Received: 18 July 2022 | Revised: 7 February 2023 | Accepted: 16 February 2023

DOI: 10.1111/josp.12513

ORIGINAL ARTICLE

JOURNAL OF SOCIAL PHILOSOPHY

Stability and disruptive speech

Carl Fox

Inter-Disciplinary Ethics Applied (IDEA) Centre, University of Leeds, Leeds, UK

Correspondence
Carl Fox, Inter-Disciplinary Ethics Applied (IDEA) Centre, University of Leeds, 17 Blenheim Terrace, Leeds, LS2 9JT, UK.
Email: ca.fox@leeds.ac.uk

KEYWORDS: disruptive speech, legitimacy, public sphere, stability

1 | INTRODUCTION

One of the big political challenges we face is deciding what to do about the explosion of disruptive speech. By disruptive speech, I mean speech that challenges or subverts widespread existing social and political norms. There are many kinds of disruptive speech, and not all of them are bad. In fact, as we shall see, some of them are indispensable to a healthy public sphere. However, fake news, where false or misleading stories are smuggled past our epistemic defenses under the cover of journalistic conventions, is one prominent example of what I shall call bad disruptive speech, and we can point to many others such as bald-faced lies, outlandish hyperbole, and hate speech.

We are learning just how corrosive bad disruptive speech can be. Increasing numbers of people appear to be turning away from core democratic principles. Foa and Mounk (2017, p. 6–7) cite evidence showing a precipitous drop in the numbers of citizens who believe that it is "essential to live in a democracy," and, hardly coincidentally, a rise in the number of people who would like to see a strong leader "who does not have to bother with elections." This chimes with the feeling that many of us have that society is becoming more polarized, and our political disagreements more fractious. Opponents cannot be persuaded or tolerated, and must simply be beaten.[1] The label Foa and Mounk attach to growing skepticism about the value of democracy is "deconsolidation," and this conveys the sense many of us have of something coming apart. Of course, there are many contributory factors that we might discuss. In particular, we might point to structural features of the global economy that concentrate wealth and power in the hands of a fortunate few.[2] Some bad behavior in the public sphere may seem of little consequence beside material conditions of inequality, but there are reasons to think that it does matter, even if there are other things that may matter more.

This paper addresses three questions. First, does it make sense to group the various kinds of bad disruptive speech together as a distinct family of related threats to broadly liberal representative democracies? Bald-faced lying and hate speech, for example, are clearly wrong for different reasons, so why treat them as if they are the same? Second, how can we distinguish bad forms of disruptive speech from good ones? If the former constitute a clear and present danger to democracy, then we will want to take action to curtail them, but we do not want to throw the baby out with the bathwater. If we are to protect and support good forms of disruptive speech, such as satire and the arts, then we must be able to tell which is which. And, third, what can we permissibly do about bad disruptive speech once we have isolated it? Facebook has appealed to the value of free speech as a justification for continuing to allow demonstrably false political advertising on their platform.[3] Is bad disruptive speech simply something that we are stuck with?

After showing why we need to employ a wide lens to capture this problem and explaining what I mean by disruptive speech, I argue in Sections 3 and 4 that the common thread linking practices that are primarily wrong for different reasons is that they all eat away at the stability of a democratic society. As I interpret it here, stability is a property of existing political communities. How much of it any particular polity possesses depends on the degree to which its citizens are disposed to play fair with one another and to refrain from imposing their particular conceptions of the good on everyone. Without a high level of stability, it would not be possible to have an open, democratic society, nor would it be possible for such a system to survive political shocks. For these reasons, citizens have a duty to refrain from engaging in bad disruptive speech, and, as a community, we have strong reasons to enact policies that will cultivate, rather than squander, stability. Section 5 illustrates how good forms of disruptive speech help to buttress and secure this form of stability.

I motivate the third question by showing in Section 6 how considerations of political legitimacy can give us pause when it comes to discouraging negative forms of disruptive speech. Since it is right to hold that an open and permeable public sphere is non-negotiable for a functioning democracy, we are left wondering what we can permissibly do to combat bad forms of disruptive speech. I argue that we are better placed to identify and evaluate the full range of options for dealing with speech-based threats to political stability when we can distinguish successfully between good and bad disruptive speech. I illustrate this point by sketching the contours of a two-pronged approach we might take. First, we could intervene to encourage good disruptive speech and do more to support those parts of the public sphere where it typically takes place. Second, we can then be more precise in our efforts to target bad disruptive speech. Even if we think that some forms of it must be permitted, it does not follow that it should be tolerated at all times and in all fora. We can, therefore, apply different standards and expectations to different activities, and this means that we can retain an open and permeable public sphere without insisting that we apply the most expansive conception of free speech to all aspects of it.

2 | DISRUPTIVE SPEECH

I have said that disruptive speech is speech that challenges or subverts widespread existing norms. What is it for speech to challenge or subvert a norm? I shall follow Jon Elster's (1989a, 1989b) account of norms as shared rules for behavior that are sustained by the approval and disapproval we direct at each other and at ourselves. Norms can thus only be said to come into

existence once they have actually been accepted by a group of people. Norms must be sharply distinguished from laws. The latter are backed up by the power of the state, but a system of norms depends on the expectation that if one does her part in upholding them then others will too, which is to say that it relies in large part on trust. To the extent that they are enforced, norms are policed informally using social sanctions such as shaming and ostracism. Elster describes norms as having "a grip on the mind" (1989a, p. 100) because the psychological process of internalizing a social grouping's rules for conduct places them under the auspices of our moral emotions. When we violate norms, we typically feel guilty, or even ashamed of ourselves. When others violate them, it usually sparks feelings of anger and indignation.

As an example of disruptive speech that violates norms, let us take the bald-faced lie. A bald-faced lie is a falsehood told with no intention to deceive the hearer.[4] According to *The Washington Post*, the former president of the United States, Donald Trump, made 16,241 false or misleading claims in his first 3 years in office, with the rate increasing markedly in 2019.[5] Many of Trump's false claims, such as that there was no "quid pro quo" in his infamous call with the president of Ukraine, can only be sensibly interpreted as bald-faced lies.[6] The transcript of the call drips with the implication that continued American aid to Ukraine was connected to the pursuit of an investigation into Trump's political enemies, and, particularly when placed in the context of subsequent revelations that Trump had frozen planned military aid to Ukraine, the truth of the matter is plain.[7] Though most observers could see that, Trump, who was facing impeachment proceedings, could not admit it without increasing his legal and political peril. The precedent that he established at the outset of the scandal inspired some Republican senators to hide behind bald-faced lies of their own.

