Can Rawlsian Containment of Hateful Viewpoints Be Effective?

Abstract: While most of the literature has attempted to justify harsh and soft containment given some fundamental commitments of political liberalism, I focus on how justified forms of containment can in themselves be deemed effective. This article shows that a reading of Rawls allows for a comparison of different containment practices based on their capacity to protect the stability of liberal democracies under serious threat. And, in making it possible to compare harsh and soft containment, I evaluate immediate stability gains against citizens' judgements about their liberal democratic institutions.

Keywords: containment, counterspeech, Rawls, stability, threats, hateful viewpoints; hate speech bans.

1. Introduction

In Rawlsian political liberalism, 'containment' is a term of art (Rawls 2005: 64). When stability is seriously at risk, institutions and citizens are justified in using otherwise unjustified noncoercive and coercive measures to block the rapid proliferation of hateful viewpoints, understood as a subset of unreasonable doctrines that reject one or more democratic freedoms and advocate the idea that some groups lack the sheer capacity to become fully cooperating members of society (Badano and Nuti 2018: 153; Quong 2004: 334).[1] Harsh containment consists of coercive state policies, such as hate speech bans. Soft containment consists of expressive measures that citizens take to combat hateful viewpoints in their intersubjective discursive interactions.[2] So far, philosophical debates on harsh and soft containment have focussed on justifiability problems. Since the rapid proliferation of hateful viewpoints puts the stability of liberal democracies under serious threat, as proponents of both soft and harsh containment argue, special responses are justified, even if they entail otherwise morally objectionable violations of the fundamental right to reciprocal justification (Badano and Nuti 2018: 150; Quong 2004: 323).

Only limited attention (Ancell 2019: 422-23; Lepoutre 2017: 880-83; Lepoutre 2019: 183-85; Howard 2021: 934-36), and rarely within the perimeter of political liberalism (Ancell 2019), has been paid to how (and on what grounds) justified harsh and soft

---

[1] Rawls refers to the containment of unreasonable doctrines that reject one or more democratic freedoms. It is important to acknowledge that in Rawlsian political liberalism, to be reasonable means to accept a host of cognitive and moral commitments which, as several scholars have aptly noticed, are far from being uncontroversial. On this issue, see, for instance, Enoch (2015) and Ancell (2019). To study the issue of containment, I focus on a subset of unreasonable citizens whose description roughly overlaps with that of haters in disputes about hate speech bans and the discursive approach to combatting hate speech.

[2] I borrow the distinction between *harsh containment* and *soft containment* from Gabriele Badano and Alasia Nuti (2020). For an analogous distinction, see Aaron Ancell (2019).

containment can be deemed effective. More generally, political philosophers tend to consider effectiveness and justification as two distinct and unrelated, or only partially related, levels of analysis. In this article, I challenge this view. At least in the case of containment, effectiveness and justification are interconnected. At the philosophical level, one should study effectiveness in the light of those same normative criteria that justify a practice in the first place. Within such boundaries, one should compare the capacity of different containment measures to preserve stability without affecting their moral justification in societies marked by the fact of pluralism.

Inside and outside the scholarship on Rawlsian political liberalism, it is commonly thought that a comparison between the degree of effectiveness of responses to the proliferation of hateful viewpoints is a very difficult matter that necessitates a great deal of interdisciplinary work (Lepoutre 2019: 183; Howard 2021: 11; Gelber 2016: 3-5). This article is an attempt in constructing a theoretical framework for studying the effectiveness of different containment practices using stability as a term of comparison. A theoretical standpoint does not substitute for systematic interdisciplinary research, especially when such a theoretical standpoint remains internal to a particular philosophical tradition. Yet, at least within the scope of Rawlsian political liberalism, such an internal and theoretical perspective helps to clarify what one can legitimately expect from justified responses to the proliferation of hateful viewpoints.

If the aim of harsh and soft containment is to protect the stability of liberal democracies under serious threat (Ancell 2019; Badano and Nuti 2018; Quong 2004), *the capacity to protect stability in extreme circumstances* (the only circumstances in which containment is a fully justified response to the proliferation of hateful viewpoints), I claim, can be considered as a philosophical criterion to evaluate the effectiveness of containment practices, at least within the normative framework that is meant to justify their specific function in otherwise just societies. Against this backdrop, I also argue that in comparing harsh and soft containment, scholars should consider three variables: that is, the number of individuals who visibly contribute to the rise and dissemination of hateful viewpoints at time t_1; the number of individuals who visibly contribute to the rise and dissemination of hateful viewpoints at time t_2; and, the number of individuals who are directly or indirectly exposed to containment. To do so, it is important to explain what 'under serious threat' means within the Rawlsian framework that proponents of harsh and soft containment accept and defend. It is also important to study how different responses that operate under the conditions that make them justified practices can *in themselves* have a positive or negative effect on the present and future stability of liberal democracies. By 'can in themselves' I mean 'considered with reference only to intrinsic characteristics and general attributes'.

The article is organised as follows. In section 2, I describe the main features of containment strategies, and explain how they are justified in Rawlsian political liberalism. In section 3, I draw upon Rawls's arguments in *A Theory of Justice* (hereafter, *TJ*) to specify when the stability of a just society is under serious threat. This analysis sets the background conditions under which one should evaluate both soft and harsh containment. Moreover, considerations about the difference between threats and actual threats in Rawls's work will allow us to single out a criterion to evaluate the effectiveness of different containment strategies (section 4). Section 5 concludes by presenting two limitations of this study.

Two preliminary remarks are on point. First, in this article, I do not offer any substantive reason to accept that Rawls has the right view here. I do not even think that the argument for containment is particularly outstanding. But my article aims to entertain a constructive dialogue with Rawlsian disputes about containment. For this reason, in the following, I shall unpack Rawls's views and take what I think is the most appropriate interpretation as an instrument to study arguments for harsh and soft containment without questioning their most fundamental normative commitments or their merit compared with alternative paradigms. Second, one may contend that in the justification of actually existing practices, such as hate speech bans, it is unusual to provide an additional argument about exceptional circumstances. I agree, but, in this article, I try to understand arguments for containment in their strongest and most credible form. For this reason, I assume that it is possible to justify harsh and soft containment with arguments about exceptional circumstances.

2. Harsh Containment and Soft Containment

The existence of unreasonable doctrines 'that reject one or more democratic freedoms', Rawls writes, 'is itself a permanent fact of life, or seems so. This gives us the practical task of containing them—like war and diseases—so that they do not overturn political justice' (2005: 64, hereby *PL*).[3] In this short passage, Rawls introduces the conceptual tool of containment. He does not elaborate on it. But the analogy with war and disease is far from being crystal clear and suggests at least two different readings. On the first reading, containment includes practices and measures that prevent something like war or disease from *materialising*. On the second reading, containment includes only the practices and measures that prevent something like war or disease from *escalating*. Rawlsian scholarship tends to accept the second interpretation (Badano and Nuti 2018: 163-64; Quong 2004: 323-24). Against this backdrop, scholars have made proposals for harsh containment and proposals for soft containment.[4]

Harsh Containment

Quong thinks of containment as a series of measures against both the proliferation and the escalation of hateful viewpoints. On this view, containment of hateful viewpoints is to be understood as an end. It justifies some forms of state interference

---

[3] It is important to observe that the sources of unreason affect *both* epistemic *and* political reasonableness. For this reason, they alter how one perceives and supports just institutions. See, Ancell (2019: 416–17).

