What's wrong with censorship?

Adam Lovett

Abstract. Regimes of censorship and propaganda suppress some messages and propagate others. The Chinese Communist Party, for instance, suppresses anti-government information and spreads pro-government messages. Such regimes wrong listeners. My goal in this paper is to explain how. I claim that they infringe on listeners' autonomy by manipulating them. The core idea is that censorship and propaganda both work by misleading people about others' attitudes. They lead us to overestimate support for certain political arrangements and to misestimate the prevalence of certain beliefs among other people. Our false beliefs about what other people think lead to other downstream beliefs: about whether protests will be successful or simply whether the government is doing a good job. Inducing inappropriate attitudes in people and then exploiting those attitudes is a form of autonomy-undermining manipulation. So censorship and propaganda are wrong because they violate listener autonomy.

Keywords: Free speech; censorship; propaganda; manipulation; deception; autonomy; higher-order evidence.

Author Affiliation: Durham University, School of Government and International Affairs.

Introduction

In 1616, the Catholic Church put Copernicus's *De Revolutionibus Orbium Coelestium* on its Index of Prohibited Books. This forbade its sale, distribution, and possession. For the Church, heliocentrism was fine as a computational device but it contradicted biblical authority as a physical claim. So the Church planned to suppress the idea that the earth genuinely orbited the sun. Such censorship is a core component of many regimes that try to mold people's beliefs. But such regimes almost always involve much more than the suppression of information. The Catholic Church had a positive doctrine it expounded weekly from the pulpit. Its missionaries spread the word of God. It is from here we get the term 'propaganda': the Church didn't just suppress, but also spread, information. The modern face of propaganda is better exemplified by the '50 Cent Army' than the missionary. The Chinese Communist Party (CCP) employs up to two million people to comment online on current affairs.[1] They cheerlead for the CCP, frame events in its favor, stoke Chinese nationalism. The goal is to mold what people believe through both suppressing and spreading information.

I think that regimes of censorship and propaganda often wrong us. My goal in this paper is to help explain how. Let me be more specific. One reason these

[1] For this number, see Fedasiuk (2021) and Lau (2016). Obviously, the number should be taken with a pinch of salt: China doesn't publish exact figures.

regimes are wrongful concerns their effects on speakers. This is clear with censorship. When the laws punish speech, they violate a speaker's autonomy rights, their right to self-disclosure, or their democratic rights.[2] But this speaker-based point doesn't exhaust our objection to censorship and propaganda. Such regimes also wrong listeners. Both of the above cases bring this out. Copernicus had been dead for almost seventy-five years when the Church banned *De Revolutionibus*. The prohibition primarily wronged potential readers of the book rather than the writer.[3] The 50 Cent Army does not stop anyone from speaking. It doesn't even really insult or argue with people online. It's a cheerleading squad (Roberts 2018, 210–15). Still, by distorting the information environment with paid commentators, the CCP wrongs its citizens. Both the Church and the CCP wrong people in their capacity as listeners as well as speakers. My goal is to elucidate how censorship and propaganda wrong listeners.[4]

I'll get right to the point. My view is that when censorship and propaganda distinctively wrong listeners, they do so by manipulating them. They induce false beliefs in them and then exploit those beliefs. Specifically, censorship and propaganda mislead their victims about other people's attitudes. They trick people

---

[2] For these views, see, respectively, Baker (1989, 2011), both Brown (2023, 37–40) and Shiffrin (2014, 88–89), and both Post (2011) and Weinstein (2011).
[3] Howard (2024, Sec. 2.1) makes the same point.
[4] For other listener-based theories, see Meiklejohn (1948), Mill (1859, ch.2), T. Scanlon (1972), and Strauss (1991). We will discuss these later. One way to see my view is as a development of Strauss's.

about what others support: the 50 Cent Army fools Chinese people into overestimating how many others support the CCP. And they trick people about what others believe. The Church's suppression of heresy fooled adherents into thinking that more people believed in, say, geocentrism than did. This affects people's other actions and attitudes. If we think other people support the government, we are less likely to protest it. If we think other people believe in geocentrism, we are more likely to believe in it. Now here's the normative nub. Getting people to think or do things via inducing inapt attitudes in them, including false beliefs, is a way of manipulating them.[5] Such manipulation violates people's autonomy. So censorship and propaganda violate the autonomy of listeners.