Bald-faced lying in politics presents a serious challenge to our collective social assumptions and expectations of honesty and accountability. Not only does the liar violate the ordinary moral duty to tell the truth, they do so flagrantly. Since everybody already knows what the truth is, their lie suggests that the truth does not matter. Whether or not wrongdoers will be held to account then comes to depend on factors that are morally arbitrary, such as the relative political strength of their supporters and opponents.[8]

Another example of disruptive speech is hate speech which explicitly or implicitly denies the fundamental moral equality of some group of people. In particular, slippery forms of hate speech such as dogwhistles and ostensibly ironic uses of racist or sexist epithets have become increasingly common in our political discourse.[9] Badano and Nuti (2018) discuss the example of the French politician Marine Le Pen who often disguises her antipathy to Muslims as a defense of the rights of women and the LGBTQ+ community. This is speech that is designed to challenge and undermine norms of tolerance and respect toward a particular group within a community. However, it is not always straightforward to say precisely why the content of such speech is objectionable. Indeed, one of the reasons why it is effective is precisely because it does such a good job of masquerading as reasonable speech by drawing on the public political culture of the French political tradition.[10]

While hate speech typically attacks the presupposition that everyone has the same basic standing, more or less directly, the supposedly ironic use of hate speech by some in the so-called "alt-right" movement does so obliquely by undermining norms like those against giving Nazi salutes or using racist epithets that protect against overt hate speech. For one thing, by defending their right to do such things "as a joke," the strength of the general presumption against doing them at all is weakened.[11] For another, it renders the idea of questioning the status of targeted groups salient. As Mary Kate McGowan (2009, p. 403) has argued, undoing the various changes to the local rules and presuppositions of particular conversations that are

enacted by racist or sexist jokes can be extremely difficult. Memorably she compares it to trying to "unring a bell."

Note that while disruption and upheaval are often intended, as it is in the case of alt-right activists spreading their beliefs or Russian bots disseminating fake news, it need not be. What matters is that the speaker behaves in such a way that she presents a challenge to the authority of current norms, either by obviously contravening them or by suggesting the adoption of alternative norms that are incompatible with the existing ones.[12] Disruptive speech is thus a very broad category. Although I believe there are reasons to think that what I say here applies to all disruptive speech, in what follows I will limit myself to discussion of public disruptive speech, which I understand as disruptive speech that is meant to be heard by many people, most of whom will not be known personally to the speaker.

There is an influential liberal tradition stretching all the way back to Mill that embraces all disruptive speech and welcomes the expression of even the most wrong-headed ideas. On the standard Millian line, "[c]omplete liberty of contradicting and disproving our opinion is the very condition which justifies us in assuming its truth for purposes of action; and on no other terms can a being with human faculties have any rational assurance of being right," (Mill, 1974, p. 79). At the very least, Mill thinks, we should recognize in even the worst forms of speech an opportunity to reinvigorate our own convictions. It is in this vein that Steven Shiffrin (1990, p. 96) asserts that "the sponsoring and protection of dissent generally have progressive implications".[13] Jeremy Waldron's (1987) thoughts on the potential benefits of the experience of being offended offer a possible reason for taking this position one step further. Going beyond the value we may derive from confronting the content of offensive speech, he argues that the experience itself can be something positive as the shock can penetrate our ideological defenses and act as a spur to genuine reflection and growth.[14]

The controversial *Harper's Magazine* open letter, which took aim at "cancel culture" and was signed by a number of prominent writers and scholars such as Margaret Atwood and Noam Chomsky, can be read as a contribution to this tradition.[15] Notably, the signatories affirm that they "uphold the value of robust and even caustic counter-speech from all quarters," suggesting that disruptive speech is to be celebrated, and not merely tolerated.

Although this approach coheres with mine in so far as it acknowledges disruptive speech as a significant category, it implies that it is unified as a positive one. It is important to see how this view is misguided and why it needs to be more nuanced. If we can successfully establish a clear distinction between good and bad forms of disruptive speech then we stand a much better chance of promoting the former while at the same time discouraging the latter.

3 | STABILITY

In this section, I will borrow and adapt John Rawls's idea of stability to argue that bad forms of disruptive speech all undermine democratic models of government by weakening citizens' sense of justice, which is to say the disposition to pass up opportunities to use the power of the state to enforce one's conception of the good on everyone. On my revised understanding, stability is an actual good that existing societies have to a greater or lesser degree, and which can be cultivated with the right policies. What those policies need to get right—and are not currently getting right—is the balance between good and bad disruptive speech. If that analysis is correct, then we face the further question of what to do about it.

This paper is an exercise in non-ideal theory and, as such, does not aim to make a contribution to Rawls scholarship. Rawls, of course, was working primarily to determine what justice would look like in an ideal society, one which is well-ordered in the sense that almost everyone in it understands the basic principles of justice and willingly complies with them.[16] Clearly, this is not the case in the actual world, and bad disruptive speech as I understand it here is the kind of problem that we have to deal with precisely because some people are not minded to treat others fairly. I borrow the term "stability" from Rawls both because it is intellectually honest to acknowledge his influence on the development of my argument, and because some of the concerns that he raised about our moral psychology seem to me to be crucial for understanding and addressing the current political moment.[17]

As noted above, we might wonder whether there is any reason to group such disparate activities as fake news and hate speech together. They are obviously wrong for different reasons. Fake news is wrong because it is a form of deception, while hate speech is wrong because it attacks the right to equal standing of some targeted, and usually vulnerable, group. As good philosophers, we ought to distinguish carefully between different categories of wrong. Further, we might worry that running them together will also blind us to important differences in their causes and effects.

However, although it is true that these reasons for holding these actions to be impermissible are clearly different, they are not the *only* reasons why these forms of speech are wrong. The bigger risk is actually that we fail to see the important similarities between them. Alan Wertheimer (1999, p. 15) notes that "when we offer a moral description of an act, we typically invoke the strongest applicable moral description." Other reasons why the act is bad tend to drop out of view. He calls this the problem of occlusion. I will argue that problematic forms of disruptive speech all share the property of undermining stability. Even though they may also be impermissible for other, often more obvious, reasons, we should not lose sight of the fact that they are wrong for this reason as well.[18] This is what will allow us to mark a distinction between good and bad disruptive speech. Although good disruptive speech also breaks conventions and challenges existing norms, by doing so, it builds up stability rather than erodes it.