[4] Political liberals, drawing upon some ideas in the late Rawls, have also argued for conjectural efforts (showing that the belief system of citizens who hold unreasonable doctrines is closer than they realise to the fundamental commitments of liberal democracy) as ways to strengthen the support to liberal and democratic institutions (Clayton and Stevens 2004 2019; Ferrara 2014). For a critique, see Badano and Nuti (2020) and Baldwin Wong (2019). Yet, arguments for conjectural reasoning are not necessarily connected with concerns with the capacity to ensure stability. On this point, see Andrew March (2009) and Micah Schwartzman (2012). If we understand conjectural reasoning as a way to ensure the stability of liberal democratic societies in front of the rise of hateful viewpoints, it is important to explain how some citizens are able to persuade other citizens by means of those same belief systems they aim to challenge. Leaving aside this criticism, I think that conjectural reasoning under serious threats can be seen as a type of soft containment.

as appropriate means to that end. According to Quong, containment strategies should tackle the phenomena and events through which hateful viewpoints gain the citizens' support (Quong 2004: 323). In the same way, containment strategies should not directly benefit or empower targets of stigmatisation and reasonable members of a society. Actually, containment strategies can violate certain basic rights that members of the society have if such violations are the only means to address the proliferation of hateful viewpoints. On this view, containment is fundamentally different from the protection of basic individual rights to freedom (Quong 2004: 323-25). Take the case of an unreasonable group such as the Ku Klux Klan that plans to lynch a Black person in the American South. Local law enforcement officials prevent the lynching and arrest the culpable members of the group. Here we are not speaking of containment. As Quong points out, preventing a lynching and arresting Ku Klux Klan members is 'an instance of basic liberal justice at work, which implies protection of basic rights and liberties as well as the delivery of the social bases for self-respect' (Quong 2004: 323).

Quong studies two examples of what he considers genuine containment strategies: state interference in parents' right to educate their children and limitations on the right to free speech. According to Quong, such violations are justified as means of containment because, in a liberal democratic society, the right to adopt and act on comprehensive doctrines covers actions only within a domain circumscribed by the normative commitment to the ideal of persons as free and equal (Quong 2004: 331). For instance, detonating a bomb in a public space, within such a normative framework, was never meant to be one of the options a person of faith could pick when exercising her right to religious freedom. It is against this backdrop that the goal of containment justifies conferring on a state the right to interfere in private community schooling and therefore to interfere with parents' right to educational choice when private schools inculcate views that clash with the normative ideal that all members of the society are free and equal. The case for limitations on hate speech follows the same line of argument (Quong 2004: 333). If we understand the right to free speech as grounded in the normative ideal that all people are free and equal, it cannot be coherently understood as entitling members of the society to disseminate ideas that violate such a fundamental commitment (Quong 2004: 334). In such cases, it is therefore permissible for liberal democratic institutions to curtail the fundamental right to freedom of speech.

Soft Containment

Gabriele Badano and Alasia Nuti defend a conception of containment centred on citizens' discursive engagement (Badano and Nuti 2018: 153). Badano and Nuti insist that to contain the rise of illiberal views, ordinary reasonable citizens must 'press the unreasonable they know (for example, relatives, friends and colleagues) on their political views to change their mind and push them toward greater reasonableness' (Badano and Nuti 2018: 157). Against this background, it is important to imagine of forms of engagement with the citizens who support hateful viewpoints to persuade them to adopt more reasonable viewpoints, thereby protecting the stability of liberal democratic societies. For this reason, Badano and Nuti (2020) defend a 'soft containment' that consists of the direct persuasive discursive engagement of reasonable ordinary citizens with people who support hateful viewpoints.

Badano and Nuti justify containment as discursive engagement through identifying an imperfect duty of pressure (Badano and Nuti 2018: 153-56). Specifically, the rise and

success of right-wing populism creates such exceptional circumstances that it becomes permissible to deviate from standard rules of respectful and reciprocal discursive engagement with other individuals. Put differently, the rise and success of right-wing populism creates such exceptional circumstances that an action some reasonable citizens may be already inclined to perform in their everyday life (talking to unreasonable members of their society) becomes an imperfect duty for all reasonable citizens. From this perspective, reasonable citizens, who should be committed to defending liberal democratic societies that are just, are morally at fault if they systematically avoid engaging with unreasonable acquaintances in informal fora (pubs, family get-togethers, workplace gatherings, or Facebook and other social networks), where they 'have the opportunity to come across supporters of illiberal views with whom they have some connection and, therefore, about whom they have the personal knowledge that boosts persuasion' (Badano and Nuti 2018: 163).

In more practical terms, the rise of right-wing populism makes permissible the use of rhetoric as a way to persuade, and effect changes in, one's interlocutors (Badano and Nuti 2018: 159). On Badano and Nuti's view, rhetoric encompasses three devices: labelling (describing a colleague in powerful terms associated with socially unacceptable viewpoints in an attempt to trigger in the hearer a process of self-reflection about beliefs and behaviours); reasoning by analogy (drawing a parallel between the status of minorities in one's own country and discrimination against another minority in other parts of the world); and appeal to trusted authorities (quoting or referring to the opinions of widely respected leaders).

Therefore, in their encounters with fellow citizens who support right-wing populist views, reasonable citizens can use both the standard deliberative attitude and the special means of persuasion. In doing so, reasonable citizens should be pragmatic and avoid overintellectualisation, especially considering how difficult it may be to empathise in a situation characterised by deep disagreement (Badano and Nuti 2020). The goal of conversation is to reach consensus on a reasonable political conception. To achieve that goal, reasonable citizens should listen to people supporting unreasonable views and express a commitment to finding common ground (Badano and Nuti 2020: 162). Further, reasonable citizens can deploy effective rhetorical tactics by playing close attention either to the ways their addressees see publicly relevant political problems or to the discrepancies between their addressees' illiberal views and a reasonable political conception.

Despite increasing attention to the conditions for justifying containment and to the best-justified institutional and noninstitutional means, there is no specific philosophical study on the effectiveness of Rawlsian containment in contexts in which hateful viewpoints are proliferating. For both harsh containment and soft containment, we lack substantial evidence describing circumstances in which the stability of just states is under serious threat. A great deal of work has been done on how speech acts can prepare the way for mass atrocities in the context of failed or about-to-fail states (for example, Wilson 2015), but we do not know much about the actual and potential effectiveness of containment within liberal and democratic states that are facing a serious threat to their stability. Facing a serious threat, we might be justified in using extraordinary means. Yet, facing a serious threat, it is more difficult to achieve the expected results. Against this backdrop, it seems worthwhile to take a step back and look at the effectiveness of containment from a theoretical perspective.

In the end, harsh and soft containment are appealing as responses to the proliferation of hateful viewpoints not so much because they have functioned well as because, given the fundamental commitments of liberal democracies, hate speech bans and an imperfect duty to speak to fellow citizens *could* function well, as a matter of normative possibility in a context in which hateful viewpoints are proliferating. In this vein, I consider several passages from *A Theory of Justice* and *Political Liberalism* in succession to define the extreme circumstances under which the stability of a just society is susceptible to an actual threat and, therefore, the same extreme circumstances in which particular types of intersubjective interaction and state-level responses are *justified* and *meant to function well* as containment measures.

3. Threats and Actual Threats

To justify harsh containment and soft containment, it is crucial to demonstrate that there are compelling reasons both to depart from standard practices of legitimate liberal democratic states and to ask for more from some members of a society. As Quong himself admits (2004: 329), 'history is rife with examples of the abuse of political power by some to suppress the ideas of their political opponents'. As for Badano and Nuti (2018: 163-64), the case for containment strategies applies to circumstances in which 'a process has started that risks leading society towards the real threat to the stability of liberal institutions'. This is the case, for instance, when right-wing populist parties are recognised as credible players in politics or reliable interlocutors for government coalitions (Badano and Nuti 2018: 163-64). Yet, free political speech, such as the speech of right-wing populist parties, should be restricted, as Rawls himself reminds us in his comments on decisions such as *Schenk*, *Debts* and *Dennis*, only if 'this is necessary to prevent a greater and more significant loss' of other liberties (*PL*, 353). This means that 'a constitutional crisis must exist requiring the more or less temporary suspension of democratic political institutions, solely for the sake of preserving these institutions and other basic liberties' (*PL*, 353).

