

Self-Censorship for Democrats¹

Matthew Festenstein, Department of Politics, University of York, YO10 5DD,
United Kingdom
Matthew.festenstein@york.ac.uk

Abstract

On the face of it, self-censorship is profoundly subversive of democracy, particularly in its talk-centric forms, and undermines the culture of openness and publicity on which it relies. This paper has two purposes. The first is to develop a conception of self-censorship that allows us to capture what is distinctive about the concept from a political perspective and which allows us to understand the democratic anxiety about self-censorship: if it is not obvious that biting our tongues is always wrong, we need a fuller account of the moral sensibility that finds it so troubling and this is elaborated here. The second is to develop an argument to the effect that this sensibility should not have the last, or only, word, but instead that self-censorship should be viewed as an 'ordinary vice' of democratic societies. The grounds for tolerating it rest on the democratic values that critics believe it threatens.

Keywords

Self-censorship, free speech, realism, democratic theory, power

I. Introduction

On the face of it, self-censorship is profoundly subversive of democracy. In current political discourse, self-censorship is normally a source of anxiety, and held up as a symptom of a climate of fear, of the tyranny of the majority, stifling conformism, groupthink, McCarthyism, political correctness, or some other malign genie of democratic politics (Robin 2004). When self-censorship is invoked, it is almost always to be condemned, along with the cowardice and dishonesty of the self-censor, as part of an explanation of why some challenging opinion or inconvenient truth is not more widely discussed.

So, for example, when the historian Tony Judt (cited in Pilkington 2007) described the “virtual silence” of the news media in the United States on the issues raised by Steven Walt’s and John Mearsheimer’s (2006) article on the “Israel lobby” he reached for the language of self-censorship: “[w]e know from De Tocqueville this country is driven by conformity. The law can’t make people speak out – it can only prevent people from stopping free speech. What’s happened is not censorship, but self-censorship”. As George Orwell acidly put it, “circus dogs jump when the trainer cracks his whip but the really well-trained dog is the one that turns his somersault when there is no whip”.² The uproar generated by the notorious cartoons first published by the Danish newspaper *Jyllands-Posten* in 2006 sparked subsequent widespread fulminations over press self-censorship. It will be recalled that these cartoons depicted Muhammad in bomb-shaped headgear, wielding a cutlass, and saying that paradise was running short of virgins for suicide-bombers. In their

coverage of the controversy stirred up by these cartoons many newspapers decided that it would be irresponsible to reproduce them, and in due course Yale University Press opted not to publish a book containing the pictures. To free-speech campaigners, “all this was seen as further evidence of self-censorship amid increasing fears of upsetting sensibilities of some Muslims” (The Economist 2009; and see Laegaard 2007). When the “Charia Hebdo” cartoons published by the provocative French periodical *Charlie Hebdo*, the French journalist Nicolas Demorand (2012) warned that cartoonists shouldn’t be confused with foreign office diplomats: “les exhorter à prendre en compte le contexte géopolitique comme s’ils étaient porte-parole du Quai d’Orsay, c’est mettre le doigt dans un engrenage dont le premier cran est l’autocensure et le dernier la capitulation.” The murder of the magazine’s staff in 2015 unleashed a torrent of understandable calls to resist the temptation to self-censor, from a spectrum of sources, some vocally incensed by a now notorious injunction on the part of the *Financial Times* for “common sense” (Barber 2015).

From this perspective, self-censorship undermines an important and vulnerable condition of democratic societies: even in a democracy with well-developed liberal protections from political domination, including freedoms of speech, “a central precondition for avoiding such domination is the existence of the public sphere, a space for the exercise of shared communicative freedoms” (Bohman 2010: 434). Self-censorship seems to pollute this space, constraining citizens’ ability to speak to each other, to speak truth to power

and freely to express themselves. What gives rise to it are unacceptable and degrading relationships of power or influence.

However, precisely what it is that is being condemned by the public rhetoric against self-censorship is not obvious. Free speech includes the option not to speak, if one wishes, and if the censorship really is censorship by and for oneself, more needs to be said about why this constitutes a problem. Most social and political discourse does not enjoy the license of automatic writing or the psychoanalyst's couch, and to lack the capacity to monitor and restrain the expression of beliefs and expressive attitudes is unfortunate. And we do not always condemn the exercise of this capacity: in social and political life, it is sensible to accommodate others, as a matter of prudence or respect. We bite our tongues, we do not say what we really mean or what we would say among friends; as the metaphor or analogy has it, we censor ourselves. The discursive turn in democratic theory emphasizes not only the importance of self-expression but also the value of mutual respect and mutual accommodation, which may require curtailing the expression of opinions.

This article sets out to achieve three goals. The first is to develop a conception of self-censorship that captures what is distinctive about the concept from a political point of view. Recent conceptual and taxonomic treatments of the topic by political philosophers have filled an important gap while failing to identify this, I suggest, notably a significant contribution by Philip Cook and Conrad Heilman (2013). The political conception of self-censorship, I argue, is constituted not only by non-performance of a speech

act but by non-performance underpinned by a certain kind of explanation, in terms of a problematic power or influence relationship. This account allows us to distinguish political self-censorship from other forms of expressive self-restraint, including prudential silence, ethical tact, and the self-restraint held to be inherent in a deliberative form of politics, which in many cases are not properly regarded as self-censorship at all: how we view these boundaries, I suggest, is significant for our sense of the importance of this concept (section II). If it is not obvious that biting our tongues is always wrong we need a fuller account of the moral sensibility that finds it so troubling and the grounds on which it does so. Second, I offer an account of what underpins this sensibility: I argue that democratic hostility to self-censorship responds both to the sources of self-censorship but also to the destructive systemic effects that it identifies, particularly on three key values for democracy: for this perspective, self-censorship corrodes individual autonomy, the quality of democratic debate, and the accountability of participants in democratic politics (section III). Third, I try to explain why self-censorship, in the political sense identified here, is a more ambivalent phenomenon than critics of its sources and effects allow. It is sometimes tolerable, as an aspect of respecting free speech and the agency of speakers, and as an acceptable protection against the harms of free speech: untrammelled speech can itself have anti-democratic effects, particularly in damaging the participatory standing of some citizens, and self-censorship can play a role in mitigating these (section IV). How instances of self-censorship in fact do function so as to undermine or support democratic conditions is an important question that depends on a range of empirical

issues that fall outside the scope of this article. My aim here is only to provide a more satisfactory mapping of its significance for political theory and of how it could function, and to unblock some assumptions about how it must work.

