Reasonably coercing the unreasonable? Investigating the challenge of unreasonable views to liberal political theory

Julia Christiane Gabriela Netter

New College

Thesis submitted in partial fulfilment of the requirements for the degree of MPhil in Politics in the Department of Politics and International Relations at the University of Oxford

Trinity 2013

Word count: 29222

Contents

Chapter 1

Introduction

> There is a nagging worry, occasionally expressed by some critics of political liberalism that the coercion of unreasonable citizens is either illegitimate or else renders political liberalism inconsistent because such citizens are coerced on grounds they do not accept. After all, the critic complains, is political liberalism not committed to the idea that the exercise of political power must be justifiable to each and every citizen?[1]

To proponents of political liberalism as, most notably, John Rawls and the above-quoted Jonathan Quong the latter question is merely rhetorical and is to be decisively denied. They argue that liberal political theory as an *ideal* theory does not need to be concerned about whether its principles, such as the requirement to justify the exercise of political power to all citizens, show respect for people who hold unreasonable views. Liberalism is only concerned with developing principles which are appropriate to structure political interaction among people who are committed to liberal values, that is, *reasonable people.* People who hold unreasonable views, by definition are considered to situate themselves outside the liberal framework, which is why they are not considered to be entitled to the same rights and privileges as reasonable people. This is also why their coercion on liberal grounds is not assumed to be problematic to political liberals. After all, the liberal principle of public justification is deemed to be applicable to reasonable citizens only,[2] thus permitting the access to what Quong refers to as 'the constituency of public justification'[3] to be restricted to the latter.

To limit the inclusion in the constituency of public justification to reasonable people, however, is to limit the scope of liberal concern in a way that conflicts with a commitment to a fundamental liberal ideal: treating human persons as ends in themselves.[4] Furthermore,

[1] Jonathan Quong. *Liberalism without Perfection*. Oxford: Oxford University Press, 2011, p. 312.

[2] Jonathan Quong. "The Rights of Unreasonable Citizens". In: *Journal of Political Philosophy* 12.3 (2004), p. 315.

[3] Ibid., p. 314.

[4] See, for example, Charles Larmore. "Political Liberalism". In: *Political Theory* 18.3 (1990), pp. 348–349.

to do justice to this commitment may be said to require coercive measures or principles to be justified to all human persons,[5] thus requiring membership in the constituency of public justification to be extended according to this criterion. Given that even unreasonable people need to be recognized as persons in this sense, there is indeed reason to consider their exclusion from public justification to be more problematic than proponents of political liberalism as an ideal theory are ready to admit. This issue is even more pressing since reasonableness itself even as a political conception may ultimately be shown to be rooted in a commitment to the moral equality of persons as ends in themselves, as I argue in chapter 2.

At the same time, considerations of stability and intra-theoretical consistency must not be dismissed when considering the inclusion of unreasonable people in the constituency of justification, maintaining support for limiting the inclusiveness of public justification as demanded by political liberalism. We are thus faced with a tension internal to liberal theory between a commitment to broaden and the necessity to restrict the scope of public justification.

My aim in this thesis is to explore a way of addressing this tension that lives up to both these extremes. I argue that one way to conceive of the tension is to regard restrictions to the membership in the constituency of public justification themselves as subject to a requirement of justification. In other words, in order to exclude unreasonable views from the public political process, those whose capacity for participating in collective self-determination is being restricted by this measure *must be shown to have reason to accept this restriction.* More precisely, it needs to be demonstrated that unreasonable people have reason to endorse the grounds upon which they are denied access to the constituency of public justification: that is, they need to be shown to have reason to endorse their unreasonableness as a valid criterion in this respect.

On an empirical level, this seemingly paradoxical requirement can hardly be expected to be fulfilled. After all, unreasonable citizens' convictions fundamentally contradict the reason they are supposed to accept. Moreover, in so far as liberal political theory eschews drawing on metaphysical foundations in order to substantiate its primary commitment to granting equal moral concern to all people by virtue of their autonomy, the reasons for endorsing such a position remain merely self-referential, as I argue in chapter 2. At first glance, thus, liberal principles appear to appeal only to those individuals who are already committed to liberal principles. This tautological conclusion does nothing to contribute to resolving the tension liberal political theory is subject to with regard to the exclusion of unreasonable views from the constituency of public justification.

To declare the attempt to resolve said tension a failure at this point would, however, be premature. It would only fail inevitably if individuals were to choose the set of their beliefs and convictions deliberately, entirely free from any constraints. Yet, epistemic con-

[5] Larmore, "Political Liberalism", pp. 348–349.

straints, mainly logical ones, play a role in framing the set, or network, of beliefs we may be said to have reason to hold – or, in other words, that we may be justified in holding, as I argue throughout chapters 3, 4, and 5. In asking whether people have reason to endorse reasonableness as a criterion for limiting the access to the constituency of public justification, we thus do not need to restrict ourselves to asking only what people deem themselves to be committed to. Instead, we may focus on the set of beliefs they may implicitly be said to have reason to hold by virtue of other factors governing what beliefs we ourselves may consistently endorse or reject. Far from irrelevant among those factors are our actions, some of which – crucially – determine what it makes sense for us to believe. This results from the fact that action may be premised upon the endorsement of certain beliefs in order to be rational. Consequently, our actions constitute a source of reasons it would be irrational to dismiss, but which does not always figure prominently within our deliberations on what we have reason to believe, as the commitments it gives rise to are mostly implicit and need to be uncovered by rational reasoning.

My argument in the following chapters takes advantage of this very fact. I intend to argue that unreasonable citizens may be said to be committed to endorsing the reasons for their exclusion from the constituency of public justification, or, more precisely, that they have implicit reasons to do so, despite the fact that they may not actually be prepared to consciously acknowledge them.

Such an argument, however, presupposes an enquiry into whether attributing reasons to others and establishing an argument for coercion is compatible with liberal values. After all, I emphasize what we, as external observers, could expect individuals to consider themselves to be rationally bound to believe, and not necessarily what they actually recognize themselves as being bound to believe. Thus, in excluding the bearers of unreasonable views from the public political discourse on such grounds, their coercion would not be sanctioned by their actual assent to said measure, but by their hypothetical endorsement, with the hypothesis being a conditional one: that they have reason to accept the grounds of their exclusion from the political realm, *if* they reasoned rationally based on all their potential sources of reasons.

This raises questions about what it means to *have* a reason: do the reasons we have consist in those we are conscious of and actively acknowledge, or do they include those we have so far failed to acknowledge, because our rationality is not perfect but bounded? Moreover, can coercion be legitimate if its justification draws on reasons we are supposed to have, but which have only been recognized by others and not by ourselves? In chapter 3, I therefore discuss the extent to which such external attributions of reasons as justifications for coercion can be considered to be compatible with the liberal commitment to treating persons as ends in themselves.