Quong, Badano, and Nuti posit that containment strategies are justified when hateful viewpoints have reached enough of a consensus to look like a threat to a liberal democratic society, but Rawls seems to set the bar very high for a danger to count as a serious threat. To call attention to the urgency of the problem in our societies, Badano and Nuti consider the rise in the systematic usage of racist messages during political campaigns and the electoral successes of right-wing populism in several European countries as proving that there is a demand for hateful rhetoric and, thus, that hateful viewpoints are gaining support to the point that they should be perceived as an actual threat to the stability of a just society (Badano and Nuti 2018: 146). To them, widespread usage of hate speech in politics is proof that there are citizens who support hateful viewpoints; resort to coded expressions or to biases, generics, and stereotypes in their everyday communication; and, eventually, agree with policies that 'clash with the equal provision of basic rights and opportunities across society' (Badano and Nuti 2018: 150).

The full (and exceptional) justification of containment is therefore linked to an ability to interpret actual threats in a nonpartisan way.[5] It is against this backdrop that it

[5] It is also for this reason, I think, that Badano and Nuti refer extensively to empirical research on right-wing populism (Badano and Nuti 2018: 146-51).

becomes very important to better understand how Rawls conceptualises supposed threats, actual threats, and the relationship between tolerant and intolerant citizens in a well-ordered society. In approaching this issue, we need to keep in mind that even if exposed to hateful viewpoints, most citizens satisfy the requirements of being reasonable citizens. Mere exposure to some hateful viewpoints, therefore, does not constitute a serious threat to stability. This simple observation reduces the number of scenarios in which one can argue in a nonpartisan way that the stability of a just society is under serious threat. One must in fact demonstrate that hateful viewpoints *have already gained a large consensus* among a significant number of citizens.

The closer Rawls gets to a full conceptualisation of the difference between threats and actual threats is in the second part of *TJ*. I suggest therefore that participants in current debates on the containment of hateful viewpoints may benefit from reading *TJ* and *PL* side by side. In this way, the few lines on containment in *PL* and Rawls's discussion of free political speech and the present-danger rule connect with the discourse on toleration and intolerant views that he develops in the sections 'Toleration and the Common Interest' and 'Toleration of the Intolerant' of *TJ*. One may object that the argument of these sections applies only to containment strategies that curtail individual liberties through physical action. This objection is incorrect (*TJ*, 217-19). These two sections prefigure the distinction between harsh containment ('Toleration and the Common Interest') and soft containment ('Toleration of the Intolerant'), where in 'Toleration of the Intolerant' Rawls explains what tolerant members, in their discursive interactions, can do to block intolerant viewpoints even in the context, such as the well-ordered society, where there is a strong presumption in favour of an unconditional and equal liberty of expression and thought. On the view I advocate, therefore, the containment of hateful viewpoints instantiates Rawls's ideas about intolerant views and intolerant people under less ideal conditions.

Toleration and the Common Interest

The argument of 'Toleration and the Common Interest' applies to a well-ordered society in which there is a large minority of citizens who hold and pursue dangerous beliefs. Rawls's research question is: in a context in which justice as fairness provides strong arguments for equal liberty of conscience, when is it permissible for a state to restrain liberty of conscience? Liberty of conscience, Rawls writes, is limited 'by the common interest in public order and security' (*TJ*, 212), where the government's right to maintain public order and security is an 'enabling right' (*TJ*, 213), 'a right which the government must have if it is to carry out its duty of impartially supporting the conditions necessary for everyone's pursuit of his interests and living up to his obligations as he understands them' (*TJ*, 213).

The argument combines two different argumentative lines (*argument 1* and *argument 2*). *Argument 1* demonstrates how limitations on liberty of conscience can be justifiable. *Argument 2* explains when such limitations are in fact justified.

> Premise: A just constitution exists[;] therefore all citizens have a duty to uphold it. Citizens are not released from this duty whenever others are disposed to act unjustly (*TJ*, 212).
> Argument 1

Postulate 1: the state can regulate liberty of conscience only in accordance with principles to which individuals themselves would agree in an initial situation of equality (*TJ*, 212).
Postulate 2: the maintenance of public order and security is understood as a necessary condition for everyone's achieving his ends whatever they are (*TJ*, 213).
Given *Postulate 1* and *Postulate 2*, all members of a well-ordered society recognize that the disruption of public order and security is a danger for the liberty of the all.
Therefore, liberty of conscience is limited, *everyone agrees*, by the common interest in public order and security (*TJ*, 212, emphasis mine).
Argument 2
Postulate 1: the state can regulate liberty of conscience only in accordance with principles to which individuals themselves would agree in an initial situation of equality (*TJ*, 212).
Postulate 2: the notion of a confessional state is rejected. The state can favour no particular religion and no penalties or disabilities may be attached to any religious affiliation or lack thereof (*TJ*, 212).
Given *Postulate 1* and *Postulate 2*, appeals to public order must be supported by common sense evidence (*TJ*, 215).
Therefore, liberty of conscience is to be limited only when there is *common sense evidence* that not doing so will damage public order (*TJ*, 214–15, emphasis mine).

The two arguments say that a fundamental interest in public order can justify limitations on liberty of conscience insofar as the vast majority of citizens feel that public order is under threat. Conversely, if just a small group of citizens pursue intolerant views and other members of their society do not consider public order to be threatened by such intolerant views, a member of the group of citizens pursuing intolerant views is protected from limitations on her liberty of conscience. Therefore, full protection of one's liberty of conscience is contingent upon the degree of which the dissemination of her views is successful. Within this framework, Rawls posits that the state should not intervene too early (otherwise it behaves like a confessional state) or too late (otherwise it may be unable to contain the threat to public order).

Toleration of the Intolerant

The argument of 'Toleration of the Intolerant' applies to tolerant citizens of a well-ordered society in which a large minority of citizens hold and pursue intolerant beliefs. Here Rawls's research question is: in a context in which justice as fairness provides strong arguments against interpreting a single moral truth as binding upon citizens (*TJ*, 217), when is it permissible for the tolerant not to tolerate those groups that are intolerant (*TJ*, 217)? For curtailments to be justified, there must be some considerable risk to the legitimate interest of preserving a just constitution (*TJ*, 219). Expressions such as 'curbing the equal liberty of the intolerant' (*TJ*, 219) overstate the coercive potential of nonreciprocal discursive interactions. Yet they are consistent with the whole of Rawls's normative argument on the topic. On his view, persuasion and conjecture can infringe the equal liberty of the intolerant because they are forms of discursive interactions in which the two parties do not conduct the discussion as equal citizens and, through such an unbalanced relation, the speaker questions the liberty of conscience of some members of the society (Rawls 1997: 786-87). When there is no serious threat to the just society, if the liberty of conscience of some citizens is to be

denied, we must give them reasons we might expect that they will reasonably accept (Rawls 1997: 771). With these preliminary remarks in mind, let me go over the key steps of Rawls's argument. His argument holds the same premise as 'Toleration and the Common Interest'. To this premise Rawls adds three postulates:

> *Premise*: a just constitution exists[;] therefore all citizens have a duty to uphold it. Citizens are not released from this duty whenever others are disposed to act unjustly (*TJ*, 219).
> *Postulate 1:* each person must insist upon an equal right to decide what his religious obligations are (*TJ*, 217).[6]
> *Postulate 2:* justice is infringed whenever equal liberty is denied without sufficient reason (*TJ*, 218).
> *Postulate 3:* justice does not require that men must stand idly by while others destroy the basis of their existence (*TJ*, 218).
> Given *postulate 1*, *postulate 2*, and *postulate 3*, the research question is: have the tolerant a right to curb the intolerant when *they are of no immediate danger* to the equal liberties of others *(TJ*, 218, emphasis mine)?
> As for *postulate 1*, each person, including the intolerant, must insist upon an equal right to decide what his religious obligations are.
> As for *postulate 2*, the equal liberty of intolerant can be denied with sufficient reasons.
> As for *postulate 3*, when citizens upholding intolerant views are about to destroy the basis of collective existence, tolerant obtain reasons that are sufficient to deny the equal liberty of intolerant.
> Therefore, there might be circumstances where the tolerant sects *sincerely and with reason* believe that their own security and that of the institutions of liberty are in danger. Under such circumstances, and only in this case, the tolerant should curb the intolerant (*TJ*, 220, emphasis mine).