II. Political Self-Censorship

The term “self-censorship” is loose, baggy and ill-defined enough to be deployed in a wide variety of ways, encompassing a range of expressive or discursive self-restraint on the part of agents. My friend censors himself when we discuss my singing ability. A doctor censors herself when she maps out prognoses to a patient’s family. A daughter censors herself when she discusses last night’s party with her parents. A newspaper censors itself when it does not publish the name of an informant. However ethically interesting these cases are, the argument of this section is that when critics identify self-censorship in a political context, as in the instances in section I, they use the concept in a distinctive way, not only picking out an instance of expressive self-control or self-restraint but ascribing to it certain characteristic and problematic properties. To understand the concept of self-censorship in this way identifies it not with mere non-performance of a speech act but with non-performance underpinned by a critical explanation. A claim that *X* is an instance of self-censorship is a claim that: *A*, the agent responsible for *X*, withheld some speech act that *A* otherwise could have been expected to express; that this withholding is explained by a power relationship; and that *A* is to be viewed as in some sense responsible for the withholding. Withholding speech acts that one would otherwise (in some sense) express does not in

itself constitute self-censorship. Rather, we ascribe self-censorship to *A* only when we explain the non-performance by reference to a problematic relationship of power or influence that leads to *A*'s action.

A first point is that in recognizing *X* as self-censorship we acknowledge the agency of the self-censor, as someone with some meaningful control over her or his expressive choices. As generally understood, self-censorship is a deliberate decision on the part of the self-censor not to express whatever is being self-censored: "if we are to talk of self-censorship then the will of the agent should have some significant non-coerced determinative role" (Horton 2011: 98). An accusation of self-censorship (such as Orwell's or Judt's) is not characteristically a claim that someone had no choice at all but to suppress her views – that the level of threat or constraint was so severe that it is inconceivable that she speak out. Rather, the claim is that the relationships underlying the self-censorship make it *very difficult* to speak out: it requires an unusual level of courage or indifference to the risks and inducements attached to self-expression. I would like to tell my boss what I think of her but am constrained by fear or incentives – I censor myself in order to keep her sweet.

To ascribe self-censorship to *A* also expresses a disappointed expectation. Where we have no expectations (e.g., of a Catholic periodical reproducing cartoons that mock the Pope) we do not think of self-censorship as a relevant concept. By contrast, the claim that publishers or broadcasters are being self-censoring in not reproducing the Danish or Charlie Hebdo cartoons does not flow from the thought that they have been persuaded to believe that these cartoons are so blasphemous that they should not be

published. Rather, the concern is that, like the obsequious and fearful subjects of “The Emperor’s New Clothes” they are buckling in the face of pressure, and hypocritically failing to act according to the standards and norms that they would otherwise support (Taylor 1982, Zerubavel 2006).

Finally, I want to suggest, self-censorship is distinguished from other forms of expressive self-restraint through its being explained specifically by a power relationship (also clearly expressed in the fairy tale). To claim that *X* is an instance of self-censorship is to imply there is a critical explanation underpinning the agent’s behavior: although *A* has sufficient agency to be thought of as the author of her action, nevertheless power and influence is being exercised over her that the attribution of self-censorship identifies. This is compatible with a variety of conceptions of power and other underpinning cognate notions such as coercion, ideological domination, adaptive preference formation, incentives, influence, and so on. Intimidating violent threats of course are a pretty clear form of *B*’s exercising power over *A* – in the sense, that is, of threatening to make *A* worse off.³ So is an environment where the fear of those threats is present, even in the absence of the explicit threats themselves. The dominance of the source may be unwitting, distributed and opaque to the agents responsible (cf. Graham 2002: 72-5; Pettit 1997: 132, 142). Cliques may not know that they are cliques, even if they exercise considerable power over non-members. It is not a necessary condition of being a clique and exercising this power that members are aware of doing so. (Their members’ consciousness too may need to be raised.)

To understand self-censorship in this way, in terms of agency, disappointed expectations and power, allows us to distinguish it from three other related forms of expressive self-restraint. In some of these cases, self-censorship may seem like an appropriate term to apply, and in others not: individual intuitions are likely to differ. My point is only that there is a significant conceptual boundary between these different usages and self-censorship in the political sense highlighted here. The first is tactical reticence, the prudential caginess about self-disclosure that characterizes social life. Deliberate, indeed calculated, expressive withholding is part and parcel of all communication, and including political communication. Bargaining, compromise and negotiation do not rely on artless self-disclosure on the part of those with interests at stake, and are not understood to do so by participants (Herzog 2006).

As well as tactical reasons for withholding, there are also ethical reasons of tact. Consider Emma Woodhouse's humiliation of Miss Bates in Jane Austen's novel. When Emma has insulted Miss Bates (for *her* irrepressible loquacity), Mr. Knightley must speak out as he rather would not ("it is very far from pleasant to me, but I must, I will") to say that here was an evaluative attitude she should have withheld. Now the interplay of the hierarchy thrown into relief by Emma's comments and Mr. Knightley's response ("Her situation being in every way below you should secure your compassion"), tact and self-disclosure in this situation does not seem to be captured by the concept of self-censorship: it is not the idea we would use to characterize Mr. Knightley's reserve, or the tact he is pressing on Emma

(Austen 1984 (1815): 335, 339-40).⁴ He is offering a moral reminder, underpinned by a view of social order, not imposing an outlook.

In addition to these cases of prudential tactics and ethical tact, theorists of democratic deliberation also stress specifically political reasons for rules and constraints on political talk, grounded in considerations of mutual respect, recognition or reciprocity. The most well-known example of this idea is in Rawls's conception of public reason. Given conditions of reasonable social pluralism, I cannot expect fellow citizens to act on the basis of reasons and arguments grounded only on my particular comprehensive conception of the good or moral outlook; so I cannot use those reasons and arguments to legitimate the use of state coercion on behalf of some particular public policy or law. Public deliberation consists only of the exchange of public reasons, that is, those reasons that do not require for their legitimation acceptance of some particular moral, religious or philosophical outlook. So Rawlsian public reason requires expressive self-restraint, in the sense that, if citizens are to offer proposals as to qualify as reasonable, they may have to hold back from offering the reasons and arguments in public that they feel truly ground their political proposals – to refrain, as Rawls puts it, from publicly appealing to “the whole truth as they see it” (Rawls 1993: 216).