Rawls introduces the notion of stability to help decide between competing conceptions of justice. When conceptions are equally just we can appeal to other advantages they may have, including their stability.[19] On this theoretical level, Rawls tells us that a particular conception is more stable "if the sense of justice that it tends to generate is stronger and more likely to override disruptive inclinations and if the institutions it allows foster weaker impulses and temptations to act unjustly" (1999, p. 398). What is it to have a sense of justice? For Rawls (1999, p. 41), a person has a sense of justice if she possesses the intellectual capacity to judge things to be just or unjust on the basis of reasons and is, crucially, motivated to act in accordance with these judgments.[20] Even under ideal deliberative circumstances we should expect deep moral disagreements, so there will be many aspects of one's conception of the good that it would not be reasonable to expect other people to accept.[21] A person who possesses a sense of justice is prepared to be bound by a system of rules that other people acting in good faith can also find to be acceptable. This is a significant commitment because it means accepting that you cannot have things all your own way, even if your view of what makes for a good society happens to have majority support. You must be prepared to pass up opportunities to further your own view of the good when it would mean violating a set of rules which could, in principle, be justifiable to everyone.

It is worth stressing that Rawls is not satisfied by a mere *modus vivendi* in which fair rules are maintained because no one group has the power to seize control. As Brian Barry (1995,

p. 881–882) points out, Rawls is committed to the idea that a truly just society must be one in which its underlying principles are freely accepted by the vast majority of citizens. Similarly, our version of stability must be for the right reasons. Any particular individual's contribution to stability is predicated on their having reasons to develop and preserve an internal disposition to play fair with others and resist temptations to pursue their self-interest when they can see that it would be incompatible with respecting their fellow citizens as political equals.

As I shall use it here, stability is *not*, as Rawls understood it, a property of principles of justice, but rather a property of actual polities. It obtains when a high proportion of the populace have a sufficiently effective sense of justice to allow democratic procedures to operate, more or less, unimpeded and to insulate those procedures from shocks. By the operation of democratic procedures, I mean that governments are elected by popular vote in accordance with the principle of political equality, that they make and successfully enact just laws,[22] and that power is transferred peacefully from one regime to the next. By shocks, I mean events such as economic depressions and natural disasters that, while difficult, do not fundamentally alter the conditions of relative scarcity. Of course, in *A Theory of Justice* Rawls sets out to establish that it is rational for a citizen to maintain a sense of justice by showing that her good is congruent with his principles of justice, but we do not need to attempt anything so ambitious here.[23] Rather, it will be enough for our purposes if we can establish that widespread possession of a sense of justice is crucial for the normal operation of democratic politics and to bolster its resilience in trying times. If we can establish that, then it would follow from a deeper duty to support (or at least not harm) democratic institutions in suitably just states that citizens have a duty to refrain from actions that would corrode stability by weakening others' sense of justice. Proving the existence of a duty to support democratic institutions would take me beyond the remit of this paper, but it is well-trodden ground and I think it will suffice to note that on grounds such as political equality it is widely recognized that such a duty does exist.

I will offer two arguments for thinking that it is vital for democratic systems that a high proportion of citizens have, and are assumed by one another to have, a robust sense of justice. In the rest of this section, I will use the example of the peaceful transition of power to illustrate the role that a sense of justice plays in underwriting cooperation between representatives of competing comprehensive doctrines. In the next section, I will examine the two threats that Rawls identifies to maintaining a healthy sense of justice and show how disruptive speech can exacerbate both of them. We will see how that, in turn, destabilizes democracies.

As Rawls observed, our societies are characterized by deep disagreements about such questions as what constitutes a good life and how we should best organize our communities to facilitate the living of it. This is one reason why politics is often so fractious—there is so much at stake. The prize is the opportunity to use the awesome power of the modern state to reshape society in line with your values. Even in a well-ordered Rawlsian society with constraints such as the demands of public reason placing significant limits on how laws and policies must be justified and applied, there is still enormous scope for office-holders to influence the character of the state and the ways in which it affects the lives of its citizens. In the world as it is, the prize is even more consequential.

Bearing this in mind, it is a colossal risk to transfer power to one's political opponents. Not only do you give up the opportunity to wield that power in pursuit of what you think to be right, you hand it over to people with whom you and your voters disagree, perhaps on fundamental moral issues. On a day-to-day basis, you will then be reliant upon the restraint of your political enemies to leave you and yours the space to pursue your conception of the good. You are also counting on them to reciprocate the next time around if, and when, the political tides

turn. The peaceful transfer of power between political opponents thus makes no sense unless there is a high degree of trust. More specifically, all sides must trust that everyone will refrain from abusing the coercive power of the state to illegitimately bring about their political goals.

What would warrant such a profound form of trust? My contention is that a well-founded belief that your opponents have a motivationally efficacious sense of justice would provide such a warrant. One is justified in assuming that citizens who are like this will not only see why they should resist temptations to abuse power, but are also reliably motivated to do so. This is to say that their possession of a robust sense of justice means that they are *trustworthy* in a political context.

4 | HOW DISRUPTIVE SPEECH THREATENS STABILITY

Let us recap the argument so far. Our concern is with how a wide range of speech acts are poisoning our political discourse. To ensure that we capture the wrong that unites all of them, and especially the slippery forms of speech that defy easy categorization, I have suggested that we focus on disruptive speech, which I define as speech that either explicitly or implicitly challenges some widely-accepted norm. My central claim is that bad forms of disruptive speech all share the property of undermining the ability of a democratic system of government to operate effectively, and to be resilient in the face of moderate shocks. In contrast, we will see that good disruptive speech has the opposite effect.

In setting out my conception of stability, I have drawn on a modified version of Rawls's notion of a sense of justice. This disposition to be fair in one's dealings with others is not easy to maintain, however. Even though I am proposing an understanding of stability that is significantly different from what Rawls envisaged, one reason for presenting this discussion as an adaptation of his idea is because the main threats he identifies to the stability of his well-ordered society are also challenges to developing and preserving an effective sense of justice in a non-ideal setting. Rawls (1999, pp. 295–296) discusses two chief causes of instability: the attraction of gaining an advantage by ignoring the established rules and the awareness that the same temptation exists for everyone else. Disruptive speech can, of course, attack one's sense of justice directly, as when hate speech is deployed to persuade the members of one group that the members of another group are less than fully human and so unworthy of being treated as equals. However, I contend that the more elusive forms of disruptive speech that I have identified here also have the effect of hollowing out citizens' sense of justice by exacerbating one or both of Rawls's mechanisms of instability. Indeed, this may be one of the primary political motivations for engaging in them.

The first cause of instability that Rawls describes is essentially the problem of free-riding. Whenever an individual can benefit from a social rule without contributing to its upkeep there will be an incentive to default. An example of this occurred in the 2019 British general election campaign. During a televised debate between Boris Johnson and Jeremy Corbyn, the Conservative Party changed the name of their official Twitter account to pose as a neutral fact-checking organization. Twitter's accreditation system at that time for official accounts did not anticipate such a move, and so the blue badge that the party had secured remained on the account even though they were now purporting to be impartial observers. To the extent that the move worked, it did so because other verified accounts did not misrepresent their identity and thus gave users reason to trust the system.