In principle, the tolerant, according to Rawls, should not curb the liberties of the intolerant. In their interactions with intolerant members of their society, the tolerant should display qualities such as reciprocity and respect unless the intolerant sect grows so quickly as to overcome the homogenising forces deriving from the fact of living under institutions shaped by a just constitution (*TJ*, 217). Therefore, exceptional justification for the curtailment of one's liberty of expression and thought is granted when there is a widespread, sincere, and reasoned belief that without curbing the equal liberty of the intolerant, the institutions of liberty would be significantly damaged (*TJ*, 219). The threat must be at hand. It cannot be a threat remembered from earlier days. It cannot be an anticipation of the future. A threat that motivates the curtailment of

[6] I do not mean to place emphasis on religious obligations. Here I quote Rawls, who discusses toleration of the intolerant in connection with religious toleration (*TJ*, 216-18). As he writes, 'I shall discuss the matter in connection with religious toleration. With appropriate alterations the argument can be extended to these other instances' (*TJ*, 216-17).

individual liberty must be perceived here and now and tested through a balance of beliefs and mutual adjustments among different individual judgements.[7]

Taken together, 'Toleration and the Common Interest' and 'Toleration of the Intolerant' limit the circumstances under which the state and ordinary citizens can take actions that are not reciprocally justified in order to curtail the liberty of individuals who support intolerant views. As the two sections show, curtailing such fundamental liberties is a last resort measure whose exceptional justificatory status depends on the presence of a sincere, reasoned, and widespread belief that a just constitution is about to be in danger.

To return to contemporary discussions about containment strategies, the above treatment of two sections from *TJ* demonstrates that the normative ideal that all members of a society are free and equal not only motivates a presumption in favour of non-interference but also significantly restricts the range of cases in which containment strategies can be in fact justified through exceptions to the requirement of reciprocal justification. A society is under actual threat when a large majority of its members have a sincere, reasoned, and empirically informed belief that something very bad is happening to the just constitution. In other words, hateful viewpoints have already started gaining momentum and are visible in the public domain. The threat, therefore, is observable, not just projected. This conceptualisation of an actual threat restricts the range of cases in which containment strategies can be sufficiently justified.[8]

Against this backdrop, one might argue that self-effacement limits the application of containment because *ought implies can*. Citizens and liberal states ought to contain hateful viewpoints only if they can, but, according to my reading, citizens and liberal states ought to contain hateful viewpoints only when it is very difficult to fulfil the duty, such as during emergencies and constitutional crises (*PL*, 354). At that point it is very late. It is also questionable that at that stage, supporters of hateful viewpoints will be submissive to agents of containment. Bearing the costs of public disapproval and contestation may be an important element of their identities. For this reason, a person who supports hateful viewpoints might resist harsh and soft containment. When someone is committed to a worldview directed against a just constitution and she bears the consequences of such a commitment, it is strange to think that she will comply with state-led containment strategies or be amenable to various stratagems of persuasion.

---

[7] It is clear that, as Rawls himself notices (*TJ*, 219), this argument relies on the assumption that intolerant sects do not grow so rapidly that the sense of justice has no time to take hold.

[8] I think that this conceptualisation of an actual threat can also be understood as a further proof to demonstrate that interventions in parental right to educational choice are hardly justifiable as containment practices. For a defence of interventions in parental right to educational choice as containment practices, see Quong (2004: 326-29). For an argument against the absolute right of parents to inculcate hateful viewpoints and for the right of democratic states to persuade citizens through long-term educational programmes, see Corey Brettschneider (2012).

One might stop here and conclude that containment has too demanding conditions to function. I believe that scholars of Rawls should take this issue very seriously because it seems to affect the plausibility of an argument for containment. In this article, in order to avoid an all-too-easy critique, I adopt a more charitable approach to the discourse on actual threats. On such an approach, I accept that the threat must be actual and therefore corroborated by a widespread, sincere, and reasoned belief. I also conceive of threats more as suggestions that something unpleasant will happen than as situations in which something bad is happening.

Such a terminological shift helps to obtain a deflationary understanding of actual threats. Actual threats are to be understood as very plausible suggestions that something bad will happen to the just constitution if a particular course of action is not followed. In order to be very plausible, threats must be corroborated by a widespread, sincere, and reasoned belief that makes one see the realisation of a particular course of action as marking the difference between a dangerous state of affairs and another, more desirable and less dangerous, state of affairs.

4. A Theoretical Framework to Evaluate the Effectiveness of Containment Strategies

If we adopt the fundamental Rawlsian assumption that reciprocity is the ideal inspiring the proper exercise of power (*PL*, xliv), to say that an actual threat justifies containment is to say that under a specific set of circumstances, with respect to Φ-ing, a state or a citizen does not owe another person a reciprocal justification, but rather owes a less morally onerous type of justification. Such a justification must be amenable to a widespread, sincere, and reasoned belief that something bad will happen to the just constitution unless one starts or continues Φ-ing.

There is nothing very surprising about the idea that in some cases, citizens and liberal states can coerce nonliberal members of a society and, in doing so, abstain from the duty of reciprocal justification. This idea is disputable on several grounds, but it has been around for a while, especially among scholars who stress a conflictual account of politics (Sleat 2012). My point is neither about the substance of a liberal democratic society nor on the justification of its coercive capacity; rather, I want to cast lights upon justified containment practices and their capacity to affect how members judge their liberal democratic institutions and, therefore, more or less consciously support their stability.

I argue that if actual threats justify an exception to the duty of reciprocal justification, liberal democratic institutions and citizens should be aware that, in such extreme circumstances, successful and unsuccessful containment strategies may have at least one side effect. Specifically, within a context in which hateful viewpoints enjoy a great deal of support, such as that in which hateful viewpoints are proliferating quickly, containment strategies show publicly that the basic right of all subjected equal and free members of a society to reciprocal justification is alienable. There might be several good reasons, such as the fundamental commitment to self-preservation, to justify such an exception. Nevertheless, the violation, however justifiable from several perspectives, may remain visible and constitute a potential source of discontent.

It is against this backdrop that I now examine liberal democratic societies as sources of subjective judgements and experience. In this way, I can consider how one can feel

about one's society and how the experience of certain social and political interactions may affect subjective judgements and reasons for action. On this view, it is critical to see how liberal democratic societies translate their fundamental normative commitments into the everyday experience of their members. Some words of caution are in order. I do not think that exposure to containment alone causes reasonable citizens to turn unreasonable. It would also be implausible to argue that exposure to containment is always detrimental to the stability of a just society. Yet exposure to containment may make citizens draw comparisons and may affect both positively and negatively their personal involvement with the project of defending the present configuration of societal and political relations. When one notices contradictions between the way our social and political interactions *should be* and *how they look in practice*, such contradictions may affect the positive aura of liberal democracies and our day-to-day evaluation of reasons for and reasons against supporting a particular social and political configuration.