Chapter 4 then goes on to explore individual action as a potential source of reasons for unreasonable citizens to consider reasonableness as a justified criterion for limiting ac-

cess to the constituency of public justification. Drawing on the assumptions a particular course of action may be premised upon in order to be rationally intelligible to others, I argue that by virtue of performing a particular action, an individual may incur commitments to beliefs. In this context, my analysis focuses on the impact of unreasonable citizens' role as aspiring members of the constituency of public justification on both the actions they may be required to perform as well as the epistemic standards they are subject to. As I am solely concerned with the reasons owed to unreasonable citizens who *intend to* achieve their aims within the realm of a justificatory political process, my enquiry restricts itself to beliefs they may be said to be committed to *as a result of acting as a regular participant in public justification.* Chapter 4 ultimately proposes an argument demonstrating that the constraints of this role may be considered to yield a commitment to recognizing other members of the constituency of public justification as agents.

Yet, this commitment, as I argue, is merely a factual one. Citizens, including those holding unreasonable views, may be said to incur a commitment to recognizing the *fact* that those individuals they interact with in the political realm are agents, but this is not equivalent to being compelled to recognizing the latter's moral equality as a result. In other words, one may recognize oneself and others as agents, but still deny to them the same moral concern one claims for oneself. Demonstrating that the recognition of others as agents entails their recognition as subjects worthy of being considered as morally equal to oneself requires a further argument linking both concepts. Again, I hold that *in a political context* – a context in which people are required to justify their proposals to all other participants of the public political discourse – citizens' requirement argue and act intelligibly renders them unable to deny the link between other individuals' agency and the need to recognize them as objects of moral concern.

My argument in chapters 4 and 5 thus heavily draws on rationality – not as an epistemic norm in general, but as a requirement imposed by the procedures of a public political discourse among members of the constituency of public justification. This limitation to the scope of the argument is not to be considered a defect, as it matches the nature of the challenge it is intended to address: a tension within liberal *political* theory, which raises concerns about how to adequately treat citizens within political frameworks characterized by liberal values. I am thus not concerned with the justifications individuals have reason to endorse in general, but with those they can be said to be committed to as aspiring participants in a public political discourse among constituents of justification.

Hence, it is the very role unreasonable people aspire to in claiming admission to the constituency of public justification that gives them reason to recognize all other members of said constituency as morally equal to themselves, and thus to reject their fundamentally unreasonable attitude to certain others as inadequate within this political realm. More generally, it gives them reason to accept their unreasonable views as valid grounds for being excluded from the constituency of public justification, since these views are not rationally reconcilable with the premises of the actions their membership in said con-

stituency requires them to perform.

Chapter 2

A tension within liberalism

2.1 Introduction

Fundamentally unreasonable views give rise to a tension within liberal political theory, as their existence among the attitudes citizens hold in public presents a challenge to its internal consistency. While, in principle, being committed to a requirement of universal justification on a procedural dimension, it also needs to ensure the protection of liberal values within the substantive outcomes yielded by these procedures. Achieving the latter aim will typically not allow unreasonable views to be taken into account, since, by their very nature, they will not give the individuals holding them reason to endorse policy proposals which foster liberal values. Yet, dismissing those views conflicts with the liberal commitment to universal justification and its underlying core principle of granting equal moral concern to all persons.

My aim in the following chapter is to explore this tension in detail, explaining why liberal theorists can neither refrain from confronting the challenge of unreasonable views by excluding them from the constituency of public justification solely by virtue of their unreasonableness, nor respond by the practical extension of said constituency to include unreasonable views. I shall argue that in order to address the tension brought about by the necessity to exclude unreasonable views and their proponents from the constituency of public justification – which qualifies as a coercive act – liberals must honour their commitment to providing justifications, in particular to those whose freedom is being limited as a result of said coercion. In other words, they must be prepared to justify to the unreasonable their subsequent exclusion from the constituency of public justification. This requires liberals to engage in an argument as to whether their criterion regulating the access to the constituency of public justification could also be made acceptable to citizens who, empirically, hold unreasonable views. In other words, they need to ask whether unreasonable citizens have reason to endorse said criterion.

To support this conclusion, I shall first set out the basic dilemma liberal political theory

is faced with, arguing that reasonableness as a criterion governing access to the constituency of public justification is ultimately self-referential, failing to engage in a justificatory discourse with those whose freedom it is bound to constrain. I shall then proceed to focus on the character of such a justificatory discourse, which I believe to be capable of successfully addressing the tension unreasonable views generate within liberal political theory.

2.2 Inclusiveness and exclusiveness in public justification

The tension within liberal political theory which I intend to explore in this chapter arises from a conflict between two of its essential traits: its maximal inclusiveness in terms of justification on the one hand, and its inability to maintain this inclusiveness in the establishment of political procedures based on these principles on the other.

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, to respecting and protecting the capacities she is endowed with by virtue of her autonomy, reason and agency.[1] What liberalism recognizes as crucially valuable to an individual is her capacity as a subject to determine her actions according to her ends.[2] Respecting this value could thus 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, as Charles Larmore emphasizes in a Kantian vein.[3]

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 conductive 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,[4] 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.'[5] 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,[6] enabling them to recognize which external claims can or cannot

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

[2] Ibid., p. 41.

[3] Larmore, "Political Liberalism", p. 348.

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

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

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

legitimately demand to be included in their realm of ends. Again, emphasizing the liberal commitment to the individual as the standard for the adequacy of intersubjective 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.'[7] 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: 'If 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 by other arrangements, for the status quo has made no claim to *his* allegiance.'[8] The purposes underpinning state action need to be acceptable to all. In order to ensure this to be the case, liberal political procedures need to address themselves to all citizens (with citizens 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',[9] to be maximally inclusive.

However, liberal political theory also relies on a certain degree of justificatory exclusiveness. Although, as Jonathan Quong argues, it may be wrong to deprive citizens of all rights and liberties they are granted by liberal political principles, liberals may need to restrict admissions to the constituency of public justification and may rightfully do so with regard to unreasonable citizens.[10] Quong holds that the reason based on which unreasonable citizens 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.[11]

Quong thus assumes that the requirement of justification does not apply to unreasonable

[7] Ibid., p. 44.
[8] Ibid., p. 44. Original emphasis.
[9] Quong, "The Rights of Unreasonable Citizens", p. 314.
[10] Ibid., p. 314.
[11] Ibid., p. 315.