Now in each case, the examples identified may invite the claim that they are in fact instances of self-censorship. However, my suggestion is that we identify them only as instances of self-censorship in the critical sense when they display the core features of self-censorship outlined here: that is, that they are examples of expressive self-restraint, meaningfully under the

control of the agent, that we would explain in terms of the various agents' responses to the specific power relations in which they are enmeshed. In this light, the famous critique of deliberation as a "gentleman's club" see it as self-censoring in this sense: that is, as flowing from relationships of power that need to be challenged and overcome and as promoting norms of speech that disadvantage some participants, and compel marginalized groups to adhere to forms of communication that suppress their own interests, opinions and perspectives (Young 2000; Tully 2002; Laden 2001; Dryzek 2000). Similarly, we may begin to view tact as self-censorship only if we think that the motives for tact should be viewed as imposed forms of power, around which we struggle to work. When we worry about the kind of authority exercised by Mr. Knightley over Emma, in other words, it becomes relevant to think of norms of tact and civility as forms of self-censorship. Finally, in the case of bargaining and negotiation, where there is an inbuilt tactical reason to withhold beliefs and evaluative attitudes, there is still scope to define a distinct realm of self-censorship, in the terms outlined here. This arises where we would explain the motives for self-restraint not – or not only – with reference to a person's view of her own tactical advantage but by reference to a problematic power relationship which leads her to not to express what we normally would expect.

Contested intuitions about whether *X* is an instance of self-censorship can be explained, then, not only through there being different interpretations of the empirical features of *X* but also by virtue of the concept's resting on three concepts each of which plainly invites challenge and conflicting theoretical interpretation, namely, agency, the normative expectations we

have of individuals and institutions, and power. Before turning to the significance of this for democratic theory, we can now highlight the significant differences between this conceptual map of self-censorship sketched here and that produced by Cook and Heilman in their important analysis. At the centre of their account is a descriptive distinction between public and private self-censorship. Public self-censorship is the accommodation on the part of an agent (“individual or corporation”) to an external censor (“a public agent, such as a government or public authority”). Private self-censorship is a type of self-censorship “where the censor and censee are the same agent, and this agent acts as censor over itself in the absence of an externally existing public censor”. So, for instance, someone who keeps her dislike of a colleague to herself out of a sense of what constitutes appropriate behaviour at the workplace is a private self-censor: if this is a matter of abiding by a social norm this is held to be private self-censorship by proxy; if it is grounded in “a personal sense of decency”, this is private self-censorship by self-constraint (Cook and Heilman 2013: 186-7). Now this descriptive distinction is thought to deliver a normative payoff. Principles of free speech only apply when there are coercive relations between agents, and so are only thought to apply where there is public self-censorship. Since private self-censorship only involves an intrapersonal relationship, coercion and principles of free speech are inapplicable, although Cook and Heilman (2013: 191) “leave open the possibility that these types of intrapersonal conflict may require a normative principle of some other kind to regulate the interests of competing considerations”.

Now the distinction between public and private that Cook and Heilman draw here is rather slippery. An initial point, as I have suggested, is that there is a more complex conceptual landscape of expressive self-restraint and civility than the category of private self-censorship seems to allow. It seems reductive to present refraining to speak ill of a colleague on the grounds of decency as a form of self-censorship or even self-restraint: at least, we would want a richer story to make this seem like the right concept to apply. More germane to the political case, public self-censorship on Cook and Heilman's model is expressive self-restraint by an agent in response to a public authority as a censor. In terms of the analysis offered here, it makes sense to think of a public censor in this sense as one source of power that can explain self-censorship. From this perspective, the paradigmatic case of self-censorship is the pre-emption of actual censorship by arbitrary state power, to avoid the punishment stemming from falling foul of the public authority (e.g., Patterson 1984). However, it is not clear what constitutes a censoring public authority, for this model. One ambiguity is that it is not clear whether this only encompasses law that restricts self-expression and which empowers public authorities to restrict expression, or also law that allows for self-expression to be restricted in other ways (for example, rules of debate, libel laws, privacy laws, or an absence of legal protection for whistleblowers). To the extent that we may think of rules of debate, for example, as involving legitimate curtailments of untrammelled freedom of expression, we would not view compliance with them as self-censorship. Rather, abiding by these rules is constitutive of participation in debate, and the kind of self-restraint they may

demand does not flow from the exercise of power and involve the disappointed expectations that characterize self-censorship. (Of course, to the extent that it does carry with it these involvements, self-censorship may be an applicable category, as we have seen in the case of skeptics about democratic deliberation.)

Public censorship alone is meant to raise the possibility of coercion, as private self-censorship cannot do so, since an agent coerce herself, on this view. What constitutes a “public” censor, then, seems to be any agency that may have this coercive effect: in their own analysis of the Danish cartoons controversy a public authority includes not only legally constituted censors, such as Danish law, but “those parts of the Muslim community that have threatened (and/or carried out) violence, legal action and public pressure” (Cook and Heilman 2013: 185). Cook and Heilman require this source of pressure to count as public: otherwise, those publishers who self-censored in response to this are not thought to do so in a way that raises an issue for freedom of speech but only “privately” and in a way that raises a different and unspecified set of normative concerns. It is not clear what defines a source of self-censorship as public beyond its having potentially coercive effects. But if that is the case, then the independent descriptive value of the distinction between private and public here is difficult to see. What matters for defining X as an instance of self-censorship is not whether we view the source of self-censorship as “public” or “private” but whether it involves an exercise of power.

III. Sources and Effects

In the light of this analysis, we can see that one important critical approach to self-censorship focuses on the sources of self-censorship, particularly on the power relations steering the motivations of the self-censor. A focus on the source of self-censorship has characterised the “neo-Roman” republican focus on it as a symptom of unfreedom. From this republican perspective, unfreedom consists not in the presence of a constraint on an agent but in the dependence of an agent on the arbitrary will of a “master”. The republican anxieties articulated by Philip Pettit (2006: 137) famously zero in on the insecurity of subjection to arbitrary domination:

The problem with the subjects of a kindly master is that while arbitrary interference in their lives may actually be unlikely, it will not be unlikely in virtue of their social standing, only in virtue of the contingent fact of the master’s goodwill or indifference or inattention. This means that if people have to rely consciously on such contingent facts they will be constrained, at whatever cost in dignity, to keep their masters sweet by practicing self-censorship and self-ingratiation; in the absence of a suitable social standing, groveling of this kind will be their only protection.