Disruptive speech like this has multiple effects. A direct consequence in this case is that users' faith in supposedly official online identities will be shaken. There are also indirect consequences when social rules are broken. For one thing, it normalizes rule-breaking behavior, both in particular and in general. Normalization robs the softer social sanctions such as shame of much of their bite and makes it less costly for others to act in a similar fashion. In turn, this changes initial calculations of risk and reward, effectively upping the incentive to break a relevant rule. For another, norm violations make options that were previously unthinkable salient. We saw this in the example of bald-faced lying which was discussed in Section 2. Being caught out in a demonstrable falsehood used to be extremely costly for politicians, and was often fatal for their careers. The only available course of action was to deliver a groveling apology and hope that the outrage would subside. Now, however, politicians are keenly aware that there are alternatives, such as doubling down or brazening it out. When options like this become live, we should expect to see them taken much more often.

The second cause of instability arises from the knowledge that temptations such as these exist. We become concerned that others might be thinking about taking us for a ride. The fear is not just that we are losing out relative to others, but that they are *taking advantage* of us and our commitment to fair play. Specifically, they are exploiting our sense of justice. Once this worry takes root in our political culture, it generates a relentless, preemptive Hobbesian logic. Since others will inevitably strike when they can to secure not only their immediate self-interest, but, ultimately, their conception of the good, the only defense is to beat them to the punch. You must get your retaliation in first.

Here is an illustration. Some on the political right argue that "political correctness" is a concerted attack on their way of life. The claim is that it amounts to "blanket condemnation of people who don't hew to progressive ideals, in a way that is inimical to free speech and ideological diversity" (Aly & Simpson, 2019, p. 125).[24] Couched in terms of stability and a sense of justice, the fear is that rather than competing fairly in the "marketplace of ideas,"[25] the opposing side have worked out how to foreclose legitimate debate and smuggle leftist ideals into public discourse as unquestionable presuppositions that are thereby imposed on everyone. Then the gloves come off. Why should you adhere to the rules of draughts when the other side is playing three-dimensional chess? If you really believe that the liberals will not stop until Big Brother is precensoring everything you say, then noxious disruptive tactics such as using hate speech "ironically" to draw opponents into embarrassing overreactions that reveal their nefarious intentions appear justifiable. In an existential conflict, the only acceptable course of action is the one that leads to victory.

It would be a fool's errand to try to list every existing or conceivable form of bad disruptive speech and then show how each individual variety enhances one or both of the causes of instability that we have been discussing. In general, we can note that paradigm forms of bad disruptive speech such as fake news, bald-faced political lies, and hate speech have local and global effects that are detrimental to stability as I have defined it here. As we have seen, these behaviors undermine particular norms, such as the norms of authenticity and honesty. More importantly, though, they have the pernicious effect of eating away at a person's belief that their fellow citizens, and especially those citizens who have worldviews very different to their own, possess an effective sense of justice. As this belief withers away, so too does that person's own sense of justice because there is no point playing by the rules if you think you are playing alone. As Rawls (1999, p. 296) says: "given circumstances of mutual fear, even just men may be condemned to a condition of permanent hostility."

5 | GOOD DISRUPTIVE SPEECH

Rawls (1999, p. 6) suggests that a mark of a stable scheme of cooperation is that "when infractions occur, stabilizing forces should exist that prevent further violations and tend to restore the arrangement." We can find a number of important stabilizing forces in the public sphere.[26] One of the chief roles of the news media, for instance, is to hold public figures and powerful organizations to account. In the traditional role of public watchdog they ferret out and expose wrongdoing. In his book *Hack Attack* about the enormous British phone-hacking scandal, the journalist Nick Davies (2015) recounts just such a process directed toward a range of newspapers and their parent companies. It might be objected that reporting criminality, corruption, and general bad behavior draws attention to the fact that there are a lot of people out there who are not playing by the rules, but this is to miss the point. Rawls's insight into our moral psychology is that we already worry about this. A watchdog press does something about it. In this section, I will discuss two kinds of paradigmatically disruptive speech that can contribute to the stability of the public sphere in different ways: satire and the arts. If I am correct, then we can use stability to distinguish between good forms of disruptive speech that should be supported and bad forms of disruptive speech which should be discouraged.

I will start with satire. Satire aims to be disruptive. Its targets are the things that have authority for us. Typically that means individual powerful people, but it can also mean institutions, practices, conventions, beliefs, and so on. Satire hobbles the authority of its subject by drawing attention to reasons that render it preposterous that it should hold whatever position of power it occupies. That is the principal reason why it is funny (when it is)—the juxtaposition of status and unsuitability—which also explains its characteristic absurdity and dark humor.[27] In particular, satire specializes in exposing deficiencies of character by plumbing the behavior of politicians, commentators, and even voters to expose injustice, hypocrisy, incompetence, and buffoonery.

An important implication of this view is that a sincere attempt at satire cannot intentionally "punch down."[28] It is, of course, commonplace to see humor used to denigrate and demean vulnerable individuals and groups,[29] but you cannot set out to undermine someone's privileged standing in the community if you know that they simply do not have it.[30] This is why I consider the general practice of satire to fall under the umbrella of good disruptive speech.[31] Though it will always be possible to mimic the conventions of satire, and so deceptively represent bad disruptive speech as satire, that will not make it so. On this understanding, genuine satire can help to strengthen and protect citizens' sense of justice in a number of ways.

In the Adam McKay film "Vice," Christian Bale portrays former US Vice-President Dick Cheney. As the film progresses, Cheney becomes obsessed with power for its own sake. That is his conception of the good and he seizes each opportunity to secure it, with no regard for the effect of his actions on the wider community. He is gradually consumed by this mission, eventually sacrificing his one and only redeeming feature on the altar of politics. Putting this story on screen is a way of holding Cheney, and others in his circle, to account. It lays out a case for all to see that he has led a shameful life and that to be like him is to be unfit for high public office. Employing the tools of exposure and ridicule, satirical movies, shows, and publications thus attach a tangible, and not inconsiderable, cost to unjust behavior. Perhaps more importantly, though, laughing along with them can be a way of rejecting such behavior and reaffirming one's own commitment to abide by norms of fairness.