To insist on the same point: since fundamental normative commitments are not self-evident, members of a society should be able to formulate experience-based subjective and positive judgements about their society. This perspective entails that a liberal democratic society is stable not only when its members have in their doctrines reasons to accept its grounding normative principles ('for the right reasons', to borrow from Rawls), but also when members find here-and-now reasons to judge that both institutions and social interactions are displaying those normative commitments that were meant to shape rules of cooperation in the first place.[9]

We know from the literature that containment is instrumental to the stability of a liberal democratic society under extreme circumstances (Badano and Nuti 2018: 146, 147, 152-55; Quong 2004: 324).[10] But, we also know that emergencies have also justified exceptions in which norms are waived and fundamental rights are curtailed. As Gelber attests in describing the post-9/11 world, 'we live in a world within which speech with only tangential and minimal risks of contributing to later terrorist-related harms can be, and routinely is, criminally prohibited in a process that is valorised as keeping us safe' (Gelber 2016: 11).[11] Containment should therefore be seen as a temporary measure. Should circumstances change for the better or threats be revealed to be less serious, harsh containment and soft containment lose at least part of their ground. Within the limited period in which they are fully justified, however, harsh containment and soft containment not only may limit the dissemination of hateful viewpoints but may also affect how liberal democratic societies present themselves to their members and, therefore, impact on their positive/negative evaluations, at least among those who are not already committed to rejecting basic liberal and democratic principles. In

---

[9] Rawls (*PL*, 163) himself says that political institutions incorporating and displaying liberal principles tend to encourage the cooperative virtues of political life.

[10] Note that even in a just society, it is very difficult to demonstrate that once in place, justified and well-functioning containment measures will target only the relevant subset of unreasonable citizens for the right amount of time. In her work on terrorism and freedom of speech in the context of counterterrorism policy since the 9/11 attacks, Katharine Gelber documents that what were meant to be emergency restrictions have become a new normal (Gelber 2016: 2, 150).

[11] On this issue, see also Jonathan White (2015: 312-13).

reality, even in that case, exposure to containment may contribute to intensifying distrust and antagonism and therefore lead to an even more unstable social environment. Exposure to containment may also enable supporters of hateful viewpoints to assume the role of victim and therefore, at least in some cases, exacerbate disagreement—even among citizens who evaluate positively their institutions and recognise the priority of self-preservation—on the way certain ideas and discourses should be treated. As a matter of fact, the fundamental commitment to self-preservation tells citizens that something needs to be done under certain circumstances; it does not tell them what is to be done here and now, and for how long. In other words, there is—as the example of committed liberal and democratic citizens, such as progressive free speech absolutists, shows us—considerable space for disagreements on means towards a certain end. One may consider both continuing harsh and soft containment as ways to disadvantage supporters of certain views and to question fundamental liberal democratic commitments to pluralism and freedom of expression.

The last observations suggest not only that containment should be understood as a temporary response, but also that, even if it is possible to demonstrate the need for extraordinary measures here and now, it seems preferable that the number of citizens directly or indirectly exposed to containment, as addressees or as bystanders, be kept low. It is important to weigh immediate gains against the possibility of motivating negative judgements and mistrust among and beyond targets of containment. Where containment aims at limiting the rise and dissemination of hateful viewpoints here and now, the immediate result of an effective containment strategy should be a decrease in the number of people who support hateful viewpoints, both among fellow citizens who voice hateful claims and among fellow citizens who more or less consciously contribute to the normalization and internalization of hateful viewpoints.

At this point of the argument, it is important to clarify what counts as having contributed to the rise and dissemination of hateful viewpoints. Following speech-act theory, I understand the discursive relationship between public speakers and their audience as a type of conversation in which two poles can give and receive something (McGowan 2004; Fumagalli 2020). On the one side, *both* speakers with a limited reach in their everyday interactions *and* speakers who have enough credibility, authority, and reach—such as politicians, members of the executive branch, representatives of political parties and fringe groups, and public intellectuals and influencers—may use hate speech, blanket statements about the members of a group, and veiled expressions. On the other side, listeners can contribute to the proliferation of hateful viewpoints in at least two ways. First, they may provide speakers with inputs about what ought to be said in a certain context (Fumagalli 2021). Second, hearers may let explicit hateful statements and coded expressions go through (Langton 2018, Maitra 2012).

Another aspect deserves further elaboration. It is not always clear whether viewpoints or citizens are the targets of containment. Rawls refers to 'doctrines' (*PL*, 64), Quong (2004: 314) refers to 'unreasonable citizens', and Badano and Nuti (2018: 115-57; 2019: 15-16) tend to refer to viewpoints and citizens interchangeably. If viewpoints were the direct target of containment, it would be very difficult to account for the fact that the same citizen may contribute to the proliferation of hateful viewpoints on multiple platforms and in multiple venues. Yet, even after accounting for existing variations in the literature, it is sound to presume that containment targets citizens. Simply put,

citizens aim at persuading other citizens through soft containment; and states penalise citizens through harsh containment. On this view, a citizen who engages with different platforms simultaneously counts as just one target of containment.

If my remarks in this section are sound, the overall evaluation of the effectiveness of containment should take a broader perspective than that commonly taken. It should also include the number of people who are directly or indirectly exposed to a violation of the right to reciprocal justification. In taking such a broader perspective, I propose a simple way to connect direct and indirect targets of containment. Specifically, the indicator C is a tool to evaluate philosophically the capacity of containment strategies to protect the stability of liberal democracies under serious threat. In light of what I have said so far, C can be expressed as follows: $\mathrm{C} = \frac{(x-y)}{z}$, where x is the number of individuals who visibly contribute to the rise and dissemination of hateful viewpoints at time t_1, y is the number of individuals who visibly contribute to the rise and dissemination of hateful viewpoints at time t_2, and z is the number of individuals who are directly or indirectly exposed to containment. The definition of C is illustrative and conceptual. For this reason, I understand x, y, and z as three complementary points of view for the philosophical study of containment and its effectiveness when the stability of a just society is under serious threat. In adopting such a conceptual standpoint, it is important not to underestimate existing practical challenges in measuring the value of x, y, and z in large societies. For the present purpose, though, one can attribute plausible and sufficiently descriptive values to these variables. And, in light of such values, one can compare different ways to contain the proliferation of hateful viewpoints. Since x counts only the targets of containment, z will always be greater than or equal to x. As for y, the value will always be between 0 and the number of residents minus the agents of containment. The greatest possible value of C will be 1, as that value entails that z includes only the targets of containment and that y is 0. The fundamental idea is that where different strategies target roughly the same amount of people and take effectiveness as the most relevant criterion, one should always prefer containment practices that in themselves have a greater C.

To be sure, C can also have a negative value. When containment practices are total failures, the value of y is greater than the value of x. For instance, C may be negative when something bad is happening to the stability of a just society to the point that it is almost impossible to contain the success of hateful viewpoints. C may also be negative when containment targets the wrong segments of the society. Once such contextual variables are set aside, there is no containment measure that can only have a negative C. Bearing this in mind, the questions for us are the following:

> *In a context in which hateful viewpoints are proliferating,* (I) to what extent *can* harsh containment or soft containment decrease the number of citizens who play a role in the proliferation of hateful viewpoints? And in this context, (II) to what extent *can* harsh containment or soft containment limit the number of people directly or indirectly exposed to such practices?

C offers a framework to address these questions. Besides the actual measurement of its value in different scenarios, something that goes beyond the spirit of this paper, C helps us to connect three variables and to link the immediate gains of containment, as expressed in the numerator, with a measure of its costs, as expressed in the denominator. In this way, it offers a new way to investigate the effectiveness of harsh

and soft containment given the fundamental normative commitment shaping social and political interactions in liberal democracies. Before embarking on such a study, I want to clarify that what I develop in the remaining discussion is just a first illustration of how harsh and soft containment can be comparatively assessed on the basis of their intrinsic capacity to ensure stability in extreme circumstances. To do so, I shall draw upon the existing literature on hate speech bans and counterspeech.

