

— "The Idea of an Overlapping Consensus". In: *John Rawls: Collected Papers*. Cambridge, MA: Harvard University Press, 1999, pp. 421–448.

Raz, Joseph. *Practical Reasons and Norms*. London: Hutchinson, 1975.

— *The Morality of Freedom*. Oxford: Oxford University Press, 1986.

Rovane, Carol. "Rationality and Persons". In: *The Oxford Handbook of Rationality*. Ed. by Piers Rawling and Alfred R. Mele. Oxford: Oxford University Press, 2004, pp. 320–342.

Sandel, Michael J. *Liberalism and the Limits of Justice*. Cambridge: Cambridge University Press, 1982.

Schwartzman, Micah. "The completeness of public reason". In: *Politics, Philosophy & Economics* 3.3 (2004), pp. 191–220.

Sher, George. *Beyond Neutrality: Perfectionism and Politics*. Cambridge: Cambridge University Press, 1997.

Simon, Herbert A. *Administrative Behaviour: A Study of Decision-Making Processes in Administrative Organization*. 3rd ed. New York: The Free Press, 1947.

Stevenson, Charles L. *Facts and Values: Studies in Ethical Analysis*. New Haven and London: Yale University Press, 1964.

Strawson, P. F. *Freedom and Resentment and Other Essays*. London: Methuen, 1974.

Ted Cruz: States should ignore gay-marriage ruling. `http://www.politico.com/story/2015/06/ted-cruz-gay-marriage-ruling-reaction-npr-interview-119559`. Accessed November 2016.

Tversky, Amos and Daniel Kahneman. "Availability: A Heuristic for Judging Frequency and Probability". In: *Cognitive Psychology* 41 (1973), pp. 306–329.

Waldron, Jeremy. "Autonomy and Perfectionism in Raz's Morality of Freedom". In: *Southern California Law Review* 62 (1989), pp. 1097–1152.

— *Liberal rights: collected papers, 1981–1991*. Cambridge: Cambridge University Press, 1993.

Wall, Steven. *Liberalism, Perfectionism and Restraint*. Cambridge: Cambridge University Press, 1998.

Williams, Andrew. "The Alleged Incompleteness of Public Reason". In: *Res Publica* 6.2 (2000), pp. 199–211.

Williams, Bernard. *Ethics and the Limits of Philosophy*. Cambridge, MA: Harvard University Press, 1985.

Williams, Jeremy. "Public Reason and Prenatal Moral Status". In: *The Journal of Ethics* 19.1 (2015), pp. 23–52.

Wittgenstein, Ludwig. *Philosophical Investigations*. New York: McMillan, 1953.

Wood, Allen W. *Kant's Ethical Thought*. Cambridge: Cambridge University Press, 1999.

Wright, Crispin and Martin Davies. "On Epistemic Entitlement". In: *Proceedings of the Aristotelian Society, Supplementary Volumes* 78 (2004), pp. 167–245.