Here is the plan for the paper. In the next section, I explore in more detail how censorship and propaganda affect our beliefs. In the second section, I look at some existing accounts of what is wrong with such regimes. I will argue that these don't fully explain the phenomena we want to explain. I will then, in the third section, explore how to think about manipulation. I will argue that autonomy-infringing manipulation involves inducing inappropriate attitudes. In the fourth section, we apply this theory to show what's wrong with censorship and propaganda. Finally, in the fifth section, we evaluate some other mechanisms by

[5] For an early version of this kind of idea see Noggle (1996). Something like it is also endorsed by Barnhill (2014), Klenk (2022), Lovett (2021), Mills (2014), Noggle (2025), and Renzo (2025a, 2025b). In other words, the idea is widely endorsed in the literature on manipulation (although not universally; see, e.g., Gibert (2023)).

which censorship and propaganda may work. I don't, to reiterate, think that the only objection to censorship and propaganda is that they manipulate their audience. But I believe this is a core objection to regimes of censorship and propaganda. My goal in this paper is to spell out that objection.

How Does Censorship Work?

Let's start with definitions. C*ensorship*, as I'll define it, consists in preventing people from encountering certain claims. *Propaganda* consists in introducing people to particular claims. I will often talk of *regimes* of censorship and propaganda because they are usually engaged in together.[6] People rarely suppress some message unless they have their own message to spread. Additionally, it seems to me that we often have the same objection to both censorship and propaganda. They seem wrong for the same reason. If so, we should treat them together. Let me make two further points about these definitions. First, these are stipulations rather than conceptual analyses of English-language terms. The aim is just to fix a target we can evaluate, not to truly capture what the English words mean.[7] Second, both identify successful attempts to suppress some information or spread a message. One can try to do such things and fail. When such attempts are wrongful, I suspect that

[6] In some cases, they are indistinguishable. In East Germany, for example, book production was wholly planned: censors decided both what was produced and what wasn't in the same process. For a good account, see Darnton (2015).
[7] 'Propaganda', for example, is pejorative in a way the definition above clearly doesn't capture. See e.g. Bonard et al. (2024).

their wrongfulness derives from the wrongfulness of the successes. It is wrong to attempt a wrong. My focus then is on what's wrong with successful censorship and propaganda. Now, of course, not all instances of censorship and propaganda (expansively defined) wrong listeners. My question is why those that do so are wrongful. My answer hinges on how they work. So let us turn to an account of how censorship and propaganda mold our attitudes.

Beliefs about Attitudes

My account depends on a claim about human cognition. The claim is that our beliefs about other people's attitudes have a big impact on what we ourselves do and think. There are two ways in which beliefs about others' attitudes matter. The first arises in the special context of co-ordination problems. Often, we will only do something if we think others are willing to act complementarily. I will only cross the road if I think drivers are willing to stop. I will only push for some political outcome if I think others are willing to do so too. The issue is especially acute when it comes to sanctionable political participation. If I protest an authoritarian government and not many other people do, I will be fired, imprisoned, sometimes killed. I will certainly not succeed in overthrowing the government. So if I believe that other people won't join my protest, I won't protest the government. My belief about whether other people will protest

crucially affects what I do.[8] The same is true for other people. Our beliefs about others' practical attitudes systematically affect our own actions.

The second way in which my beliefs about others' attitudes matter concerns my own beliefs. It is impossible for human beings to come to remotely accurate beliefs about the world from perception and reason alone. I could not establish that the earth orbits the sun or that Paris is the capital of France from merely looking out of my window and applying the probabilistic calculus. I need to hear from other people. Most of our knowledge of the world comes from listening to what other people think. We learn from parents, peers, teachers, colleagues, from television, films, books, and much else besides. Accordingly, what we think other people believe substantially influences what we ourselves believe. In my view, this process is entirely rational. It is rational to accept testimony from others; to conform what we believe to what others believe.[9]

It is useful to survey some evidence for the powerful influence of others' attitudes on our own beliefs. One class of evidence comes from the classic Asch conformity experiments (1956). In these experiments, a subject is put in a room with several of the experimenter's confederates and shown a line. They are asked to judge which of three further lines is the same length as it. The experimenter's

[8] For a good account of how this works see Kurtulmus (2025). For a classic articulation of the underlying idea see Kuran (1989, 1995).
[9] For a general overview of the literature on the rationality of such adjustment, see Frances and Matheson (2024).

confederates all announce the wrong answer. In about a third of the trials, the subject announces the wrong answer too. Subjects do repeated trials; three-quarters announce a wrong answer at some point.[10] Subjects in such experiments seem to genuinely shift their beliefs. One of Asch's subjects, for example, explained their conformity by saying that '140 million Frenchmen can't be wrong' (1956, 28). Subjects also report wrong answers when written privately, insulated from direct peer pressure (1956, 56–58).[11] Seemingly, we sometimes give more weight to what other people believe than our own perception.