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 realizing 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, 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 their origin in the distinctly human capacities of autonomy, rationality and agency and the liberal commitment to treating persons as ends. In other words, it claims that the liberal principle of justification only needs to be applicable to liberals, while disregarding the fact that this very principle is rooted in the liberal commitment to treating persons as ends – a commitment which refers to their quality as persons, not their attitudes to others. Citizens deserve to be offered justifications, not by virtue of their own willingness to do so to others, but owing to their quality as persons. References to the nature and the content of citizens' unreasonable doctrines themselves are hence unconvincing as 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 any less than 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.*'[12] They hold that the inclusion of those citizens in said constituency is more compatible with ensuring 'the greatest range of equal rights and liberties for all'[13] than denying to them the right to be offered justifications.[14]

It is, however, doubtful whether the toleration of such attempts to promote illiberal val-

[12]Erin Kelly and Lionel McPherson. "On Tolerating the Unreasonable". In: *Journal of Political Philosophy* 9.1 (2001), p. 54. Original emphasis.
[13]Ibid., p. 42.
[14]Ibid., p. 55.

ues by liberal means actually remains compatible with the protection of equal rights and liberties in the long run. It is this danger that provides liberals with another, more compelling argument for limiting the inclusiveness of the constituency of public justification. If they could not prevent unreasonable doctrines from being included in the constituency of public justification at all, liberal polities might be barred from effectively defending their fundamental ideas and principles in public justification. While a prudential consideration in practice,[15] avoiding such a internal conflict becomes subject to a principled argument on the theoretical level: if the principle of universal justification were to be universally and unconditionally valid, it would require illiberal views and proposals to be treated with the same respect as those in line with liberal ideals. As a result, liberals could only consistently defend their ideals on a procedural meta-level, being barred from principled discussion as far as the content of individual policies is concerned. Given that the latter might ultimately also question said procedural principles, liberals would be trapped in a fundamental conflict: true commitment to liberal ideals in terms of both procedures and the realization as well as the protection of substantive liberal values in the rights, liberties and benefits that form part of any particular policy cannot be upheld without setting boundaries to either the content of individual legislation, or to participation in the constituency of public justification. Without such limits, a commitment to liberal values would thus be led *ad absurdum*, with people ultimately being bound to respect illiberal positions out of liberal motives.

In order to maintain consistency within liberal political theory, 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 characterised by a commitment to the moral equality of all people as well as a disposition for toleration is one such theoretical safeguard.

2.3 The self-referentiality of reasonableness

It is these dispositions that are embodied by the idea of 'reasonableness', an ideal characteristic which is common in liberal political theory as a criterion for regulating access to the constituency of public justification, most notably in John Rawls's *Political Liberalism*.[16]

[15] 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).

[16] See also Barbara Herman. "Pluralism and the Community of Moral Judgment". In: *Toleration. An Elusive Virtue*. Ed. by David Heyd. Princeton, N.J.: Princeton University Press, 1996, pp. 60–80; Barbara Herman.

Two dimensions may be distinguished with regard to Rawls's conception of 'reasonableness': a cognitive and a moral one.[17] For Rawls the cognitive dimension dimension of reasonableness is expressed in individuals' 'similar powers of thought and judgment' and their ability to 'draw inferences, weigh evidence, and balance competing considerations' – capacities that people share by virtue of a 'common human reason'.[18]

Going beyond these basic cognitive capacities, Rawls also conceives of reasonableness as a richer, more substantial moral concept. He associates this moral dimension of reasonableness with 'first, the willingness to propose and honour fair terms of cooperation, and second, with the willingness to recognize the burdens of judgment and to accept their consequences.'[19] 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 justifiable to them; and they are ready to discuss the fair terms that others propose.[20]

This moral definition of reasonableness reflects Rawls's fundamental idea of persons as free and equal. Reasonable people are willing to show respect to others in the above sense, because they recognize 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 judgment, thought, and interference connected with these powers)'[21] and equal in being owed just treatment. In his original treatment of this topic in *A Theory of Justice*, Rawls emphasizes that '[t]hose who can give justice are owed justice.'[22] Reasonable people can thus be considered to share a commitment to treating all those others as ends who are capable of recognizing others as ends. They regard others as beings capable of determining their actions based on their purposes according to their conceptions of the good. They also deem them to be able, by virtue of their powers of reason and their capacity for a sense of justice, to discern and object to any kind of treatment that does not pay sufficient respect to their and others' status thus defined. Rawls's conception of persons as free and equal is explicitly concerned with moral personality as a potentiality. Citizens respect for each others as free and equal persons is therefore independent of the degree to which

"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.

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

[18] John Rawls. *Political Liberalism*. 2nd ed. New York: Columbia University Press, 2005, p. 55.

[19] Ibid., p. 49.

[20] Ibid., p. 49.

[21] Ibid., p. 19.

[22] John Rawls. *A Theory of Justice*. Oxford: Oxford University Press, 1972, p. 510.

a person realizes 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',[23] arguing that only very few human beings can be considered to lack this attribute.[24] It is because of this quality that they assign to themselves as well as their co-citizens that reasonable people are committed to proposing fair terms of cooperation and to honour the burdens of judgment where disagreement about fundamental comprehensive issues persists.[25]

It seems to be this moral dimension of reasonableness that serves as the criterion 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, for Rawls, is crucial in 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.'[26] Relating to political constructivism, Rawls also affirms that '[r]easonableness is its standard of correctness, and given its political aims, it need not go beyond that.'[27] The latter statement refers to reasonableness in its moral dimension, regarding 'the principles of practical reason in union with *conceptions of society and person*'[28] as the only necessary prerequisites of 'a reasonable and workable political conception'. Reasonableness is thus defined in terms of the particular moral attitude to social interaction that Rawls considers to be appropriate within a liberal society.

The Rawlsian public political discourse thus yields reasonable political principles and policies as the only legitimate outcomes. It is therefore hardly surprising that Rawls explicitly limits the range of permissible attitudes and positions within the liberal political discourse to reasonable (comprehensive) doctrines. According to Rawls '[t]hese are the doctrines that [...] political liberalism must address.'[29] Unreasonable views, in contrast, do not merit inclusion in the constituency of public justification. In fact, Rawls asserts that their presence in society entails on liberals the 'practical task of containing them – like war and disease – so that they do not overturn political justice.'[30]

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 necessity to exclude those individuals who pose a threat to the realization 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, which arises from the

[23] Ibid., p. 505.
[24] Ibid., p. 506.
[25] Rawls, *Political Liberalism*, p. 49.
[26] Ibid., p. 90. Emphasis added.
[27] Ibid., p. 127.
[28] Ibid., p. 127. Emphasis added.
[29] Ibid., p. 36.
[30] Ibid., p. 64.

merely self-referential nature of the liberal ideal of reasonableness, precluding its merit from being comprehensible and defensible externally. It is hardly possible to *explain* – rather than justify – 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 objective 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, on Rawls's account, for a person to reason practically is to ask how her reasonable disposition requires her to act. This is because Rawls defines 'the principles of their common practical reason'[31] 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.'[32] 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 the reasonableness, i.e. the dispositions which political liberalism considers to be appropriate in political interaction among free and equal persons. Hence, practical reason in a Rawlsian sense cannot provide an external observer with independent variables accounting for the normative significance of reasonableness, since it is itself defined in terms of liberal values.

As a result, reasonableness is basically incontestable from within a liberal framework, but, apart from self-referential affirmation, remains unable to provide reasons to those who, so far, do not endorse liberal political ideals. A commitment to, or, at least, a thorough understanding of liberal values, thus remains the prerequisite for being able to appreciate the normative significance of reasonableness in a political framework.

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

> political liberalism must find some way to penetrate this plurality of insular groups.[33] 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

[31] Rawls, *Political Liberalism*, p. 90.