For this republican perspective, subjection to arbitrary power is constitutive of unfreedom, considered as dependence “on the dispositions – however benign or gentle – of our princes” while self-censorship is a symptom of this condition (Laborde and Maynor 2008: 7). Quentin Skinner (2008: 93-4) stresses the way that dependence, even without overt interference, nevertheless interferes

in agents' actions, through the self-censorship they engage in once they know themselves to be dependent: exponents of republican liberty "agree that anyone who reflects on their own servitude will probably come to *feel* unfree to act or forbear from acting in certain ways. But what actually *makes* them unfree is the mere fact of their living in subjection to arbitrary power". The wish to keep the dominating party sweet is sufficient to explain self-censoring behaviour. The key source of self-censorship is not more or less explicit threats of harm or specific interferences (although these may continue to have their place) but the absence of protections against the master's will.

Where the exemplary source of self-censorship in neo-Roman republicanism is the slaveowner wielding arbitrary power, for the tradition of anxiety about self-censorship that comes down from Tocqueville, the source of self-censorship is the stifling force of the majority, which is unleashed by the egalitarianism of modern democratic societies.⁵ For this version of the worry about self-censorship, it has its source in the potential despotism of the majority. To return to Judt's invocation of Tocqueville, while communication can of course be used instrumentally to deceive, instill false beliefs, and marginalize dissident points of view, the accusation here is significantly different. In this case, what is thought to constrain the expression of non-standard opinions (truths, in Judt's view) is not deception but the stifling force of "common sense". The concern of this tradition is with the climate of opinion that allows for free discussion, not merely with those moments when state power is invoked. While not formally constraining the individual, a conformist public opinion makes the expression of dissenting views unbearably risky.

Thought is an invisible power and one almost impossible to lay hands on, which makes sport of all tyrannies. In our day the most absolute sovereigns in Europe cannot prevent certain thoughts hostile to their power from silently circulating in their states and even in their own courts. It is not like that in America; while the majority is in doubt, one talks; but when it has irrevocably pronounced, everyone is silent, and friends and enemies alike seem to make for its bandwagon.

(Tocqueville 1969: 254)

This works through the fear of ostracism and humiliation:

The master no longer says: "Think like me or you die". He does say: "You are free not to think as I do; you can keep your life and property and all; but from this day you are a stranger among us. You can keep your privileges in the township, but they will be useless to you, for if you solicit your fellow citizens' votes, they will not give them to you, and if you only ask for their esteem, they will make excuses for refusing that. You will remain among men but you will lose your rights to count as one. When you approach your fellows, they will shun you as an impure being, and even those who believe in your innocence will abandon you too, lest they in turn be shunned. Go in peace, I have given you your life, but it is a life worse than death. (Tocqueville 1969: 255-6)

In democracies, the courtier spirit identified by the republicans is not eliminated but democratized, and put in reach of greater numbers, and "the master" is relocated in public opinion. These republican and liberal accounts are not of course the only diagnoses of the sources of self-censorship.

However, drawing attention to them serves to highlight the significance of self-censorship for political theory: self-censorship in each case is viewed as a symptom of an underlying power relation that the theory identifies and dissects.

The other side of the democratic rejection of self-censorship focuses on its political effects for key democratic values. Three in particular stand out: self-censorship diminishes accountability by stifling the sincerity on which this relies, it erodes autonomy, and it dilutes the epistemic quality of democratic debate. On the first count, self-censorship seems to undermine accountability and trustworthiness. According to what Elizabeth Markovits (2006) calls the sincerity ethic for political deliberation, democratic deliberation requires sincere self-disclosure. From a Habermasian perspective on deliberation, the claim to sincere self-disclosure is an underlying presupposition of communicative action and discourse ethics, taken as a normative backdrop to democratic deliberation. Regardless of whether it is explicitly at issue – “thematized”, in Habermas’s term – the participant’s sincerity is implicitly claimed in the act of communicative participation (Habermas 1990: 136; Habermas 1996: 318-9). The failure of sincere self-disclosure expressed in self-censorship undermines the conditions of democracy. If we do not sincerely express our points of view and reasons for policies – if our expressed views aren’t genuinely ours – then we cannot be held to account for them, in the same way as we could if these reasons were genuinely ours. Further, as Jack Knight and James Johnson (2011: 139) suggest, an awareness of this can fuel various sorts of reaction formation which add to the

pollution of public discourse, from the cynical dismissal of all public discourse as inevitably corrupted to a kind of arrogant nonconformity, “breeding detachment, reinforcing contempt, and inflating self-importance” (cf. Loury 1994).

Second, self-censorship erodes autonomy: the “tyranny of prevailing feeling and opinion”, in Mill’s (1991 (1859): 9; see Urbinati 2007) famous formulation, penetrates “deeply into the details of life ... enslaving the soul itself”. The practice of self-censorship is not only a symptom of an underlying power relationship but is thought to be problematic in that it distorts our characters. Self-censorship may be accompanied by the “sour grapes” form of adaptive preference formation as the servile come to identify their own beliefs and interests with those of the master or majority (Elster 1983). For the tradition of political thinking that comes down from Tocqueville this conformism is a likely and terrifying possibility. Beyond this, however, we may worry that the disposition to self-censor can outrun any immediate occasion and become a settled feature of an individual’s character.⁶

The third damaging effect on democratic values is the impact of self-censorship on the epistemic and deliberative quality of democratic debate. From this perspective, democratic debate is enriched by a diversity of opinions and challenges (Anderson 2006; Landmore 2012). It is impoverished when points of view are ruled out not as a result of reasoned argument but as the effect of conformist responses to power relations. Arguments cannot be properly understood and assessed, and truth cannot be spoken to power, in these circumstances. The epistemic damage that flows

from what Mill (1991 (1859): 21; cf. White 2012) calls “the peculiar evil of silencing the expression of an opinion” flows just as much from self-silencing, reducing the diversity of voices and opinions that improves the epistemic quality of discussion and decision.

In sum, then, we can see self-censorship as raising for the democratic moral sensibility a range of concerns, not only as a symptom of unfreedom but also as a corrosive of the wider set of social conditions that are thought to be important constitutive elements of democracy. The political conception of self-censorship outlined seems to squeeze out room for any form of self-censorship that is compatible with democracy. From a democratic perspective, then, although we tolerate and sometimes require other forms of self-restraint (such as the discursive norms proposed by the deliberative democrat, perhaps), self-censorship, in the specific political sense delineated here, seems by contrast intolerable, flowing from objectionable power relations and corroding important democratic values.

IV. Self-Censorship for Democrats

In this section, I want to argue that this skepticism about self-censorship should not have the last word, as it is a more ambivalent phenomenon than these criticisms suggest: the democratic values which condemn self-censorship can also support it. Let us explore this claim with respect to the sources and effects of self-censorship.

The challenge that this view of self-censorship gives rise to is to find a way of both acknowledging the agency of the self-censor and viewing that

agency as significantly impaired by relations of power and influence.