Satirical works can encourage us to reflect on our own conduct, which is why its proponents often reach for shocking language and imagery.[32] The makers of the cartoon "South Park" are

past masters at using offense to cut through rationalizations and obfuscations.[33] The program has a ridiculous or grotesque character to represent the dangers of almost any deficit or excess of character, with particular scorn reserved for self-righteousness. Done well, satire has the potential to inspire a little humility, which can be helpful in appreciating the significance of the burdens of judgment and so conducive to the creation of an atmosphere of tolerance and inclusion. Humor also provides a release valve for some of the fear and insecurity that comes with seeing your political opponents in power. By puncturing rage, laughter can reduce the likelihood that understandable fears about the strength of other people's sense of justice will lead an individual to give up her own sense of justice.

The arts can also make an important, and patently disruptive, contribution to stability. Ai Weiwei's "Soleil Levant" installation in Copenhagen is a case in point. The salvaged lifejackets crammed in so tight that they are almost bursting out at pedestrians are a potent argument against prevailing attitudes to the migrant crisis. The artwork challenges belief in an "us" and a "them," using tactile objects to summon up visceral feelings about what it would be like to wear one of these damp, smelly lifejackets in cramped and life-threatening conditions. As these are not conditions that it would make sense for any rational person to voluntarily choose, it forces questions on passers-by about their tacit acceptance of complacent political norms that distinguish between, for example, refugees and economic migrants, and thus license us to treat desperate people as being somehow responsible for their plight.

One of art's most powerful effects is to entice us away from our normal standpoint so that we can experience what it might be like to be someone else. This perspective-taking can enliven considerations of justice of which we are dimly aware, but fail to accord their due weight. In the case of Weiwei's artwork, this is particularly true of the urgency of basic needs. It can also strengthen our commitment to treat others as equals more generally. By emphasizing the ways in which our fragile bodies are fundamentally the same, the piece has the effect of making it easier to believe that migrants are capable of having and sustaining a sense of justice, and so are equally suited to bearing the responsibilities of citizenhood. As we have seen, motivation matters for stability. Art can speak to people on an emotional level, buttressing stability by leveraging empathy against suspicion and mistrust.

Of course, not all art or satire is good. Some artworks are vapid and some satire is dull. Another way in which such work goes wrong, however, is when it weakens the bonds of solidarity that ought to be promoted between citizens. In the case of art, this is most clearly seen in state-sponsored propaganda pushing narrow, exclusionary agendas.[34] Even if you reject the account of satire I offered earlier, on my view of disruptive speech we can still use stability to distinguish between good and bad speech within categories of speech that are presumptively beneficial. In the next section, I consider how this distinction can help us to respond to the threat posed by bad disruptive speech.

6 | A MORE NUANCED APPROACH TO DISRUPTIVE SPEECH

There is, however, reason to think that, in spite of all the reasons I have outlined for concern about the pernicious effects of bad disruptive speech, we are constrained in what we can permissibly do to combat it. As Jürgen Habermas (1996, p. 359) argues, the public sphere provides a "sounding board for problems that must be processed by the political system because they cannot be solved elsewhere." In order to work out what it is that we should do about the kinds

of sticky problems that require collective action, we need to talk it out. On this line, the public sphere is like an incredibly fast computer where our individual rational capacities combine to provide the processing power. The more citizens involved in thinking about and debating the issues of the day, the more powerful the processor. Because all of us can enter the public sphere, it is always permeable to new information. As individuals encounter and experience problems that require collective action, they can bring them to the attention of the rest of us. Habermas (1996, p. 381) cites issues such as environmental degradation and the treatment of women as illustrations of how issues are raised at the periphery of the public sphere by activists and scholars and gradually gain traction. Further, having public fora for discussion of both the political agenda and the potential policies that might be enacted is a way of opening up alternative channels through which people can get involved in the business of government on an ongoing basis. The more that the discussion and planning phase of concrete decisions can be made accessible to ordinary citizens, the less worried we should be that politicians, civil servants, lobbyists, etc., are more equal than the rest of us. The existence and smooth functioning of an open public sphere are, therefore, vital to substantiating and securing the equal status of all citizens within the political process. Both procedural and outcome legitimacy thus furnish strong reasons for states to refrain from policing people's contributions to the public sphere.[35]

What does this mean for disruptive speech? There is a stronger and weaker interpretation of the point. On the stronger interpretation, we should welcome all disruptive speech. As we saw in Section 2, for Mill everything is grist. Even completely wrongheaded contributions have value because they force us to rediscover the truths buried in our dogmas. On the weaker interpretation, even though some speech may not actually be helpful to our enquiries, we risk too much by deliberately excluding particular perspectives—not least because of the chilling effect it may have on other potential participants. We would do better to trust that in the fullness of time the process will deal with fringe views that do not have anything worthwhile to offer. This suggests that even patently noxious speech must simply be tolerated as the cost of doing business. Bad disruptive speech, from late-stage Morrissey to Russian bots, might be a problem, but to avoid the downsides associated with heavy-handed state intervention in the public sphere, perhaps we simply have to live with it.

Of course, this is not a decisive argument, and the very emergence of bad disruptive speech gives us reason to be sceptical of the Millian line. Wherever you think the balance of reasons lies here, though, there are at least some costs attached to any effort to restrict speech rights so an unpalatable trade-off is clearly in the offing. Fortunately, the distinction I have drawn between good and bad disruptive speech can provide some assistance. In the rest of this section, I will briefly sketch the outlines of two complementary approaches we might employ that are suggested by it. First, we can try to shore up, and even increase, stability by cultivating good forms of disruptive speech, and, second, we can be much more discerning in how we understand and apply the demands of legitimacy. A lot of philosophical discussion about the limits of speech and expression orbits the general question of whether we should acknowledge a moral and/or legal right to say obnoxious, or even dangerous, things.[36] This level of abstraction can, however, be misleading. Most of the time we do not face a simple binary choice between allowing or prohibiting some form of undesirable speech all across the public sphere. Rather, what we have to decide is where, when, and how to intervene to discourage it. Even if we have to allow people to engage in bad disruptive speech, we do not have to allow them to engage in it wherever and whenever they choose. Different permissions and requirements can be applied to the various parts of the public sphere depending on the function we want them to serve. We could take advantage of this flexibility to make it considerably harder, though not impossible, for bad disruptive speech to chip away at stability.

The idea that certain activities can make special contributions to our public discourse should not be especially surprising. In discussing the significance of state speech, Corey Brettschneider (2012) raises the example of public statues, which are often used to promote democratic ideals by signaling that someone who is closely associated with them is worthy of respect and emulation. It is thus relatively uncontroversial to use statues and other forms of commemorative public art to foster democratic values. Now that we are better equipped to identify and appreciate the significance of good disruptive speech, we could be much more discerning, imaginative, and ambitious in how we deploy state funds and other resources, such as national broadcasters, public spaces, educational curricula, etc., to encourage good disruptive speech, and, ultimately, direct public attention and conversation in ways that are conducive to the development of a strong sense of justice. Even for those citizens who prefer not to engage with these activities, a significant investment of resources by the state sends its own message.