[32] Ibid., p. 107.

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

> the ones with rejection rights and we would not. The question is [...] whether we can or cannot suspend judgment on it. We cannot, since suspending judgment would leave us with a plurality of insular groups, none evidently having a better claim to be authoritative than any other.[34]

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. In conjunction with liberalism's universalist commitments to justification, this elusiveness is problematic, for – to echo Waldron's statement – the lack of a justificatory intention deprives liberal polities of a legitimate claim to the allegiance of those citizens outside the insular group of the reasonable.[35]

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 well not violate the broader liberal requirement to provide justifications to all individuals. In such an ideal society consisting 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 not live up to their potential as moral persons and do, in fact, hold unreasonable doctrines, the failure to provide these citizens with reasons for their coercion – and, more specifically, reasons that could be valid for them (and not only for those who already share liberal values) – violates liberalism's universalist commitment to justification which is based on its inclusion of all people in the realm of ends. The latter commitment which is also acknowledged by Rawls, recognizing all human beings as free and equal only in virtue of their *potential* for moral personhood.[36]

However, given this conflict, how can liberalism morally cope with the condition of pluralism in modern societies? For modern liberal, but non-ideal, pluralistic societies, it is empirically highly unlikely to consist solely of reasonable citizens. So, empirically, liberals cannot avoid being confronted with citizens holding unreasonable views. If their liberal ideals are to be of any practical relevance to them, they need to allow them to effectively protect and promote their liberal values, without, in doing so, failing to live up to some of them. The latter, however, appears to be hardly avoidable, revealing a fundamental tension within liberalism: for the sake of their ability to remain internally consistent in practising their liberal ideals, liberals need to set limits to the inclusiveness of the constituency of public justification, barring all unreasonable doctrines and their proponents from participating in the public political discourse.[37] At the same time, their fundamental commitment to treating all persons as ends imposes upon them a requirement to

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

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

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

[37] See section 2.2.

justify their actions to all individuals whose freedom is being constrained by them. This latter norm, requiring the constituency of public justification to be maximally inclusive, however, clashes with the need to limit its inclusiveness, because the limiting criterion – reasonableness – is merely self-referentially justified. As I have argued in this section, it does not address itself to, or provide reasons to, anyone outside the group of reasonable people as to why it should be furnished with the authority to regulate admissions to the constituency of public justification. Yet, this is exactly what the liberal norm of universal justification requires by virtue of its concern for treating others as ends. Consequently, liberals are caught in a dilemma between a normative requirement for the constituency of public justification to be universally inclusive, a prudential and rational need to limit its inclusiveness, and the fact that the criterion they need to draw on in order to achieve the latter, by its very nature, violates the former norm.

Superficially, the self-referential justification of reasonableness as a principle regulating access to the liberal political realm may be considered an asset in terms of inclusiveness, because it renders liberal principles self-sufficient, maintaining their independence of any comprehensive metaphysical doctrines. It thus renders liberal principles accessible to a variety of people who draw their essentially reasonable convictions from a variety of sources, some of which might consider themselves to be unable to endorse liberal principles if they claimed to be based on specific, comprehensive metaphysical foundations. However, not to recognize the effect of said self-sufficiency on its prospects for achieving support outside the community of the reasonable is to perpetuate the tension within liberalism between the poles of justificatory in- and exclusiveness.

Given these conclusions, liberals need to regulate access to the constituency of public justification in a way that maintains the self-sufficiency of reasonableness as a criterion governing said access (thus remaining on a political level, eschewing controversial philosophical foundations), while also allowing them to live up to their universalist commitment to justification.

If the latter commitment requires coercive measures to be justified to all people, liberals need to provide unreasonable people at least with a justification for their exclusion from all further justifications. In other words, liberals 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 constitute an expression of disrespect for their equal moral value, for their entitlement to be treated as ends, not means. Yet, if unreasonable people themselves recognized 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 solely 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 liberals to live up to their

universalist commitment to justification.

Since the self-sustainability of reasonableness does not allow liberals to draw on any potentially shared comprehensive doctrines or other fundamental convictions which might induce an unreasonable individual to recognize reasonableness – or the values it incorporates – as the only adequate attitude within liberal political discourses and processes, the only remaining source of reasons justifying reasonableness as criterion for their exclusion consists in these very practices themselves. Hence, reasonableness needs to be justifiable to unreasonable people from within the political framework. As liberals, we thus need to search for reasons that their inclusion 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 would hence be based on the following conditional: *if* unreasonable people sought admission to the constituency of public justification[38], 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.

In the following chapters, I aim to demonstrate that such an argument justifying reasonableness to unreasonable citizens as the grounds for their exclusion from the constituency of public justification can indeed be made. I shall argue that their participation in the public political discourse as members of the constituency of public justification would provide unreasonable citizens with reasons to act reasonably within this very role. More precisely, my argument intends to demonstrate that the actions they would be required to perform in this capacity yield a commitment to the values underlying reasonableness: the recognition of others as ends and their moral equality (i.e. their equal worthiness of being granted moral concern). This argument is both hypothetical and epistemic: it is concerned with the conclusions unreasonable people would be rationally committed to draw, given the actions they would be required to perform as hypothetical members of the constituency of justification, irrespective of whether they would actually do so. Consequently, it will *attribute* reasons to them – reasons they should be able to endorse if they were rational.

There is, however, potential for controversy regarding the question whether an argument which is relying on citizens' hypothetical endorsement of certain reasons – reasons which are thus (merely) being externally attributed to them – can be considered sufficient in terms of justification for coercive measures. The following chapter therefore investi-

[38]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.

gates whether, and subject to which conditions, coercive measures based on externally attributed reasons may be said to respect individuals who are subject to these measures to a sufficient degree for them to be considered to be treated as ends instead of means. The latter is crucial, as any argument attempting to ease the tension within liberalism would ultimately end up perpetuating it, being guilty of the same lack of respect to universalist liberal values as the position it is designed to overcome. Therefore, I proceed by first discussing the character of the reasons unreasonable citizens may be supposed to have for recognizing reasonableness as an appropriate criterion regulating access to the constituency of public justification, before going on to examining said reasons' origin and content in chapters 4 and 5.

Chapter 3

Attributing reasons to others

3.1 Introduction

This chapter is concerned with the character of reasons we may draw on when arguing that unreasonable people have reason to act reasonably while participating in the public political discourse. As these reasons are being attributed externally, it is important to ask about the standard for judging whether the reason can actually be considered to apply to the individual in question and thus be a reason *for her*. In other words, what criteria 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 unreasonable citizens be said to have reason to be reasonable in political interaction, because *it is the case* that reasonableness is the only appropriate attitude in this context? Or must the reasons a person is supposed have bear some relation to a person's internal mental set-up? In this case, we might only be justified in coercively excluding unreasonable people from the constituency of public justification if the reasons we offer them can be shown to be supported by other beliefs they hold. One crucial aspect in determining the appropriate character of the justifications unreasonable people are offered consists in its compliance to the liberal commitment to treating all people as ends. If an argument that attempts to justify to unreasonable people their exclusion from the constituency of public justification out of a commitment to their entitlement to be offered justifications is to be convincing, it must ensure that the character of the reasons it deems to be acceptable to these people does not itself violate said commitment.