Harsh containment

The key intuitions behind harsh containment echo expressive arguments for hate speech laws. In assessing harsh containment, we can therefore draw upon arguments for and against bans on expressive grounds.[12] Expressive arguments claim that sanctions can in themselves deter present and future hate speakers from engaging (or continuing to engage) in hate speech (Bollinger 1988; Gelber and McNamara 2015). On this view, the very presence of hate speech laws, given the potential force and reach of institutional discourses, contributes to removing hateful utterances from the public sphere (Delgado 1982: 148).

Available data do not tell us clearly the extent to which hate speech bans have helped reduce the proliferation of intolerant views (Heinze 2016; Lepoutre 2020). As Brown (2015: 246) notes, 'There is a dearth of useful evidence comparing the extent of hate speech in countries that do possess hate speech law with the extent of hate speech in countries that do not'. Moreover, even if bans do help reduce the incidence of hateful expressions, significant methodological obstacles hinder our comparisons between different contexts and times: different countries define hate speech in different ways; hate speech often goes unreported; and there are so many cultural, social, and political differences among countries that, as Maxime Lepoutre aptly puts it, 'it would remain extremely difficult to establish a causal connection between bans and reductions in hate speech' (Lepoutre 2020: 279). Public hateful utterances are often deep-seated in hearers' minds, and even bans, as David Partlett says, do little to contain the activities of 'those citizens bent on disseminating racial defamation' (Partlett 1989: 467). For instance, Gelber and McNamara report that despite the presence of hate speech laws in Australia, the 9/11 attacks triggered a wave of verbal abuse of Arabs and Muslims (Gelber and McNamara 2015: 631-64). A problem cutting across arguments for bans is that they tend to underestimate how language shapes both affective and cognitive aspects of hateful viewpoints and, as Lynne Tirrell argues, how 'people find social interactions, reading the daily news, surfing the internet, or engaging in social media to be laced with minefields of fear, hate, anxiety, and despair' (Tirrell 2018: 117). For instance, social psychologists have consistently demonstrated that implicit racist prejudices are so tightly linked to deeply entrenched habits developed through socialisation that they are very difficult to change through occasional actions (Lai and Hoffman and Nosek 2013).

Despite the lack of conclusive evidence, it remains plausible that harsh containment measures, such as bans and related legal regulations of dangerous speech, could erase speeches publicly transmitting viewpoints that contradict the fundamental normative commitments of a liberal democratic society. The fact that in many cases regulations do not remove all traces of hatred from the public sphere may depend on a host of

[12] I draw upon the excellent work of Lepoutre (2020).

contextual variables that do not necessarily relate to what those measures can do, if considered without other related situations or inhibiting factors.

To address the second part of the question, we must dig more into expressive arguments for bans. In liberal democracies, hate speech laws, like all other democratic laws, affect all residents within a certain territory. It is for this reason that advocates of hate speech laws on expressive grounds say that such laws convey against hateful viewpoints authoritative and intense public utterances that speak to the entire democratic community (Waldron 2012). Such public utterances can be powerful exactly because the condemnatory message is voiced by the state and received by the whole demos. For this same reason, opponents of hate speech bans argue that the reach and force of public utterances conveyed by hate speech laws can be counterproductive (Strossen 2018). Democratic states cannot control the number of people directly or indirectly exposed to the bans. The whole process of designing and enacting a democratic law, including procedures governing prosecution and defence, not only allows a platform for arguments that would not qualify for public debates (something that does not necessarily affect the palatability of containment measures), but also can expose the entire demos to containment actions, where some members of the demos may perceive such actions as violations of some normative commitments upon which the democratic society was built (Strossen 2018). These remarks suggest that in a context where hateful viewpoints are proliferating, harsh containment can decrease the number of citizens who play a role in the proliferation of hateful viewpoints but cannot in itself control the number of people directly or indirectly exposed to containment. If we therefore assume that hateful viewpoints are on the rise but still held by a minority, the value of C describing the intrinsic capacity of harsh containment to protect stability under extreme circumstances can be either low or very low.

The observation that everyone 'hears' the expressive dimension of potential sanctions is valid also for certain cases of state speech, such as the usage of the state's power to grant or withdraw tax-exempt and tax-deductible nonprofit status from groups that advocate hateful viewpoints. In so doing, as Brettschneider argues, the democratic state actively promotes the value of equal and free citizenship and responds to the duty to 'make clear that it is not complicit in their opposition to the ideal of free and equal citizenship' (Brettschneider 2012: 71, see also, 16, 128-29). Scholars have already argued that in liberal and democratic societies marked by the fact of pluralism and by disagreements on the meaning of free and equal citizenship, this kind of state speech can be seen as an instance of coercion (Billingham 2019). In the same way, from the standpoint of this paper, given both the expressive capacity of democratic institutions in pluralistic democracies and the reliance of certain groups on a privileged fiscal status to sustain their activities, withdrawing fiscal benefits should be read as an instance of harsh containment. If decisions on state subsidies, just like hate speech laws, have the expected level of publicity, such instances of harsh containment can have a low or very low C.

The last claim does not apply to the most intuitive forms of state speech, such as public declarations, conjectural arguments, building of public monuments to honour democratic activists, and naming of national holidays after prominent democratic activists. In such cases, liberal and democratic institutions assert their conception of free and equal citizenship without using any punishment or threat (Brettschneider

2012: 95-6). For this reason, I believe that these forms of state speech count as instances of soft containment. Yet, unlike in the case of one-on-one interactions, all citizens are meant to be exposed to democratic persuasion, and state speech cannot avoid the fact that several citizens may see conjectural and noncoercive attempts at democratic persuasion as violations of other fundamental commitments, such as pluralism, upon which democratic societies were built. Therefore, even accepting that state speech is so authoritative as to cause a significant decrease in the number of individuals who visibly contribute to the rise and dissemination of hateful viewpoints between t_1 and t_2, the potential value of C remains low because democratic persuasion is meant to speak to all citizens and therefore exposes the entire demos to containment actions.

Soft containment

Philosophical defences of soft containment in the Rawlsian tradition echo arguments for counterspeech, understood as discursive actions that citizens, and, as we have just seen, states, take in the first person to undermine offensive or dangerous communicative acts and the proliferation of hateful viewpoints (Fumagalli 2021; Howard 2021). Seen through such lenses, continuous discursive engagement with citizens who support hateful viewpoints is meant to alter the conversational context to the point that it becomes difficult for hateful viewpoints to be transmitted from one citizen to another.

While bans have an impact mainly on public speakers and mainly after the fact and therefore, as Lepoutre (2019: 180-81) has recently argued, are unlikely to succeed in countering deep-rooted attitudes, persuasion, which is a continuous process and is extended over time, can precede and follow single hate speech events and therefore affect conditions for uptake (Fumagalli 2021; Lepoutre 2019; Tirrell 2018). Yet, compared with arguments for hate speech bans, it is more difficult to evaluate whether the presence of an imperfect duty of pressure can in itself decrease the number of citizens who, both as speakers and addressees, play a role in the proliferation of hateful viewpoints and can in itself limit the number of people directly or indirectly exposed to discursive engagement.

So far, demonstrating the immediate positive effect of sustained discursive engagement with citizens displaying different kinds of unreasonableness has proven to be very difficult. There is an increasingly widespread feeling that counterspeech might be effective in the long run, but we are very far from having conclusive evidence (Fumagalli 2021; Howard 2021; Lepoutre 2019; Tirrell 2018). We know that wholesale changes in attitudes are rare: attitudes evolve slowly, develop with the accumulation of experience, and become nuanced over time (Dovidio and Kawakami and Gaertner 2000). Devine and colleagues have also demonstrated that the effect of interventions against implicit biases becomes more pronounced over time as people become more aware of their own spontaneous biases (Devine et al. 2012). There is also conflicting evidence about the best means to the end of persuasion. Because of public shaming, some speakers who do not perceive themselves as racist are quick to correct their hateful tweets, while other speakers escalate their hateful rhetoric and double down on what they said (Benesch et al. 2016). Scholars have also shown that in many cases, attempts at persuading people to change their political views through facts and evidence are severely limited by confirmation bias and defensiveness (Ancell 2019: 423). Against this backdrop, it seems safe to say that we do not know whether soft

containment has reduced the number of people who contribute to the proliferation of hateful viewpoints.