Once we distinguish clearly between good and bad disruptive speech it is open to us to find creative ways to nurture the former. But we can also be more forensic in dealing with the latter. Taken to its logical conclusion, the arguments advanced above for an open and permeable public sphere insist that people be legally, if not morally, permitted to say whatever they like, no matter how toxic or hateful. Even if we accepted that conclusion, which of course we may not, it would not follow that they must be allowed to do so absolutely everywhere. A quiet carriage on the train is not the same kind of space as an online comment thread, which is not the same kind of space as the front page of a newspaper of record. We might, for instance, reserve places on the internet as arenas for almost[37] unrestricted free speech or designate locations in the physical world where fringe groups could meet, while imposing a blanket prohibition on hate speech in city squares and parks, news broadcasts, dominant social media sites, and so on.

Someone who values all disruptive speech might object that measures like this are, in effect, a form of quarantine, which violates the requirements of both procedural and outcome legitimacy. Restricting the participation of some citizens to particular sub-domains undermines their ability to feed into the real political agenda and prevents whatever occasional insights or discoveries do arise there from spreading to the mainstream discourse. The public sphere, therefore, would neither be open to everyone nor to all new information and ideas. However, legitimacy cannot require that we establish a right for everyone to be heard. What matters is that we ensure possession of a substantive right to speak. Although it is necessarily true that any restrictions would make it harder to use that right to engage in bad disruptive speech, it would by no means make it impossible. Further, individuals would remain free to participate across the public sphere so long as they follow the relevant local rules. And while the goal is indeed to mark out the boundaries of the various kinds of communicative practices that are going on, there is no reason in principle why these boundaries should not be sufficiently permeable to allow genuinely useful ideas to cross over if they attract substantial attention and debate.

Of course, this is not likely to mollify a true disruptive speech absolutist, but there is a limit to far we should go to appease such a person in any case. What matters is that we can provide a justification that gives appropriate weight to the real concerns that the position highlights. If this is right, then we are not faced with a simple trade-off between legitimacy and stability. The two complementary strategies I have floated in this section are unhappily vague, and further development would require considerably more philosophical reflection, as well as substantial engagement with a range of empirical questions, but hopefully they serve to illustrate the value of the distinction between good and bad disruptive speech, and point toward potential courses of action we can undertake to preserve stability that will not jeopardize the legitimacy of a democratic state.

7 | CONCLUSION

The jumping-off point for this paper was the observation that our public sphere is struggling to cope with various kinds of disruptive speech, which is to say speech that contravenes and calls into question existing social and political norms. These forms of speech are clearly having a deleterious effect on the quality and tenor of our public discourse, but, arguably also constitute a serious threat to democratic political systems. I have shown that although many of the paradigm forms of bad disruptive speech are wrong for different reasons, they are also wrong because they share the feature of undermining the capacity of democratic states to govern in accordance with democratic procedures and to cope with the kinds of political shocks we have become familiar with over the last number of years. They have this effect because they directly or indirectly weaken and erode citizens' sense of justice. When a person's sense of justice fails, then they can be expected to view politics as a zero-sum game and act accordingly.

Good disruptive speech, on the other hand, has the opposite effect. I offered examples from satire and art to illustrate how unjust and inegalitarian norms can be constructively challenged in ways that encourage us to hold fast to a disposition to treat others fairly, and to retain the belief that they can, at least in benign circumstances, be trusted to treat us fairly in return. If I am right that good disruptive speech contributes to the success of a democracy by scaffolding its stability, then it is crucial to be able to distinguish it. This is a nuance that many theories of the value of free speech coming from the Millian tradition miss. With this distinction in hand, we are better placed to identify and evaluate the full range of options for discouraging bad disruptive speech, on the one hand, and encouraging good disruptive speech, on the other. If we wish to preserve, protect, and, indeed, even to promote the stability of democratic political communities, then one of the things that we should do is to tilt the balance toward good forms of disruptive speech.

ACKNOWLEDGMENTS

For helpful comments and feedback, I am indebted to audiences at the Dublin Political Theory Workshop in UCD, the Legitimacy and Stability in Fractious Times Workshop at the University of Leeds, the IDEA Centre Work in Progress group, the Global Ethics Conference at the University of Birmingham, and the 2018 Society for Applied Philosophy Annual Conference in Utrecht. For comments on various drafts, I would also like to thank Jessica Begon, Richard Healey, Jan Kandiyali, Joey Lacey, Jonathan Parry, Angie Pepper, Kristin Voigt, and the anonymous reviewers.

CONFLICT OF INTEREST STATEMENT

The author declares that there are no conflicts of interests.

ORCID

Carl Fox https://orcid.org/0000-0001-5657-3586

ENDNOTES

[1] This is reflected in the evolution of electoral strategies. Political parties now spend most of their time and resources ensuring that voters who are already sympathetic to their message actually turn out and vote at election time, rather than attempting to persuade more skeptical citizens to change their minds. See Issenberg (2013).

[2] For instance, see Chambers and Kopstein (2001) for an argument that bad civil society is primarily a function of socioeconomic injustice.

[3] https://www.theguardian.com/technology/2020/jan/09/facebook-political-ads-micro-targeting-us-election.

[4] For a discussion of bald-faced lies see Fallis (2009).

[5] https://www.washingtonpost.com/politics/2020/01/20/president-trump-made-16241-false-or-misleading-claims-his-first-three-years/.

[6] His false claims about imaginary electoral fraud in the 2020 presidential election are a trickier case. Although there was no remotely plausible evidence to support them and much of his behavior at the time—particularly behind the scenes—appeared to show that they were a transparent pretext for abusing his power to alter the result, a key element of the strategy involved bringing grassroots pressure to bear on local and national officials. For that reason it also relied on convincing a large proportion of Republican voters that the claims were, in fact, true.

[7] The transcript of the call is available here: https://edition.cnn.com/2019/09/25/politics/donald-trump-ukraine-transcript-call/index.html.

[8] Another species of Trumpian deception that I cannot explore here is the deliberate creation of confusion about some issue or event. Writing about the release of the Pentagon Papers, Arendt claims that it is counter-productive to confuse people without convincing them. She thinks that the efficiency of lies "depends entirely upon a clear notion of the truth that the liar and deceiver wishes to hide," and that for this reason truth "possesses an ineradicable primacy over all falsehoods," (Arendt, 1973, p. 30). It is not clear to me that this holds in a world where the pace of the news agenda moves so much faster. Indeed, when polarization and partisanship are so entrenched that bald-faced lies are a viable political strategy, general confusion may provide an even more effective fig leaf.