I begin 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 possible inferences within a person's existing, internal set of reasons and beliefs as the standard for the applicability of reasons. I finally 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, and conclude that this is not the case because intersubjective intelligibility presupposes a common set of those norms.

3.2 Strong externalism of reasons

I begin this enquiry with an assessment of strong externalism of reasons. According to this perspective, whether or not a person – I will call her 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 set-up, but rather is determined by the correspondence of *X* to a fact that obtains objectively and independent of Alice's individual perspective and is therefore entirely external to Alice.[1] 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.'[2] 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.*'[3]

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 case that we know[4] *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, we would agree on the fact that the chair in front of me is blue to be a good reason to believe that the chair in front of me is blue. Now, it is hardly surprising that, by virtue of this knowledge ("*X* is true"), we come to doubt 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), it is certainly sensible for me to judge 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 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:

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

[2] Ibid., p. 33. Emphasis added.

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

[4] 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.

> 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 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.[5]

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 particular person such as Alice *has* a reason to believe that the chairs are blue, and thus to *attribute* such a reason to her?[6] 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 which she should adapt her actions to, and what she herself is able to recognize 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 recognized 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. Neither did, for example, liberals, who are committed to a strongly externalist conception of reasons, if they claimed that the reason for Alice to be reasonable if she desires to be admitted to the constituency of public justification consists in it being the case that it is only appropriate for reasonable people to be part of said constituency. Now, in attributing such a reason to her when arguing that her exclusion from the public political discourse is justifiable to her, they assume that Alice must recognize their claims about what they deem to be objective facts as authoritative for her choice of action. At the same time, they deliberately refuse to offer Alice any supporting evidence for the truth of their claim, because, as strong externalists, they do not believe she even needs to be

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

[6] The distinction between the reasons there are 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).

aware of or recognize a reason in order for there to be one for her.

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 to them on their own terms, but only by reference to external authority. For liberals, however, this is a deeply problematic stance to take, given their commitment to treating others as ends, not means.

As Rainer Forst points out, such 'respect for moral persons as "ends in themselves means" that one recognizes their right to justification and the duty to be able to give them appropriate reasons.'[7] According to Forst,

> justifying reasons must in principle be accessible and agreeable to every reasonable person. In other words, a moral person must be able to *take responsibility for* his or her actions before affected others [...].[8]

A person cannot live up to this responsibility if she is unable to assess the reasons which are supposed to guide her actions in terms of their validity to her. 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 accept 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* recognize 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 recognize the reason of the intervention as a valid end to herself. Since this makes *her* judgment the standard for the legitimacy of coercion, her judgment on the validity of the reason justifying said coercion cannot be dispensed with. To do so would only reiterate the process of justification on another level: the level of reasons. 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 a judgment 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 set-up – the context which provides the criteria based on which individuals judge the validity of a reason. Strictly externalist

[7] Rainer Forst. *The Right to Justification: Elements of a Constructivist Theory of Justice*. New York: Columbia University Press, 2012, p. 21.

[8] Ibid., 19. Original emphasis.

attributions of reasons cannot insist on such relations as a necessary condition for a person to be considered to have a reason. For, as soon as the reason for a person to reject or endorse a given statement, consists in *her* recognition of the validity of a statement and not in its truth independent of her judgment, it is no longer strictly external.

However, proponents of a strongly externalist conception of reasons deny that treating other people with respect requires the reasons they are offered to be 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.[9]

An explanation 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 upon whom the speaker's action is inflicted. Therefore, providing an explanation for our actions is not (necessarily) an expression of respect.[10] As Gaus remarks, 'murderers often explain *their* reasons, but we can hardly take this as showing respect for others.'[11] Explanation is merely an effort to set out one's own reasons for believing or acting, whereas justification asks whether one's interlocutors have reason to consider a given belief or action acceptable. 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 it does not provide a satisfactory approach to discerning what reasons may be *accessible* to other people when considering whether to endorse or reject the claims they are confronted with.

It is for this reason that strong externalism of reasons must be rejected for the purpose of my main enquiry: 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 motivate my emphasis on the necessity of justification: the respect for people's autonomy, their ability to reason, as well as their capacity for individual self-determination.

[9] William A. Galston. *Liberal Purposes: Goods, Virtues, and Diversity in the Liberal State*. Cambridge: Cambridge University Press, 1991, p. 109.
[10] Gaus, *Justificatory Liberalism: An Essay on Epistemology and Political Theory*, p. 141.
[11] Ibid., p. 141.

3.3 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 civilizations, 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 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 that the chairs are blue, even though this belief would violate a long-held and so far valid assumption. 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 equal, from his point of view there is nothing to count in favour of believing my statement rather than Bob's, since its very substance contradicts everything that until now has proved to be sound for Alice to believe.

Drawing on this rather crude example, I illustrated the reliance of what most would consider to be a sound attempt to evaluate one's beliefs and the reasons one is presumed to have for holding them on the set of beliefs that currently provide us with an understanding of our environment. As Gaus notes, 'to have justified belief [i.e. a belief one has reason to hold], one has to make the connection between the belief and the relevant considerations.'[12] Such considerations are beliefs which, as individuals, we have already come to recognise as plausible *explanantia* of other phenomena, and which, therefore, can

[12]Gaus, *Justificatory Liberalism: An Essay on Epistemology and Political Theory*, p. 19.

be considered as beliefs that apply to us insofar as they provide the underpinnings of our understanding of the world: we have thus internalized them. In turn, they enable us to scrutinize 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 have 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 scrutinize the beliefs and reasons that are supposed to apply to her. 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 subject to these acts on their own terms – hence being made accessible to them – we must not attribute to them reasons and beliefs that clearly violate this condition in the way described above. Attributed beliefs and reasons must therefore 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.

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, as 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 a relevant set of beliefs she has come to accept in the past. So, 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 hardly intelligible to her on her own terms. However, upon further inspection of Alice's set of beliefs, we might realize that, in fact, she can be said to have several potentially relevant beliefs she did not take into account when making her initial judgment, 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 the first-hand experience of having received this bar of chocolate may be said 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 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 or not 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 she does not reflect all beliefs which are relevant to the claim she intends to evaluate is neither surprising nor need it be a sign that Alice 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 recognized than relations among beliefs within one compartment'.[13]. Consequently, individuals may not typically 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 can no longer be totally unintelligible 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

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

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.'). The acceptance of Y and Z does not by 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 its seeming contradiction to 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 recognizing as part of her belief system. Such efforts to identify possible reasons the person in question may have for endorsing or rejecting certain beliefs are part of a process which Gaus refers to as 'open justification'.[14] 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 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.[15]

In the previous example, I showed that Alice indeed seems to be committed to accepting some new information. By 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 recognizes, can it then be considered disrespectful to her

[14] See Gaus, *Justificatory Liberalism: An Essay on Epistemology and Political Theory*, p. 31.
[15] Ibid., p. 32.