According to the account of self-censorship outlined, this agency is impaired through the stacking of incentives or threats such that we should view the self-censoring agent as subject to a power relationship which means she does not act as she would in the absence of this relationship. To the extent that we view the self-censor as lacking agency, as merely being guided by duress or incentives, then self-censorship seems intolerable by democratic lights, for the reasons outlined in the previous section. Yet, as we have seen, the self-censoring agent is not viewed as entirely lacking agency. Identifying self-censorship involves ascribing to the self-censor enough agency over her action to distinguish it from merely constrained or compelled action. If we lay stress on the agency of the self-censor, then self-censorship seems tolerable as a form of free speech. The democratic concern with agency and autonomy that leads the critic of self-censorship to condemn it also provides a reason to accept it. Since self-censoring agents are viewed as expressive agents, as outlined in section II, the reasons we have for respecting free speech generally apply in this case. Just as we tolerate other acts of free expression of whose content we disapprove, we should tolerate self-censorship (cf. Horton 2011: 101; Sedler 2012).

Understood in this way, we may regard an act of self-censorship as contemptible in the sense that we find the specific content of what is articulated contemptible (dishonest, for example) but no more so than other contemptible uses of free speech. Indeed, Orwellian venom seems to presuppose the responsibility and blameworthiness of those at whom it is

directed. The more we ascribe responsibility for self-censorship to the agent, the more it is tolerable, as an exercise of free speech. Conversely, the less we view the agent herself as responsible and the more we ascribe responsibility to the background power relationships, other agents, etc., the less applicable the concept of self-censorship, as a choice on the part of an agent, seems to be.

We can also see this ambivalence when we consider the effects of self-censorship. The democratic response to the democratic critique of self-censorship here starts from the idea that restrictions on speech may be in some cases be instrumental in removing obstacles to expression or (an important theme in the recent literature) obstacles to reception of speech.⁷ In real discursive conditions free speech can be a medium for bullying, public contempt, insinuation, and humiliation. The experience of this threatens the autonomy and capacity for participation of those on the receiving end of this speech, especially if they are relatively powerless and lack the discursive and other resources to respond, and can pollute the quality of democratic discussion. The idea that some citizens are excluded from effective participation by stereotypes, cultural norms, and lack of discursive and material resources is a familiar one. So is the idea that untrammelled free speech on the part of a powerful group can serve as a medium to enforce this exclusion.

Self-censorship can play a role in countering this: the racist member of a dominant group who keeps his opinions to himself out of fear of the social and professional consequences that flow from unabashed self-expression

helps to maintain an atmosphere in which all can participate freely and equally. We can note this while accepting that this is an instance of self-censorship, in the specific sense outlined in section II. However, the claim here is that this is tolerable when it promotes parity of participation, in circumstances when no other means of countering discursive inequalities is at hand. It is better that informal censorship plays a role than that nothing does – better, that is, for ensuring the equal participation of those who would otherwise be disadvantaged by a public sphere with a more relaxed attitude toward racist self-expression.

From this perspective, self-censorship is an “ordinary vice” of democratic societies, which can serve to protect, as well as diminish, the quality of democratic participation and decision-making. The depth of disagreement in beliefs and interests, together with mutual wariness in democratic societies, means that these societies “cannot afford public sincerity”, as Judith Shklar (1984: 78) puts it in her well-known discussion of hypocrisy:

Honesties that humiliate and a stiff-necked refusal to compromise would ruin democratic civility in a political society in which people have many serious differences of belief and interest. Our sense of public ends is so wavering and elusive because we do not even see the same social scene before us. We do not agree on the facts or figures of social life, and we heartily dislike one another’s religious, sexual, intellectual and political commitments – not to mention one another’s ethnic, racial and class character.⁸

The “democracy of everyday life ... does not arise from sincerity. It is based on the pretense that we must speak to each other as if social standings were a matter of indifference in our views of one another” (Shklar 1984: 77). A form of public life without self-censorship of this sort would not be one that allows for more honesty or truthfulness but one that opens up a more unfettered expression of mutual loathing, or of the loathing of the strong for the weak.

This claim about the potential of self-censorship is fairly minimal. It does not require us to believe that self-censorship always has beneficial effects or even that other modes of regulating expression may not be superior, where they are available. It should be distinguished from two more problematic ways of addressing the democratic critique, which can usefully be discussed in a bit more detail. The first dissolves the worry about the effects of self-censorship by supporting what could be called the no-sincerity in politics thesis. Let us consider again the claim that public discussion should be governed by a norm of sincerity which self-censorship undermines. On the face of it, this claim about sincerity seems too strong, however, and difficult to reconcile with the acknowledgement that trusting in others’ sincerity, like other forms of trust, is inherently difficult even in democratic political arenas (cf. Warren 1999; Festenstein 2009). A therapeutic response is to dissolve the worry about self-censorship by the rejecting the very idea that sincerity, truthfulness and cognate concerns have a place in democratic communication. A concern about self-censorship is a concern that a speaker is not truly saying what is on her mind or expressing her genuinely held evaluative attitude. Perhaps, however, this is too much to demand. We may

view politics as a realm of rhetoric, opinion, and appearance, where what matters is the publicly expressed face and views of the speaker, not what we imagine her genuine beliefs and attitudes are. On a very ludic interpretation, worrying about self-censorship in politics is to make a category mistake like worrying about sincerity of poker players. Political speech is a realm of rhetorical speech, play and dissimulation. The claim to be removing the mask and speaking the whole truth, or from the heart, is itself only another move in the game. As Martin Jay suggests in his interesting resuscitation of this line of argument, this is parabasis (Jay 1999, 2010). Near the end of a classical Greek comedy, the chorus, unmasked, steps forward to address the audience directly in a speech that contains the author's views on some topical matter. Of course this too is part of the play. From this perspective, the sincerity ethic for political deliberation sets a standard of access to beliefs and attitudes which is impossible to meet but also, thankfully, irrelevant to political deliberation.

The standard is impossible to meet since sincerity, in Hannah Arendt's words, is "truly unknowable ... A fellow citizen's soul remains opaque and unreadable to us; part of the beauty of democracy is that, unlike the totalitarian state, democracy gives us the freedom to have a private life" (Markovits 2006: 267; Arendt 1977: 96-8). As this way of putting the point suggests, this unknowability claim has an epistemic and a normative strand. The epistemic strand of the claim is to the effect that identifying authentic motivations and beliefs through speech is a futile task; they remain hidden to the agent as well as to those who try to interpret her actions. The "human

heart” of political agents, is “a place of darkness which, with certainty, no human eye can penetrate” (Arendt 1977: 96). The normative strand is the claim that all that matters is the public person, and that attempting to dig deeper in order to unearth others’ real views and attitudes is a violation of the scope for privacy that democracy allows. The standard is irrelevant, because what matters is only the factual truth or normative appropriateness of public speech, not the sincerity of the speakers: if someone lies, then we focus on the content of the lie, the consistency with previous utterances, and so on, but not on whether or not what she has said expresses what she truly believes. And this focus provides enough for democratic politics. In particular, it provides an adequate basis for accountability and trustworthiness.