[9] On dogwhistles, see Saul (2018).

[10] On Lynne Tirrell's (2017, 2018) account of toxic speech, we can explain how Le Pen's speech is intended to work. Her rhetoric operates in a complex linguistic context and in that context her denunciations of a caricature of the Muslim faith are understood by many of her listeners as a speech act licensing various real-world acts of discrimination against Muslims. In the terminology Tirrell uses, these are "exit moves" from a language game. Tirrell compares speech that engenders harmful exit moves to toxins that interact with our biology in ways that damage our bodies. However, Tirrell admits that "[i]t is not clear how to identify the toxicant in discursive toxicity," (2018, p. 131). As in the medical case, we must rely on our observations of the harmful effects that have followed from various speech types in the past. As it is difficult to tell in advance which substances will be dangerous to human beings, it can be difficult to predict when, or if, specific harms will follow from new tactics and patterns of speech. The account I construct tells us to look for speech that undermines stability by attacking the hearer's sense of justice. I will explain this in more detail in Section 4, but bad disruptive speech does this directly by rejecting the idea that some vulnerable group have the same standing as the rest of us, or indirectly, either by making it more attractive for that individual to break social rules, or by exacerbating her fear that others will break them. Similarly we can use the ideas of a sense of justice and stability to be more precise about the ways in which good disruptive speech can play the role of "inoculations" or "antidotes" (Tirrell, 2017, p. 146) such as egalitarian rhetoric that challenges unjust norms.

[11] In this sense we may view the weaponisation of provocative, edgy humor as a complementary strategy to the traditional, overt hate speech that is also often deployed by the alt-right. I thank an anonymous reviewer for pushing me on this point.

[12] It may be that these alternative norms are already accepted and upheld among some communities. What matters for my purposes is simply that the speech challenges at least one widespread norm.

[13] Shiffrin is not blind to the fact that dissenters are often wrong and wish to bring about undesirable change, but like Mill he believes that truth and wisdom will ultimately prevail. He also echoes another of Mill's arguments for free speech when he writes that "[p]romoting dissent can promote its own pathologies, but the sponsoring of dissent is necessary for any healthy individuality to flourish," (1990, p. 93).

[14] I am not claiming that Waldron himself, given his other commitments, should be included in this tradition, but his argument in that paper does fit rather neatly with it. It is also worth noting that Waldron's position

here is not inconsistent with his more recent book on hate speech. There he explicitly distinguishes between causing a person offense and undermining their dignity. Laws that prohibit hate speech ought only to protect against the latter. See Waldron (2012, pp. 105–111).

[15] https://harpers.org/a-letter-on-justice-and-open-debate/.

[16] A well-ordered society is "a society in which everyone accepts and knows that the others accept the same principles of justice, and the basic social institutions satisfy and are known to satisfy these principles," (Rawls, 1999, p. 397).

[17] One of the few things that almost all political philosophers agree on is that in terms of realizing the demands of justice we have an awfully long way to go. There is, therefore, an interesting debate to be had about whether Rawlsians should adopt or reject stability as I envision it here as an intermediary goal within a larger transitional plan to bring about a well-ordered society. This raises questions about the relationship between one's ideal and non-ideal theories, and whether commitments in one domain can have implications for the other. Though I cannot explore these questions properly here, I thank an anonymous referee for pushing me to acknowledge them. For solid staring points and helpful discussion on some of these questions see Erman and Möller (2022), Hamlin and Stemplowska (2012), Sen (2009), Simmons (2010), Valentini (2012), and Wiens (2012).

[18] I will not here consider comparative questions about which wrongs are more grievous. My argument does not rely on the claim that undermining stability is more serious than other potential wrongs committed in the course of bad disruptive speech. Indeed, it seems likely that occlusion has taken place precisely because oftentimes it is not. However, since undermining stability is incompatible with the duty to support democratic institutions it does seem to me to be a serious wrong. I thank an anonymous reviewer for pushing me on this point.

[19] Rawls (1999, pp. 5–6) also discusses coordination and efficiency.

[20] See also Rawls (2005, p. 19).

[21] See Rawls (2005, pp. 56–57) on the burdens of judgment.

[22] It should be emphasized that communities that are stable in this way are not static. The principles underpinning democratic systems of government mandate progress and reform when justice has not been realized. Indeed, as we shall see, promoting good disruptive speech is one important way in which states can do this. I thank an anonymous reviewer for pushing me to clarify this point.

[23] See Barry (1995) and Weithman (2013) for two painstaking discussions of the transition in Rawls's thinking about stability from *A Theory of Justice* to *Political Liberalism*.

[24] Aly and Simpson (2019, pp. 126–129) argue that this view is mistaken, and, in fact, it is simply politics as usual to fight over the boundaries of acceptable speech. On this point, see also Fish (1994).

[25] See Sparrow and Goodin (2001) on the limitations of this metaphor.

[26] For a complementary idea to good disruptive speech see what Lepoutre (2019) calls "positive forms of counterspeech". He focuses on how individuals and, in particular, the state can combat misinformation and disinformation by "affirming a correct vision of the world that is incompatible with the falsehoods at hand," (2019, p. 167). Although this approach is limited to dealing with what he calls "ignorant speech" as opposed to the wider category of bad disruptive speech, it strikes me as a very worthwhile strategy.

[27] For a contrary view, see Declercq (2021), who holds that humor is not a defining feature of satire and argues that it must merely aim to entertain while engaging in serious moral critique.

[28] Further, satire that punches down unintentionally will fail on its own terms simply because it is directed at an inappropriate target. A comedian might attack an underprivileged group because of a mistaken belief that they hold a position of undue privilege or power, but such a performance must fail because, as a matter of fact, there is no special position of authority that the targets hold for which they are unsuitable. The real absurdity is actually in the conception and execution of the performance.

[29] Though see Basu (1999) and Neu (2008, p. 228) on this point. They note how efforts to deploy humor to bully and oppress often seem strained.

[30] For an account of what it means to be "marked out" for systematic oppression see Haslanger (2000, pp. 40–41).

[31] I thank an anonymous reviewer for pushing me to clarify these points.

[32] For an insightful treatment of the benefits of moral distress see Waldron (1987).

[33] I thank Josh Hobbs for a helpful discussion of this example.

[34] For an account of propaganda see Stanley (2016, esp. Ch. 2).