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 being 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. 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 recognize the reasons she is attributed, 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 recognize as a reason. Yet, it is still adopting an external perspective inasmuch as the authority judge which potential reasons may be said to fulfil this criterion rests with the external observer. It 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.'[16] 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 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 judgment. Adopting Gaus's terminology, I refer to the external attribution of reasons within a framework of open justification as *weakly externalist*.[17]

In asserting that weakly externalist attributions of reasons may be considered to demonstrate respect for the autonomy and capacity for self-determination of those individuals who are supposed to recognize said reasons, 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 with regard to reasons she is supposed to have for endorsing a given belief, if she could rightfully argue that, *to her,* the justification she is presented with for accepting said belief is unintelligible. Thus, if an individual could convincingly argue that epistemic norms may be considered subjective rather than objective rules of reasoning, claiming to be committed to an idiosyncratic set of such norms, externally attributed

[16] Gaus, *Justificatory Liberalism: An Essay on Epistemology and Political Theory*, p. 32.
[17] Ibid., p. 32.

reasons could not be deemed valid any more, 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 could be sustained.

3.4 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 recognize 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 others'. What might seem to be a valid external inference within her belief system might thus still fail to make any sense to her, rendering this line of reasoning unintelligible to her.[18]

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 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:*

[18] Consequently, she would also be rendered unable to access the reasons that might have identified in this process.

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 provide her with a reason to believe that the chairs in the library are blue. Instead, she claims that 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 requires inferences within her belief system to 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 with the person in question.

If individuals could sustain 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. There are, however, 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 recognize an utterance (A) to have a particular meaning, to designate a particular phenomenon (B), we must recognize the norm that we ought to take A for B.[19] 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 must be 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

[19] 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.

me is right.'[20] The inferences a private language draws lack a 'criterion of correctness',[21] 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 characterizing the relations between a denomination and a perceived phenomenon that a private reasoners' language establishes renders said relations devoid of any meaning we could recognize 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 draw based on what we just said. However, the fact that 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 recognize – 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'[22] – an initial successful interpretation of their utterances, actions or reactions that indicate what might be their actual epistemic norms – providing 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 does not contradict our initial interpretation of their reasoning (the bridgehead), we must assume their epistemic norms to be similar to ours. For, if they were not, our interlocutors' reaction, that is, their behaviour and reasoning, would have to remain wholly unintelligible to us, as 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.'[23] 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.'[24]

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

[20] Ludwig Wittgenstein. *Philosophical Investigations*. New York: Mcmillan, 1953, p. 92.
[21] Ibid., p. 92.
[22] Martin Hollis. "The Limits of Irrationality". In: *Rationality*. Ed. by Bryan R. Wilson. Oxford: Blackwell, 1970, pp. 214-215.
[23] Martin Hollis. "Reason and Ritual". In: *Rationality*. Ed. by Bryan R. Wilson. Oxford: Blackwell, 1970, p. 232.
[24] Ibid., p. 232.

understand each other, i.e. 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[25] – 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 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 inappliccable 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 genuinely reflections of some subjects' mental set-up, they could not even hope to attract the reasoned endorsement of their interlocutors. Since the latter cannot perceive vastly deviant reasoning as intelligible to them at all, they cannot be expected to acknowledge subjectivist epistemic norms based on reasoned conviction.

The cases which are relevant to my discussion of the external attribution of reasons

[25]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.' (Hollis, "Reason and Ritual", 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.

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. Even 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 which 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 recognize to constitute an end to herself. This is because weak externalism operates on the reasonably well-grounded assumption that the person in question – if presented with the complete inferential supposedly applicable to her – can retrace, evaluate and rationally agree to any step in the external observers' reasoning.

Chapter 4

Factual commitments to agency

4.1 Introduction

Throughout the previous chapter, my concern about the compatibility between (i) the external attribution of reasons for beliefs to others and (ii) the liberal commitment to treating people as ends has been rather abstract. I argued that it does not need to be disrespectful to persons as ends to attribute to them reasons which they supposedly have for endorsing a certain belief. I emphasized that this can only be said to be true if the reason in question consists in, or can be inferred to 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 recognize 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 my initial question: is it possible to show that unreasonable people 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 recognize 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 recognizing other persons as ends to be the only adequate attitude for citizens to adopt

[1] Rawls, *Political Liberalism*, p. 49.
[2] As I have argued in section 2.3.

towards others as constituents of justification. As set out in section 2.2, I am only concerned with 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 recognize 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 recognizing others as ends if they were members of the constituency of public justification. Hence the following two chapters aim to answer the following key question: would even unreasonable citizens have reason to recognize 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 recognize 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, it 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 recognize 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 recognizing 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 recognize other persons as ends. Drawing on the conclusions of the previous section on externally attributing reasons to others, in this chapter, I explore elements in unreasonable citizens' individual belief system 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 recognizing the resulting *moral norm* to treat all people with equal moral consideration. These are separate issues. At this point, it still

[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.

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 as such not sufficient to 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. This is my main concern in chapter 5. In this chapter, however, I shall 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 human 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 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 has so far been 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 it to be a valuable attitude.

Is Bob nevertheless justified in attributing to Alice a commitment the value of charity by drawing on her charitable actions, despite her explicit refusal thereof? The reason he

can do so – and indeed Alice – 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 realize. 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 they do not 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 tried to demonstrate in section 3.4, 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. How-

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

[5] See ibid., pp. 39–41. I am discussing this argument in more detail in chapter 5.

ever, 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 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 believe X. Yet, 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 only which renders 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 they can be expected to be committed to in a liberal but pluralistic society.[9] O'Neill is thus right in claiming that

> If the elusive definitive analyses of personhood, agency, subjecthood and the like remain unavailable, it won't help to demand that agents base their views

[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 2.3 and 2.4.

> 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 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, 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

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

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) judgment, 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, with regard to activities which affect others 'commitments to others' ethical standing are taken on as soon 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, rationalize – 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 rationalize 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 latter: 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

[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.

> [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]

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 to them, 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 realizations 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 realizing or acting upon their wills. However, it is not the perception of an *attitude* of resentment as such that compels a person to draw

[20] O'Neill, *Towards justice and virtue: a constructive account of practical reasoning*, p. 103. Original emphasis.
[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.

this conclusion.[26] Rather, it is the need to reflect upon her attitude, to rationalize 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 rationalization of attitudes, O'Neill is right to emphasize that an attitude of resentment is premised upon an assumption of agency, rendering individuals 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 recognizing others' agency to the moral implications of recognizing 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 recognized 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 recognizing normative prescriptions with regard to the moral status

[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 rationalize it.

[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 rationalize 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.

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 emotions in general. In order to be resilient against this kind of objection, O'Neill's practical account of individuals' factual commitment 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. Chapter 5 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 rationalizing 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

[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 free-standing emphasizes 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.