Judge-advisor experiments provide a second class of evidence. In these experiments, a subject (a 'judge') gives a numerical estimate of some value, such as the distance between two cities or how many coins are in a jar. They then see an estimate from someone else (the 'advisor') and are asked if they want to change their own estimate. A meta-analysis of 346 effect sizes from such experiments shows that, on average, people bring their estimate around 40% closer to the advisor's (Bailey et al. 2023, 24533). People conform their credences to those of the advisor. A third class of relevant studies looks at how we respond to presentations of aggregated opinions, such as opinion polling or online reviews. Aggregated opinions often substantially impact our own judgments (Oktar et al.

[10] For an almost exact, recent replication of these results see Franzen and Mader (2023). For a systematic review of the recent literature see Capuano and Chekroun (2024).
[11] In other words, Asch's results are partly explained by what Deutsch and Gerard (1955) called 'informational social influence'.

2024). One striking study concerns conspiratorial claims (e.g. 'The U.S. government planned the 9/11 attack on the World Trade Center'). Subjects were asked both how likely these claims were to be true and how many other people believed them. They were then told that belief in the claims was 40% more prevalent than the subject had predicted. This increased subjects' confidence in the relevant claims by about a third (Orticio et al. 2022). It seems we often take what other people believe as evidence. We often conform our beliefs to those of others.

Censorship and Propaganda

My conjecture is that a core mechanism through which censorship and propaganda work is via affecting what people believe about others' attitudes. This affects what we ourselves think and do via the two routes just outlined. In this section, I want to explore a range of cases in which this conjecture is plausible.

Let's start with the state. China, to take an important example, uses both censorship and propaganda to create the appearance of widespread support for the government.[12] When it comes to censorship, it blocks internet access to material critical of the government, takes down social media posts containing such

[12] I think there *is* widespread support in China for the government (Edelman Trust Institute 2025). So my point is not that the CCP is creating a total fiction: only that it is leading people to overestimate even high levels of support.

material and throttles its distribution (Roberts 2018, 113–89). As per propaganda, it instructs media companies to write copy supportive of the government (Roberts 2018, 201–9) and employs an army of commentators to cheerlead China and the CCP on social media (Griffiths 2021, 212–15; Roberts 2018, 210–15). The aim of both is, in part, to lead people to overestimate how many others support the government.[13]

How does this affect what people think and do? Via the two mechanisms identified above. To begin with, getting people to overestimate others' support for the government disrupts anti-government coordination. People are less likely to protest the government if they think that other people support it, because protests are only effective if others join in. Haifeng Huang and Nicholas Cruz (2022) suggest that this happens because people think others are especially susceptible to propaganda. They find experimental evidence that, when contemporary Chinese people think that others are exposed to propaganda, that reduces their own willingness to protest. Yating Pan and Zhan Shu (2022) find that seeing uniformly pro-regime online comments on an article makes expressions of dissent very unlikely. They suggest people think others support the regime, and this dissuades them from expressing opposition. There is good

[13] For an argument that this is the aim, see Masterson (2025).

evidence that Chinese censorship and propaganda reduce each individual's disposition to dissent by affecting their beliefs about others.

Additionally, getting people to overestimate how many other people support the government likely affects their own attitudes towards it. We have fairly direct evidence of this happening in Russia. Noah Buckley and his co-authors find that inducing Russians to think support for Putin is lower reduces their own support for him (Buckley et al. 2024). Their study exploits the fact that, after an unpopular pension reform, polled support for Putin dropped from 80% in 2018 to about 66% in 2020-21. To some respondents, they frame this as Putin having the lowest support for many years. To others, they frame it as a strong majority of Russians supporting Putin. They find that support for Putin among the former was 6-11 percentage points lower than that among the latter. Outside authoritarian regimes, Sirianne Dahlum and her co-authors find similar evidence in a Norwegian survey experiment: they find that when people are told that support for an authoritarian strongman is higher among others, they report higher support for strongmen themselves (Dahlum et al. 2026). Affecting what we believe about others' political beliefs seems to affect our first-order beliefs.