Such inconclusive empirical results make a theoretical analytical standpoint even more necessary. A theoretical perspective allows scholars to imagine what the best possible deployment of soft containment can do given its most fundamental characteristics. From the revival of speech-act theory in contemporary political philosophy we know that among their felicitous conditions, successful persuasive speech acts need at least the presence of some kind of common ground and an appropriate configuration of authority relations within the conversation (Langton 2012: 90-93). To be sure, hearers may adjust the context of an act to make it appropriate, but the adjustment is motivated by hearers' tacit acquiescence to what speakers aim to do with their speech (Ayala and Vasilyeva 2016: 265-67; Maitra 2012: 106). Unless we imagine a very malleable group of hearers, it is odd to believe that citizens who hold hateful viewpoints will be ready to accommodate and include in the common ground exactly those presuppositions questioning directly their viewpoints and therefore will let persuasive speech go through.

When we turn to authority relations within the conversation, it is important to keep in mind that speakers can come to have authority even when they lack it prior to speaking (Maitra 2012: 96). For this reason, it is plausible to think that in certain conversations, despite disagreements, unreasonable friends, relatives, and colleagues can recognise the authority of their reasonable peers. Nevertheless, the problem is exactly that a speaker comes to have authority, in that hearers refrain from challenging the speech, stay silent, conform to an assertion. In one way or another, hearers play an active role in facilitating the conversation (Maitra 2012: 105-6). Therefore, the authority that agents of soft containment require in order to have a chance to persuade citizens who support hateful viewpoints is not intrinsic to the practice of discursive engagement. It is against this backdrop that advocates of soft containment should recognise that the presence of an imperfect duty to persuade cannot in itself reduce the number of citizens who contribute to the proliferation of hateful viewpoints. Such a reduction also depends on actions performed by hearers.

The point of view of the hearers helps me to address the second part of the question. In contexts in which unreasonable worldviews are expanding quickly but remain a minority point of view, the presence of a duty of pressure does not in itself imply that the violation of the right to reciprocal justification will be visible to the whole democratic community. As an imperfect duty, the duty of pressure admits exceptions and does not require 'reasonable citizens to react to each and every relevant unreasonable comment' (Badano and Nuti 2018: 164). Moreover, unlike democratic laws, the scope of sustained discursive interactions with friends, colleagues, and relatives correlates with the number of meaningful discursive relationships an agent of soft containment can realistically have. In this way, soft containment does not in itself entail that a large number of members of the demos will be directly or indirectly exposed to what they perceive as a violation of a fundamental moral right. Actually, z, even if we assume that some third parties will be indirectly exposed to containment, cannot be much greater than x (the number of people who are the target of containment at t_1), suggesting therefore that even under these circumstances, soft containment, understood as an imperfect duty of persuasion, can in itself have a greater

C than harsh containment, understood as bans on speech expressing hateful viewpoints.

Exposure to soft containment, as an anonymous reviewer put it, may in fact push nonhateful citizens to reflect on their values and therefore strengthen their commitment to supporting a just society. This observation, I think, adds to the idea that at least from the point of view of effectiveness, justified soft containment is preferable to justified harsh containment. Not only does soft containment keep the number of bystanders indirectly exposed to soft containment low, but among those exposed to containment, it also multiplies opportunities for bystanders to deliberate.

There might also be cases in which both harsh containment and soft containment have a C of the same value.[13] According to the perspective of this article, one should consider how harsh and soft containment can in themselves guarantee a well-considered guess that targets of containment have stopped supporting hateful viewpoints. Hate speech laws can correlate with a significant reduction in the number of people who visibly contribute to the proliferation of hateful viewpoints but cannot in themselves ground a well-considered guess that all targets (speakers and addressees) of containment have given up their views.

Hate speech bans and state speech through public declarations face analogous obstacles. In both cases, it is very difficult to conclude that targets have given up their views. So when state speech and hate speech bans can be equally effective at containing threats to the stability of a just society, reasons for preferring one strategy over the other should be found elsewhere. The story is different for soft containment through persuasion. In their interpersonal interactions, citizens can reach firmer ground in concluding that targets of containment have changed their views. One-on-one interactions, even if much seems to rely upon performative signals, allow citizens to monitor their results, adapt their conversational moves, and, if necessary, continue engaging with supporters of hateful viewpoints. For this reason, soft containment can provide agents of containment with a firmer basis for claiming that targets have changed their views.

5. Conclusion

In this article, I have analysed Rawlsian literature on containing the rapid proliferation of hateful viewpoints. While most of the literature has attempted to justify harsh and soft containment given some fundamental commitments of political liberalism, I have focussed on how justified forms of containment can in themselves be deemed effective. My article has shown that a reading of *TJ* and *PL* in continuity allows for a comparison of different containment practices based on their capacity to protect the stability of liberal democracies under serious threat. And, in making it possible to compare containment practices, I have illustrated how to evaluate immediate stability gains against citizens' judgements about their liberal democratic institutions. Within this framework, it is plausible to say that even in extreme circumstances, soft containment can in itself be either as effective as harsh containment or more effective than harsh containment. Even if harsh containment can significantly decrease the number of citizens who visibly contribute to the proliferation of hateful viewpoints,

[13] I wish to thank an anonymous reviewer for pressing me on this point.

soft containment has in itself a heightened capacity to limit the number of citizens directly or indirectly exposed to a violation to the right of reciprocal justification.

I want to conclude by expressing two limitations of my study. My argument has tried to proceed internally to the premises adopted by advocates of both soft and harsh containment measures. However, there are many other considerations that can be relevant to selecting a containment strategy, especially from an interdisciplinary perspective. This article does not capture several of such considerations. I recognise that other variables, such as reduction in the authority of speakers (Wilson and Kiper 2020: 99), receptiveness of the audience (Lepoutre 2019; Fumagalli 2021), and audience capacity to deal with the negative effects of hate speech (Tirrell 2018), should go into assessing the effectiveness of containment practices. Without studying the context of speech, it remains difficult to predict how soft and harsh containment practices can in themselves empower the audience and, therefore, lead to a reduction in the number of people supporting (and advocating for) hateful viewpoints. Yet, it is plausible to hold that respectful forms of discursive engagement, which aim to make the audience ready to reject or block hateful viewpoints without necessarily contesting listeners' worldviews, seem to be more likely to reaffirm those same civic bonds that are necessary for a liberal and democratic society to be stable over time. Within this framework of ideas, this article has added to ongoing scepticism towards the effectiveness of coercive measures (Brettschneider 2012: 77-78; Lepoutre 2017; Talisse 2009: 60-61). I have argued that even under extreme circumstances, which may seem to be an ideal context for drastic actions, there are good reasons to be sceptical about the deterrence effects of justified coercive measures.

One may object that in both theory and practice, harsh and soft containment work in tandem. In practice, it is true that laws shape different stages of private interactions (Tirrell 2019). Yet, at the theoretical level, soft containment is presented as a more effective alternative than harsh containment (Badano and Nuti 2018: 153-54). My argument adds to this thesis. It shows that a study of serious-threat situations gives us reasons to continue thinking that soft containment is in itself more effective than harsh containment.

Acknowledgement: many thanks to Valeria Ottonelli and three anonymous referees for extremely helpful guidance in revising the article.