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 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 recognize the inferences which link said individual's reasons to her endorsement of the proposal to be valid, 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 criticizing the reasons other citizens have so far deemed themselves to have for opposing these proposals.

So, 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.

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

[32] See, for instance, Gerald 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.

[33] Ibid., p. 279.

[34] Ibid., p. 279.

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 subject 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 would they impose upon unreasonable citizens' ability to sustain unreasonable views within justificatory procedures among constituents of justification? I begin to discuss 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, and the fact of reasonable pluralism.[36] In doing so, they may also be considered to reject the idea underlying these beliefs that all human persons equally deserve to be treated as ends in themselves.[37] If they were allowed to enter the constituency of public justification, unreasonable people would be required to justify such an attitude, given that they intend to draw on it in the course of public justification. Assuming that they are equipped with a moral capacity at all, 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 human 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. Such an attitude of resentment alone is not as such unreasonable. Yet, Alice is

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

[36] See Quong, "The Rights of Unreasonable Citizens", p. 315.

[37] See section 2.3.

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 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, scrutinizing her reason in terms of its intelligibility to themselves and its acceptability to her.[38] As I set out 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 recognized 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,[39] 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.[40] 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 acknowledg-

[38] See also section 3.3.

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

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

ing 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 emphasize the following: that, in rationalizing 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 inferring from a factual commitment to a person's agency to 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 of offering *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 am now considering 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 recognizing their objects as agents. Rather, she might emphasize 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 con-

stituency 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.[41]

Certainly, in our case, an unreasonable person who denies her interlocutors' agency would not interact with them because she expected herself to be able to convince them by any 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, however, cannot be interpreted but 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 in its mere locutionary function,[42] 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 recognize said reason. Moreover, as I have emphasized 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 recognize the reason in question as a valid purpose for her, and to act upon this purpose by giving her assent to the proposal the

[41] 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. 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 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.

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

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'[43] a particular set of 'consequential effects upon feelings, thoughts, or actions of the audience',[44] 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 recognizing the intentions which are embedded in the utterance of a justification. Or, as Donald Davidson holds '[w]hat we seek are intentions characterized in non-linguistic terms – ulterior purposes in uttering sentences.'[45] Hence, to 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'[46] 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 recognize 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, however, 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 either 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 a perlocutionary act would be to admit to their inherent futility. So, to engage in a justificatory dialogue with a person whose agency one does not acknowledge would 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

[43] Austin, *How to Do Things with Words*, p. 101.
[44] Ibid., p. 101.
[45] Donald Davidson. *Inquiries into Truth and Interpretation*. 2nd edn. Oxford: Clarendon Press, 2001, p. 272.
[46] Ibid., p. 277.

rationally deny the agency of any of her co-citizens she would engage 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 whose exclusion from the constituency of public justification is at all controversial.[47] 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 were 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 their convictions they intend to draw on in public justification. At this level, these constraints must only be perceived as procedural not 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.

[47] We do not need to consider any other group of unreasonable people, as I set out in section 2.2.

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.

Chapter 5

From factual to moral commitments: agency and moral equality

5.1 Introduction

The key question I shall discuss in this chapter is whether it is possible to draw an inference from a factual commitment to another individual's agency to a moral commitment to considering 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 from 'is' to 'ought'. 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 recognize 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 non-trivial 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.

5.2 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 nonlogical, leaving no room for a rational inference from factual to moral commitments.[1] 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 draw an inference 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.'[2] The emotivist non-cognitive view recognizes 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.[3] It rather 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.[4]

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 raising claims of the sort 'agency may be a reason for *you* to consider others as morally equal, but I cannot recognize any link between this fact and its moral implication.' As I have emphasized before, I do not intend to criticize the viability of this sort of ethical statement on a general, meta-ethical level.[5] 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 rationalization 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

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

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

[3] Ibid., p. 84.

[4] Ibid., p. 85. Emphasis added.

[5] 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.

are rationally deducible from their respective premises, given that rationality can be considered to constitute the shared standard of mutually intelligible communication.[6] The public political discourse among members of the constituency of public justification thus cannot tolerate them asserting non-rational, nonlogical 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.

5.3 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'.[7] 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'[8] and 'the agent acts for some end or purpose that constitutes his reason for acting'.[9]

The fact that purposiveness is one of the 'features distinctively characteristic of the whole genus of action'[10] 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

[6] See section 3.4.
[7] Gewirth, *Reason and Morality*, p. 63.
[8] Ibid., p. 27.
[9] Ibid., p. 27.
[10] Ibid., p. 27.

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,[11] 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.'[12] 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 his actions.'[13] 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 goods'[14] to consist in the voluntariness or freedom of a person's actions, as well as in their purposiveness.[15] Lacking the former, she 'would not be able to act for any purpose or good at all',[16] while the latter desire to increase the 'level of purpose-fulfilment'[17] (or well-being) is a logical correlate of valuing one's premises. 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.[18] 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

[11] This scenario is not to be mistaken for the *decision* to choose randomly, that results from the conscious realization that there *is* no best choice, that all alternatives are of equal value.
[12] Gewirth, *Reason and Morality*, p. 40.
[13] Ibid., p. 52.
[14] Ibid., p. 52.
[15] Ibid., pp. 52-53.
[16] Ibid., p. 52.
[17] Ibid., p. 52.
[18] Ibid., pp. 52-53.

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.'[19]

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 recognize the obligation others' agency imposes on her. The latter need results from the epistemic force of what Gewirth refers to as 'the criterion of relevant similarities'.[20] 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 recognize 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 universalizability:

> if some predicate P belongs to some subject S because S has the property Y (where the 'because' is that of sufficient reason or condition), then P must also belong to all other subjects S_1, S_2, ..., S_n that have Q. If one denies this implication 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.[21]

A person who recognizes herself as an agent is thus committed to treating others in the same way as she herself cannot but demand others to 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 recognizes herself as an agent. This commitment itself is not a moral one, but a mere requirement of rationality. One person's necessary realization 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

[19] 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.

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

[21] Ibid., pp. 104-105.

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 recognizing these prescriptions. Yet, upon universalization, 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 universalizability, change its (prescriptive) character only because it is being abstracted from the individual herself.

Gewirth's account has been criticized for making to hasty a transition from a person's prudential 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 universalizable as the fact that as an agent she needs to make this claim.[22] 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 *prescription* is not a universal one, and therefore not moral.'[23] Gewirth responds to this criticism by pointing out that, by acknowledging other agents' prudential requirement 'to seek the necessary 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.'[24] It is the notion of the process of universalization which leads to the assumption of a favourable attitude towards others' requirements that allows for the 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 universalization 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 criticized 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.'[25] Korsgaard does not deem an argument based on the universalization of self-interested and essentially private reasons to be

[22] 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, London: University of Chicago Press, 1984, p. 56.

[23] Ibid., p. 56.

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

[25] 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.