Now this is a complex and challenging set of claims. There is no space fully to address them, but on the face of it there seems to be a lot in the thought that the inner views and attitudes of political actors may indeed be irrelevant to how we evaluate their actions and speech, and that they are also in many cases inaccessible, including to the agents themselves (cf. Dunn 2000). However, this epistemic skepticism may not in itself eliminate our interest in sincerity. There seems to be a genuine residual issue, for example, of whether a politician is attempting to censor her own racism which is nevertheless “leaking” into the public realm, whether she is subtly projecting a coded appeal to racist voters, whether this is a matter of over-interpretation, and so on. Further, and more importantly here, epistemic skepticism about sincerity does not dissolve the democratic worries about self-censorship. Rejecting the idea that the conditions of democracy must include the publicly

accessible sincerity of participants does not address the concerns about power and its effects that are core to the democratic anxiety about self-censorship. This anxiety does not spring from the thought that self-censorship perverts the expression of *A's sincerely held* points of view but only that it blocks *A's* point of view that otherwise would have been expressed in the absence of the relationship with *B*. This point of view itself may be held quite insincerely. *A* may gleefully spread a rumor about a colleague, which he does not sincerely believe at all, among co-workers – but be intimidated from gossiping in this way in front of someone on whom he wants to make a good impression. Similarly, a political candidate may offer one view of the relationship between wealth creation, entrepreneurship and taxes to a closed audience of wealthy donors, and another on national television – but we do not need to assume that the first audience is necessarily receiving this candidate's sincere opinions to view her as self-censoring in the latter case, only that her relationship to the wider audience in the second case is an important part of the explanation of the difference between her utterances in these two contexts.

The other position which needs to be distinguished from the minimal case made here is argument ingeniously developed by Jon Elster (1986, 1997) and others on behalf of the so-called civilizing force of hypocrisy in democratic deliberation (cf. Sunstein 1993; Dryzek 2000; Williams 2000). This takes self-censorship to be a necessary part of a normatively reliable mechanism for delivering one important goal for deliberative democrats, namely, impartial consensus. While an honestly expressed motive for a

proposal may be “this is good for my group”, the requirement to set out arguments for policies in the public forum compels speakers to frame their proposals in such a way as to appeal to the interests of all. As Elster (1986: 111) puts it, “publicity does not eliminate base motives, but forces or induces speakers to hide them”. In particular, “[t]he presence of a public makes it especially hard to appear motivated merely by self-interest. Even if one’s fellow assembly members would not be shocked, the audience would be. In general, this civilizing force of hypocrisy is a desirable effect of publicity”. Now this is a mechanism of self-censorship, in the sense we have set out here: the power of the majority “forces or induces” me to frame my utterances in terms that won’t trigger their rejection or expose me to shame or ridicule. Over time, it is argued, this strategic adoption of impartial principles for political action becomes a genuinely held commitment, through a psychological mechanism of “dissonance reduction”, according to which it is difficult for me to live with a set of public justifications that diverge from my real motives. So, while the sources of this self-censorship and its consequences lie in compulsion or inducement, the workings of this mechanism result in the sincere adoption of an outlook that enhances the quality of democratic debate.

Both the stability and normative attractiveness of the purported mechanism outlined by this line of thought are problematic. First, the adaptation of preferences forced or induced by this process may only be superficial, lasting only as long as the perceived sanctions of stepping out of line are thought to apply (Johnson 1997; Knight and Johnson 2011: 136-144). When these conditions change, what appeared to be a thorough identification

with social interests may quickly alter. As Uriah Heep puts it, once he has risen in the world, “I am very umble to the present moment, Master Copperfield, but I’ve got a little power” (Dickens cited Shklar 1984: 56). With a little power, the mask of humility drops, and the loyalty that seemed constitutive of his identity (at least to the gullible, and perhaps to Uriah Heep himself, if not to David and the reader) evaporates as pertinent social norms cease to exercise any sanctioning force. Second, identification with impartial standards that is constructed in this way seems to be undermined by the conditions of its own production. If these standards are arrived at through force and inducement, then this is how they may appear to those who end up adhering to them. The claims made on behalf of this mechanism assume that hypocrisy in this guise has civilizing effects, when it may breed disillusionment, cynicism or contempt about the language of public discourse. To the extent that thinking in terms of a wider interest is viewed as forced on us, we may reduce dissonance by growing more cynical about the general interest, rather than internalizing it. The general interest will seem only to be the expression of the power of the majority.

In addition to being unstable, this mechanism seems normatively dubious. The compulsion underlying the transformation of preferences is at odds with the commitment to the autonomy of the agents involved. The workings of this mechanism rests on psychological processes that are themselves shielded from critical scrutiny, occurring behind the backs of the agents affected, as it were. This undermines the autonomy of preference formation, and recalls the democratic worries about conformism.⁹ The form of

impartiality delivered is not valuable from the perspective of this conception of democracy. It is, then, overreaching to think of self-censorship as a necessary part of a normatively reliable instrument for delivering a deliberatively valuable impartiality.

Unlike the civilizing force of hypocrisy thesis, the minimal claim I am making on behalf of self-censorship addresses only “speech” or publication. There is no assumption that there is a benign feedback mechanism that goes to work on the underlying attitudes of participants. Nevertheless we may think that the minimal claim falls foul of the objection to the mechanism underlying the civilizing force of hypocrisy argument. First, it is natural for a supporter of a “talk-centric” conception of democracy to hold that the “only remedy for false or invalid arguments is criticism [...] The only cure for false, manipulative or inappropriate talk is more talk that exposes or corrects it, whether as a string of reasons, a mode of recognition, a way of making points, or a narrative” (Young 2000: 79). Accepting the instrumental value of self-censorship closes off the possibility of publicly ventilating and correcting invalid and destructive views. A second response points to the malign side-effects of tolerating self-censorship. In such an environment, politics can then become a matter of (not always subtle) signaling and “dog whistles”, allowing participants to shelter behind ambiguity and the familiar claim about the construction aggrieved parties are putting on legitimate expressions of opinion. Awareness of self-censorship may lead to a self-righteous refusal to respect public standards and a desire to probe the limits of tolerance (Knight and Johnson 2011: 140). However these responses are subject to the same difficulty that afflicts the

civilizing force of hypocrisy argument, namely, they help themselves to particular mechanistic assumptions that may not manifest themselves in practice. It is not necessarily the case that the only remedy for bad talk is more talk of higher quality, even if the latter is available. And while self-censorship may provoke cynicism or self-righteousness it may not, any more than it automatically guides an agent to a more socially minded point of view.