[35] There is an enormous literature on the concept of legitimacy and it is not possible to explore the idea in any depth here. However, for some background see Buchanan (2002), Estlund (2008), Nagel (1987), Quong (2004), and Simmons (1999).

[36] See, for instance, Howard (2019). This is not to deny the value of this level of analysis, but merely to emphasize the moral significance of the differences between different kinds of spaces and activities in the public sphere.

[37] I will not here deal with threats or crowded theaters, but I assume that at least some basic restrictions on speech are uncontroversial.

REFERENCES

Aly, W., and R. M. Simpson. 2019. "Political Correctness Gone Viral." In *Media Ethics, Free Speech, and the Requirements of Democracy*, edited by C. Fox and J. Saunders. Abingdon: Routledge.

Arendt, H. 1973. *Crises of the Republic*. London: Penguin.

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

Barry, B. 1995. "John Rawls and the Search for Stability." *Ethics* 105(4): 874–915.

Basu, S. 1999. "Dialogic Ethics and the Virtue of Humor." *Journal of Political Philosophy* 7(4): 378–403.

Brettschneider, C. 2012. *When the State Speaks, What Should it Say? How Democracies Can Protect Expression and Promote Equality*. Princeton, NJ: Princeton University Press.

Buchanan, A. 2002. "Political Legitimacy and Democracy." *Ethics* 112: 689–719.

Chambers, S., and J. Kopstein. 2001. "Bad Civil Society." *Political Theory* 29(6): 837–65.

Davies, N. 2015. *Hack Attack: How the Truth Caught up with Rupert Murdoch*. London: Vintage.

Declercq, D. 2021. *Satire, Comedy and Mental Health: Coping with the Limits of Critique*. Emerald Publishing: Bingley.

Elster, J. 1989a. *The Cement of Society: A Study of Social Order*. Cambridge: Cambridge University Press.

Elster, J. 1989b. "Social Norms and Economic Theory." *Journal of Economic Perspectives* 3(4): 99–117.

Erman, E., and N. Möller. 2022. "Is Ideal Theory Useless for Nonideal Theory." *The Journal of Politics* 84(1): 525–40.

Estlund, D. 2008. *Democratic Authority: A Philosophical Framework*. Princeton, NJ: Princeton University Press.

Fallis, D. 2009. "What Is Lying?" *Journal of Philosophy* 106(1): 29–56.

Fish, S. 1994. *There's No Such Thing as Free Speech, and it's a Good Thing, Too*. Oxford: Oxford University Press.

Foa, R. S., and Y. Mounk. 2017. "The Signs of Deconsolidation." *Journal of Democracy* 28(1): 5–16.

Habermas, J. 1996. *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*. Cambridge: Polity Press.

Hamlin, A., and Z. Stemplowska. 2012. "Theory, Ideal Theory and the Theory of Ideals." *Political Studies Review* 10: 48–62.

Haslanger, S. 2000. "Gender and Race: (What) Are They? (What) Do We Want Them to Be?" *Nous* 34(1): 31–55.

Howard, J. W. 2019. "Dangerous Speech." *Philosophy and Public Affairs* 47(2): 208–54.

Issenberg, S. 2013. *The Victory Lab: The Secret Science of Winning Campaigns*. New York: Broadway Books.

Lepoutre, M. 2019. "Can "More Speech" Counter Ignorant Speech?" *Journal of Ethics and Social Philosophy* 16(3): 155–91.

McGowan, M. K. 2009. "Oppressive Speech." *Australasian Journal of Philosophy* 87(3): 389–407.

Mill, J. S. 1974. *On Liberty*. London: Penguin.

Nagel, T. 1987. "Moral Conflict and Political Legitimacy." *Philosophy & Public Affairs* 16(3): 215–40.

Neu, J. 2008. *Sticks and Stones: The Philosophy of Insults*. Oxford: Oxford University Press.
Quong, J. 2004. "The Rights of Unreasonable Citizens." *Journal of Political Philosophy* 12(3): 314–35.
Rawls, J. 1999. *A Theory of Justice: Revised Edition*. Harvard University Press: Cambridge, MA.
Rawls. 2005. *Political Liberalism: Expanded Edition*. Columbia University Press: New York, NY.
Saul, J. 2018. "Dogwhistles, Political Manipulation, and Philosophy of Language." In *New Work on Speech Acts*, edited by D. Fogal, D. W. Harris, and M. Moss. Oxford: Oxford University Press.
Sen, A. 2009. *The Idea of Justice*. Cambridge, MA: Harvard University Press.
Shiffrin, S. H. 1990. *The First Amendment, Democracy, and Romance*. Princeton, NJ: Princeton University Press.
Simmons, A. J. 1999. "Justification & Legitimacy." *Ethics* 109: 739–71.
Simmons, A. J. 2010. "Ideal and Nonideal Theory." *Philosophy & Public Affairs* 38(1): 5–36.
Sparrow, R., and R. E. Goodin. 2001. "The Competition of Ideas: Market or Garden?" *Critical Review of International Social and Political Philosophy* 4(2): 45–58.
Stanley, J. 2016. *How Propaganda Works*. Princeton, NJ: Princeton University Press.
Tirrell, L. 2017. "Toxic Speech: Toward an Epidemiology of Discursive Harm." *Philosophical Topics* 45(2): 139–62.
Tirrell, L. 2018. "Toxic Speech: Inoculations and Antidotes." *The Southern Journal of Philosophy* 56(S1): 116–44.
Valentini, L. 2012. "Ideal vs. Non-ideal Theory: A Conceptual Map." *Philosophy Compass* 7(9): 654–64.
Waldron, J. 1987. "Mill and the Value of Moral Distress." *Political Studies* 35(3): 410–23.
Waldron, J. 2012. *The Harm in Hate Speech*. Cambridge, MA: Harvard University Press.
Weithman, P. 2013. *Why Political Liberalism? On John Rawls' Political Turn*. Oxford: Oxford University Press.
Wertheimer, A. 1999. *Exploitation*. Princeton, NJ: Princeton University Press.
Wiens, D. 2012. "Prescribing Institutions Without Ideal Theory." *Journal of Political Philosophy* 20(1): 45–70.

AUTHOR BIOGRAPHY

Carl Fox is a lecturer at the IDEA Centre in the School of Philosophy, Religion, and History of Science at the University of Leeds. His main interests are in political philosophy, and he works primarily on the ethics of the public sphere. He is the co-editor of *Media Ethics, Free Speech, and the Requirements of Democracy* (Routledge, 2020).

How to cite this article: Fox, Carl. 2023. "Stability and Disruptive Speech." *Journal of Social Philosophy* 1–17. https://doi.org/10.1111/josp.12513