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 recognize as adequate with regard to their agency. Genuinely moral reasons, according to Korsgaard, need to provide an individual with obligations owed to his fellow individuals, not herself.[26] 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.'[27]

Rather, Korsgaard argues that reasons are never private, but essentially public by nature.[28] As soon as reasons are being exchanged, they cannot remain subjective, but need to abstract from the individuals' 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.[29] In other words, we can only be intelligible to ourselves in the same way as we are intelligible to others. A reason we recognize 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 whose normative force may be shared with others.'[30] 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 recognize 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 recognize 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 individ-

[26] Ibid., p. 134.
[27] Ibid., p. 134.
[28] Ibid., pp. 134-135.
[29] Ibid., pp. 136–138.
[30] Ibid., p. 136.

ual, first-personal perspective it can achieve the 'self-conscious awareness'[31] 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.[32]

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 recognizes 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 recognize the 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 emphasizes, '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.'[33] What is crucial is that as soon as a person is confronted with both her first-personal and their 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 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 universalizing one's claims. Contrary to Korsgaard's criticism of Gewirth's argument, the reasons an agent recognizes 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 universalize 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 of the concept of agency (which is

[31] Thomas Nagel. "Universality and the reflective self". In: *The sources of normativity*. Ed. by Onora O'Neill. Cambridge: Cambridge University Press, 1996, p. 203.

[32] Ibid., p. 203. Original emphasis.

[33] Korsgaard, "The origin of value and the scope of obligation", p. 143. Original emphasis.

shared) exerts upon her reasoning, rather than upon the fact that *she* happens to be an agent herself. To universalize means to subject one's reasoning to publicly, i.e. socially, recognized standards. In so far as all reasons we may consider to be meaningful must be universalizable (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 recognizes 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 claims 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 the sharedness of reasons, their being subject to the normative force of rationality, which only 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 recognizing themselves as agents and thus as subjects who cannot but demand other agent's moral concern.

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 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 chapter 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 would need to 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.

This line of reasoning allows us to attribute to unreasonable citizens a weakly externalist reason to recognize reasonableness as a valid criterion for regulating the access to the constituency of public justification. The liberal commitment to the moral equality of persons underlying reasonable attitudes[34] 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 recognized as a meaningful instance of those actions she is supposed to perform. It is for this reason that the argument I presented throughout this and the preceding 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 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.

[34] See section 2.3.

Chapter 6

Conclusion

Liberal political theory is rightfully dismissive of the claims of unreasonable people to be included in the constituency of public justification. However, this is not merely because, as an ideal theory, liberalism does not need to address itself to people who reject its fundamental values. In this thesis, I have argued that these are insufficient grounds for excluding unreasonable citizens from the constituency of public justification. To exclude unreasonable citizens because they do not recognize the values which liberals consider to be foundational to the necessity to structure their interaction on the basis of justification, is to misconceive these values, disregarding their roots in a fundamental liberal commitment to the value of human personhood as such. The latter requires justifications to be offered to all human beings.

Nevertheless, as I argued in chapter 2, liberal societies cannot but exclude unreasonable views from the constituency of justification in order to maintain stability, just as liberal theory needs to limit membership in said constituency to reasonable people in order to maintain theoretical consistency. To do so, however, without paying due respect to its parallel commitment to universal inclusiveness is to perpetuate, not to resolve the tension between these poles of liberal commitments. One way to allow for liberal theory to exclude unreasonable people from the constituency of justification, while still living up to its requirement to provide justifications to all, is to demonstrate that unreasonable people can be shown to have reason to accept their exclusion from the constituency of justification. It is such an argument that I have offered in this thesis.

By virtue of its theoretical and hypothetical nature, this argument could only engage with the reasons we might assume unreasonable citizens to have, thus requiring us to attribute reasons to them. In chapter 3, I discussed the character such externally attributed reasons would need to possess if their attribution is to be compatible with the liberal commitment to respecting persons as ends. Rejecting a strongly externalist conception of reasons on the grounds that it is does not pay due respect to said commitment, as 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 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 externally attributed inferences within a person's belief system, as people's mutual intelligibility in communicative situations is indicative of their commitment to shared epistemic norms.

Drawing on the idea of open justification in conjunction with the shared nature of epistemic norms, in chapter 4, I argued that a person's actions may serve as a source of beliefs, and thus 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, as I set out in section 4.2 and section 4.3. 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 to all other participants (see section 4.4). In seeking support for a proposal, participants in public justification may further be expected to be prepared to justify said proposal 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 furthermore yielding a commitment to the agency of those whose moral equality the argument is ultimately intended to deny.

In chapter 5, I finally assessed whether a commitment to recognizing 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 universalize the demands to be attributed moral concern which she herself is bound to demand by virtue of her agency. As such, she is required to affirm the value which is foundational to an attitude of reasonableness, thus giving her reason to deem said attitude to be adequate for the domain of public justification.

This conclusion is political, not moral in nature, as it is driven by constraints which are internal to procedures of public justification. It is hence situated on the same level as

the identified tension within liberal political theory, the challenge which it is intended to address. To show that unreasonable people have reason to endorse reasonableness (and reject unreasonable dispositions) as a requirement for meaningful participation in public justification enables the scope of liberal political theory to be broadened without surrendering necessary restrictions to the exclusiveness of the constituency of justification. Liberals may thus insist that unreasonable views must be dismissed from procedures of public justification, while still living up to their commitment to treating all persons as ends.

As a next step, one might be compelled to assess the applicability of this account of unreasonable views and their implications for procedures of public justification beyond the theoretical domain. An empirical case study analysing the views of real-world unreasonable actors might shed light on the accuracy of the assumptions made about their belief systems in this work, but is outside its immediate scope.

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.

Cherniak, Christopher. *Minimal Rationality*. Cambridge, MA: MIT Press, 1986.

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

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. *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.

— *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.

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, London: University of Chicago Press, 1984.

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, N.J.: 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.

Kelly, Erin and Lionel McPherson. “On Tolerating the Unreasonable”. In: *Journal of Political Philosophy* 9.1 (2001), pp. 38–55.

Korsgaard, Christine M. “The origin of value and the scope of obligation”. In: *The sources of normativity*. Ed. by Onora O'Neill. Cambridge: Cambridge University Press, 1996.

Larmore, Charles. “Political Liberalism”. In: *Political Theory* 18.3 (1990), pp. 229–360.

— *The Morals of Modernity*. Cambridge: Cambridge University Press, 1996.

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.

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

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*. Oxford: Oxford University Press, 1972.

— *Political Liberalism*. 2nd ed. New York: Columbia University Press, 2005.

Raz, Joseph. *Practical Reasons and Norms*. London: Hutchinson, 1975.

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

Strawson, P. F. *Freedom and Resentment and Other Essays*. London: Methuen, 1974.

Waldron, Jeremy. *Liberal rights: collected papers, 1981–1991*. Cambridge: Cambridge University Press, 1993.

Williams, Bernard. *Ethics and the Limits of Philosophy*. Cambridge, Massachusetts: Harvard University Press, 1985.

Wittgenstein, Ludwig. *Philosophical Investigations*. New York: Mcmillan, 1953.