V.

The goal of this article has been to set out a concept of self-censorship that captures the specific political concerns of those who use it, and to articulate the concerns and issues that it raises for democratic theory. These, I have argued, are less clear-cut than the democratic criticisms of the sources and effects of self-censorship suggest. Our relation to self-censorship is (or should be) ambivalent, since the democratic values that underpin the criticism of self-censorship also suggest reasons it should be tolerated in some circumstances.

This is a qualified claim. This article does not offer an endorsement of self-censorship, or an injunction to it, and is not intended to license bullying by gentlemen's (or anyone else's) clubs, or to promote a cowed and secretive media. Nor does this remove the taint from self-censorship: its sources in power and influence mean precisely that we tolerate rather than endorse or vindicate it. Rather, the point is that we need to be able to distinguish tolerable and intolerable forms of self-censorship, while recalling that an action's being self-censorship does not in itself establish its intolerability. The grounds for

this tolerance, to be identified token by token, are that it can help to sustain the same democratic values that critics invoke for rejecting it as a type. Equally, identifying an instance of self-censorship is not in itself an evaluative conversation-stopper: its critics cannot help themselves to the assumption that it must play this role – itself sometimes a useful rhetorical blanket over careful scrutiny of the sources and effects of speech – but need to make good on their assessments in each case.

References

- Allen B (1991) The Spiral of Silence and Institutional Design: Tocqueville's Analysis of Public Opinion and Democracy. *Polity* 24: 243-67.
- Anderson E (2006) "The Epistemology of Democracy". *Episteme* 3 (1-2): 8-22.
- Arendt H (1977) *On Revolution*. New York: Penguin.
- Austen J (1984 [1815]) *Emma*. Oxford: World's Classics.
- Barber, T. (2015) The Gunmen in Paris Attacked More than a Muslim-Baiting Magazine. *Financial Times*, 27 January. Available at: <http://www.ft.com/cms/s/0/9f90f482-9672-11e4-a40b-00144feabdc0.html#axzz3Yab2wM6C> (accessed 27 April 2015).
- Bohman J (2010) Democratising the Global Order: From Communicative Freedom to Communicative Power. *Review of International Studies*, 36: 431-47.
- Colburn B (2011) Autonomy and Adaptive Preferences. *Utilitas*, 23: 52-71.

Cook P and Heilman C (2013) "Two Types of Self-Censorship: Public and Private". *Political Studies* 61: 178-96.

Davidson J (2004) *Hypocrisy and the Politics of Politeness: Manners and Morals from Locke to Austen*. Cambridge: Cambridge University Press.

Demorand, N. (2012) Libre: Editorial. *Libération*, 19 September, Available at: http://www.liberation.fr/societe/2012/09/19/libre_847506

Dryzek J (2000) *Deliberative Democracy and Beyond*. Cambridge: Cambridge University Press.

Dunn J (2000) *The Cunning of Unreason: Making Sense of Politics*. London: HarperCollins.

The Economist (2009) Self-Censorship in the West: Speech Impediments. Available at: <http://www.economist.com/node/13130069> (accessed 27 April 2015).

Elster J (1983) *Sour Grapes*. Cambridge: Cambridge University Press.

Elster J (1986) The Market and the Forum. In J. Elster and A. Hylland (eds) *The Foundations of Social Choice Theory*. Cambridge: Cambridge University Press, pp. 103-132.

Elster J (1997) Deliberation and Constitution-Making. In J Elster (ed) *Deliberative Democracy*. Cambridge: Cambridge University Press, pp. 97-122.

Elster J (2009) *Alexis de Tocqueville, the First Social Scientist*. Cambridge: Cambridge University Press.

- Festenstein M (2009) Truth and Trust in Democratic Epistemology. In R. Geenens and R. Tinnevelt (eds) *Does Truth Matter? Democracy and Public Space*. New York: Springer, pp. 69-80.
- Foerstel H (1998) *Banned in the Media: A Reference Guide to Censorship in Press, Motion Pictures, Broadcasting and the Internet*. Westport: Greenwood Press.
- Graham K (2002) *Practical Reasoning in a Social World: How We Act Together*. Cambridge: Cambridge University Press
- Grant R (2012) *Strings Attached*. Princeton: Princeton University Press.
- Habermas J (1990) *Moral Consciousness and Communicative Action* tr C Lenhardt and SW Nicholson. Cambridge, MA: MIT.
- Habermas J (1996) *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy* tr W Rehg. Cambridge, MA: MIT.
- Herzog D (2006) *Cunning*. Princeton: Princeton University Press
- Horton J (2011) Self-Censorship. *Res Publica* 17: 91-106.
- Jay M (1999) Mendacious Flowers. *London Review of Books* 21 (15): 16-7.
- Jay M (2010) *The Virtues of Mendacity: On Lying in Politics*. Charlottesville and London: University of Virginia Press.
- Kingwell M (1995) *A Civil Tongue: Justice, Dialogue and the Politics of Pluralism*. University Park, Pa: Pennsylvania University Press.
- Knight J and Johnson J (2011) *The Priority of Democracy: Political Consequences of Pragmatism*. Princeton: Princeton University Press.

- Laborde C and Maynor J (2008) The Republican Contribution to Contemporary Political Theory. In C. Laborde and J. Maynor (eds) *Republicanism and Political Theory*. Oxford: Blackwell, pp. 1-21.
- Laden AB (2001) *Reasonably Radical: Deliberative Liberalism and the Politics of Identity*. Ithaca: Cornell University Press.
- Laegaard S (2007) The Cartoons Controversy: Offence, Identity, Oppression?. *Political Studies* 55: 481-98.
- Landmore H (2013) *Democratic Reason: Politics, Collective Intelligence and the Rule of the Many*. Princeton: Princeton University Press.
- Langton R (2009) *Sexual Solipsism: Philosophical Essays on Pornography and Objectification*. Oxford: Oxford University Press.
- Loury G (1994). Self-Censorship in Public Discourse: A Theory of 'Political Correctness' and Related Phenomena. *Rationality and Society* 6 (4): 428-61.
- Maitra I and McGowan MK (2012) (eds) *Speech and Harm: Controversies over Free Speech*. Oxford: Oxford University Press.
- Marczewski P (2010) Ancient Threats to Modern Liberties: Tocqueville and the Contemporary Debate Between Republicans and Liberals. *Ethical Perspectives* 17: 449-74.
- Markovits E (2006) The Trouble with Being Earnest: Deliberative Democracy and the Sincerity Norm. *The Journal of Political Philosophy* 14 (3): 249-269.
- Mearsheimer J and Walt S (2006) The Israel Lobby. *London Review of Books* 28 (6): 3-12.
- Mill JS (1991 (1859)). *On Liberty*. In J. Gray (ed.) *JS Mill: On Liberty and Other Essays*. Oxford: Oxford University Press, pp. 1-128.