References

Ancell, Aaron. 2019. "The Fact of Unreasonable Pluralism," *Journal of the American Philosophical Association* 5 (4): 410-28.

Ayala, Saray and Nadya Vasilyeva. 2016. "Responsibility for Silence," *Journal of Social Philosophy* 47 (3): 256-72.

Badano, Gabriele and Alasia Nuti. 2018. "Under Pressure: Political Liberalism, the Rise of Unreasonableness, and the Complexity of Containment," *Journal of Political Philosophy* 26 (2): 145-68.

Badano, Gabriele and Alasia Nuti. 2020. "The Limits of Conjecture: Political liberalism, Counter-Radicalisation and Unreasonable Religious Views," *Ethnicities* 20 (2): 293–311.

Benesch, Susan et al. 2016. *Considerations for Successful Counterspeech*. Cambridge MA: Dangerous Speech Project.
Billingham, Paul. 2019. "State Speech as a Response to Hate Speech: Assessing 'Transformative Liberalism," *Ethical Theory and Moral Practice* 22: 639-55.
Bollinger, Lee Carroll. 1988. *The Tolerant Society Freedom of Speech and Extremist Speech in America*. New York: Oxford University Press.
Brettschneider, Corey. 2012. *When the State Speaks, What Should It Say? How Democracies can Protect Expression and Promote Equality*. Princeton: Princeton University Press.
Brown, Alexander. 2015. *Hate Speech Law: A Philosophical Examination*. London: Routledge.
Clayton, Matthew and David Stevens. 2004. "When god commands disobedience: Political liberalism and unreasonable religions," *Res Publica* 20 (1): 65–84.
Clayton, Matthew and David Stevens. 2019. "Further Thoughts on Talking to the Unreasonable: A Response to Wong," *Res Publica* 25: 273-81.
Delgado, Richard 1982. "Words That Wound: A Tort Action for Racial Insults, Epithets, and Name-Calling," *Harvard Civil Rights-Civil Liberties Law Review* 17: 133-81.
Devine, Patricia G. et al. 2012. "Long-term Reduction in Implicit Race Bias: A Prejudice Habit-Breaking Intervention," J*ournal of Experimental Social Psychology* 48 (6): 1267–78.
Dovidio, John, Kerry Kawakami and Samuel L. Gaertner. 2000. Reducing Contemporary Prejudice: Combating Explicit and Implicit Bias at the Individual and Intergroup Level. In *The Claremont Symposium on Applied Social Psychology 'Reducing prejudice and discrimination,* ed. Stuart Oskamp. Lawrence Erlbaum Associates Publishers: 137-63.
Enoch, David. 2015. Against Public Reason. In *Oxford Studies in Political Philosophy, Volume* 1, eds. David Sobel, Peter Vallentyne, and Steven Wall. Oxford: Oxford University Press.
Ferrara, Alessandro. 2014. *The Democratic Horizon: Hyperpluralism and the Renewal of Political Liberalism*. Cambridge: Cambridge University Press.
Fumagalli, Corrado. 2020. "Populist Appeals and Populist Conversations," *Global Justice: Theory Practice and Rhetoric* 12 (2): 72-93.
Fumagalli, Corrado. 2021. "Counterspeech and Ordinary Citizens: How? When?" *Political Theory* 49: 1021-47.
Gelber, Katharine. 2016. *Free Speech After 9/11*. Oxford: Oxford University Press.
Gelber, Katharine. and Luke McNamara. 2015. "The Effects of Civil Hate Speech Laws: Lessons from Australia," *Law and Society Review* 49: 631–64.
Heinze, Eric 2016. *Hate Speech and Democratic Citizenship*. Oxford: Oxford University Press.
Howard, Jeffrey 2021. "Terror, Hate and the Demands of Counter-Speech," *British Journal of Political Science* 51 (3): 924-39.
Lai, Calvin, Kelly Hoffman, and Brian A. Nosek. 2013. "Reducing Implicit Prejudice," *Social and Personality Psychology Compass* 7 (5): 315–330.
Langton, Rae. 2012. Beyond Belief: Pragmatics in Hate Speech and Pornography. In *Speech and Harm. Controversies over Free Speech*, ed. Ishani Maitra and Mary Kate McGowan. Oxford: Oxford University Press, 72-93.
Lepoutre, Maxime. 2017. "Hate Speech in Public Discourse: A Pessimistic Defence of Counterspeech," *Social Theory and Practice* 43 (4): 851–83.
Lepoutre, Maxime. 2019. "Can 'More Speech' Counter Ignorant Speech?" *Journal of Ethics and Social Philosophy* 16 (3): 155-91.

Lepoutre, Maxime. 2020. "Hate Speech Laws: Expressive Power Is Not the Answer," *Legal Theory* 25 (4): 272-96.
Maitra, Ishani. 2012. Subordinating Speech. In *Speech and Harm. Controversies over Free Speech,* ed. Ishani Maitra and Mary Kate McGowan. Oxford: Oxford University Press, 94-120.
March, Andrew. 2009. *Islam and Liberal Citizenship: The Search for an Overlapping Consensus*. New York: Oxford University Press.
McGowan, Mary Kate. 2004. "Conversational Exercitives: Something Else We Do with Our Words," *Linguistics and Philosophy* 27: 93-111.
Partlett, David. 1989. "From Red Lion Square to Skokie to the Fatal Shore: Radical Defamation and Freedom of Speech," *Vanderbilt Journal of Law* 22: 431-77.
Quong, Jonathan. 2004. "The Rights of Unreasonable Citizens," *Journal of Political Philosophy* 12 (3): 314-35.
Rawls, John. 1971. *A Theory of Justice*. Cambridge MA: Harvard University Press.
Rawls, John. 1997. "The Idea of Public Reason Revisited," *The University of Chicago Law Review* 64 (3): 765-807.
Rawls, John. 2005. *Political Liberalism. Expanded Edition*. New York: Columbia University Press.
Schwartzman, Micah. 2012. "The Ethics of Reasoning from Conjecture," *Journal of Moral Philosophy* 9 (4): 521–44.
Sleat, Matt. 2013. "Coercing Non-Liberal Persons: Considerations on a More Realistic Liberalism," *European Journal of Political Theory* 12 (4): 347–67.
Strossen, Nadine. 2018. *HATE. Why We Should Resist it With Free Speech, Not Censorship*. New York: Oxford University Press.
Talisse, Robert B. 2009. *Democracy and Moral Conflict*. Cambridge: Cambridge University Press.
Tirrell, Lynne. 2018. "Toxic Speech: Inoculations and Antidotes," *Southern Journal of Philosophy* 56 (S1): 116-44.
Tirrel, Lynne. 2019. "Toxic Misogyny and the Limits of Counterspeech," *Fordham Law Review* 87 (6): 2433-52.
Waldron, Jeremy. 2012. *The Harm in Hate Speech*. Cambridge MA: Harvard University Press.
White, Jonathan. 2015. "Emergency Europe," *Political Studies* 63 (2): 300-18.
Wilson, Richard Ashby. 2015. "Inciting Genocide with Words," *Michigan Journal of International Law* 36 (2): 277-320.
Wilson, Richard Ashby and Jordan Kiper. 2020. "Incitement in an Era of Populism: Updating *Brandenburg* after Charlottesville," *Journal of Law and Public Affairs* 5(2): 57-121.
Wong, Baldwin. 2019. "Conjecture and the Division of Justificatory Labour: A Comment on Clayton and Stevens," *Res Publica* 25 (1): 119-25.

Corrado Fumagalli
Assistant Professor in Political Philosophy
Università degli Studi di Genova
Dipartimento di Antichità, Filosofia, Storia, Geografia
corrado.fumagalli@unige.it
Via Balbi 30, 16126 Genova (GE) Italy