Mill JS (1991 (1861)). *Considerations on Representative Government*. In J. Gray (ed.) *JS Mill: On Liberty and Other Essays*. Oxford: Oxford University Press, pp. 203-467.

Newey G (2007) (ed.) *Freedom of Expression: Counting the Costs*. Cambridge: Scholars Press.

Patterson, A (1984) *Censorship and Interpretation: The Conditions of Writing and Reading in Early Modern England*. Madison: University of Wisconsin Press.

Pettit P (1997) *Republicanism: A Theory of Freedom and Government*. Oxford: Oxford University Press.

Pettit P (2006) Freedom in the Market. *Politics Philosophy Economics* 5 (2): 131-49.

Phillips A (1995) *The Politics of Presence*. Oxford: Oxford University Press.

Pilkington, E (2007) Articles of Faith. *The Guardian* September 15. Available at: <http://www.theguardian.com/world/2007/sep/15/usa.israel> (accessed 27 April 2015).

Rawls J (1993) *Political Liberalism* New York: Columbia University Press.

Robin C (2004) *Fear* New York: Oxford University Press.

Runciman D (2008) Political Hypocrisy: The Mask of Power from Hobbes to Orwell. Princeton: Princeton University Press.

Sedler R (2012), Self-Censorship and the First Amendment. *Notre Dame Journal of Law, Ethics and Public Policy* 25 (1): 13-45.

Shklar J (1984) *Ordinary Vices*. Cambridge: Harvard University Press.

Skinner Q (2008) Freedom as Absence of Arbitrary Power. In C. Laborde and J. Maynor (eds) *Republicanism and Political Theory*. Oxford: Blackwell, pp. 83-101.

Sunstein C (1993) Democracy and Shifting Preferences. In D. Copp, J. Hampton and J. Roemer (eds) *The Idea of Democracy*. Cambridge: Cambridge University Press, pp. 196-230.

Sunstein C and Thaler R (2008) *Nudge: Improving Decisions About Health, Wealth and Happiness*. New Haven: Yale University Press.

Taylor DG (1982) Pluralistic Ignorance and the Spiral of Silence. *Public Opinion Quarterly* 46: 311-35.

Tocqueville A (1969) *Democracy in America* ed. J. P Mayer, tr. G. Lawrence. New York: Anchor Books.

Tully J (2002) The Unfreedom of the Moderns in Comparison with their Ideals of Constitutional Democracy. *Modern Law Review* 65: 204-228.

Urbinati N (2007). The Many Heads of the Hydra: Mill on Despotism. In N. Urbinati and A. Zakaras (eds) *J. S. Mill's Political Thought: A Bicentennial Assessment*. Cambridge: Cambridge University Press, pp. 66-97.

Waldron J (2012) *The Harm in Hate Speech*. Cambridge: Harvard University Press.

Warren M (1999) Democratic Theory and Trust. In M. Warren (ed) *Democracy and Trust*, Cambridge: Cambridge University Press, pp. 310-45.

White RG (2012) Self-Censorship and the Constriction of Thought and Discussion under Modern Communication Technologies. *Notre Dame Journal of Law, Ethics and Public Policy* 25 (1): 123-142.

- Williams M (2000) The Uneasy Relation of Group Representation and Deliberative Democracy. In W. Kymlicka and W. Norman (eds) *Citizenship in Diverse Societies*. Cambridge: Cambridge University Press, pp. 125-54.
- Wingo AH (2003) *Veil Politics in Liberal Democratic States*. Cambridge: Cambridge University Press.
- Young IM (1990) *Justice and the Politics of Difference*. Princeton: Princeton University Press.
- Young IM (2000) *Inclusion and Democracy*. Oxford: Oxford University Press.
- Zerubavel E (2006) *The Elephant in the Room: Silence and Denial in Everyday Life*. Oxford: Oxford University Press.

¹ This article's origins lie in a talk given at a conference on self-censorship organized by Jonathan Parkin at the University of York in 2007, under the auspices of the Morrell Trust. I am particularly grateful to Dr Parkin, and, for comments at the time and in the many years since, to Phil Cook, Alan Haworth, John Horton, Sue Mendus, Mihaela Mihai, and Tim Stanton. This journal's referees were also immensely helpful.

² George Orwell, cited (in the context of contemporary political theorists' anxieties about self-censorship) in Wingo 2003: 103. See Foerstel (1998) for a wider study of media self-censorship in the United States.

³ On incentives as power, see Grant (2012).

⁴ On this theme, see also Kingwell 1995; Davidson 2004. Different examples of ethical tact are discussed under the heading “self-censorship good” in Sedler (2012). These include, for instance, a newspaper’s self-imposed judgement not to publish the name of a victim. This kind of case falls outside of the scope of self-censorship in the sense identified here.

⁵ From the vast secondary literature, particularly relevant is Allen 1991; Robin 2004; Elster 2009; Marczewski 2010.

⁶ This article is not focused on institutional responses to the fear of conformism, but it is worth noting that such different critics of self-censorship as the imposition of majority values as John Stuart Mill (1991 (1861)) and Iris Marion Young (1990, 2000; cf. Phillips 1995) converge on mechanisms of group representation in order to provide a protected sphere for vulnerable discursive agents – although they have different views of who those agents are and the conception of vulnerability that triggers group representation.

⁷ See, for example, Newey 2007; Langton 2009; Waldron 2012; Maitra and McGowan 2012.

⁸ For a development of Shklar’s treatment of hypocrisy, see Runciman 2008; Jay 2010.

⁹ This is probably not a problem for Sunstein, for whom it can serve another item in the arsenal of libertarian paternalism: Sunstein and Thaler 2008. On the relationship of autonomy and Elster’s conception of adaptive preferences, see Colburn 2011.