More direct evidence in the Chinese context is a little scant. There is evidence that exposure to Chinese propaganda does change recipients' attitudes. One experimental study suggests that showing people a replica propaganda newsreel

affected their beliefs (J. Pan et al. 2022). A different case concerns Chinese propaganda around the Qingming festival. This is a festival for honoring the dead. Inconveniently for the CCP, it provides a time at which the tombs of those killed by the CCP (e.g. in 1989) are salient. To address this, the CCP has used both formal media and the 50 Cent Party to reframe the festival. The goal is to have journalists and internet commentators link Qingming to the recognition of Communist 'martyrs': soldiers who died during the revolution. Studies find success: even in international media, people increasingly talk about Qingming in the 'martyr' framing (Roberts 2018, 218–21). These studies don't definitively show causation via beliefs about what other people believe in China. But, in light of the psychological evidence discussed in the pervious subsection, that hypothesis is a good explanation of why such propaganda is effective.

So, plausibly, state censorship and propaganda often work by affecting what we believe about others' attitudes. The same is true for non-state censorship and propaganda. Consider religious censorship. The Church developed the Index of Prohibited Books to address heresy.[14] The critical Tridentine Index, published in 1564, forbade works written by heretical leaders – people like Luther, Calvin, Zwingli. The non-religious work of other heretics was permitted only if approved by Catholic theologians. Vernacular bibles were forbidden unless permitted by a

[14] For a systematic account of how Church censorship worked in Venice, see Grendler (1977).

bishop or inquisitor. This was the stuff of counter-reformation: the largely successful project of stemming the tide of Protestantism.[15] Partly, I conjecture, all this worked via the second mechanism above. We treat what other people believe about the world as evidence about how the world is. The more people encountered those with heretical beliefs, the more evidence they had against doctrine. By prohibiting books expressing such differing beliefs, Rome could minimize such encounters. So it could stymie heresy. If this conjecture is true, then Church censorship worked by inducing false beliefs about others' beliefs.

How plausible is this conjecture? We do not have systematic data on belief formation from the 1500s, so our evidence has to be indirect. Partly, that evidence comes from the picture of human cognition just discussed. But we also have some more specific evidence about religious belief. Adolescents, for instance, tend to adopt the level of religiosity of their friends (Cheadle and Schwadel 2012). Several scholars have found evidence that religious belief is promoted by 'credibility enhancing displays': if an observer sees others engaging in costly religious behavior that leads them to infer others genuinely hold religious beliefs. That makes the observer more religious themselves (Gervais and Najle 2015; Henrich 2009; Ishii and Watanabe 2024). This is evidence that we conform our religious

[15] Famously, this also involves the creation of the *Sacra Congregatio de Propaganda Fide*, from which our term 'propaganda' derives. Censorship and propaganda went hand-in-hand in the Early Modern Church.

beliefs to those we think other people hold. Thus, if the Church did suppress expressions of Protestantism, this would likely fortify Catholics' beliefs.

Neither CCP nor Catholic censorship had very sympathetic goals. But regimes with much more appealing ends can employ the same means. Consider hate speech laws. These exist in almost every developed democracy. They prevent, in part, the 'stir[ring] up' or 'promot[ing]' of hatred.[16] How do they do that? According to Jeremy Waldron, a key factor is that 'the publication of hate speech…is a way of providing a focal point for the proliferation and coordination of the attitudes that [it expresses]' (Waldron 2012, 95). Waldron suggests that 'the whole point' of hate speech laws is to 'drive hate underground' (2012, 95). His idea is that public bigotry allows bigots to realize that others share their attitudes. Laws banning public bigotry stop that realization. That stops bigots coordinating, for example by forming political parties. So Waldron recommends hate speech laws in part as a way of misleading people about what others believe.[17] Hate speech laws have much more appealing goals than the Counterreformation Church or the CCP. But, according to Waldron, they affect what people do in much the same way that CCP propaganda suppresses protests. They disrupt co-ordination by causing bigots to underestimate the prevalence of bigotry.

---

[16] The language is, respectively, from Part 3 of the UK's Public Order Act 1986 and Section 319(2) of Canada's Criminal Code.

[17] This isn't the only part of his view. He also thinks hate speech undermines its victims' sense of assurance of being treated justly; hate speech laws stop this undermining (Waldron 2012, 81–89).

One might resist Waldron's suggestion. Perhaps hate speech laws simply deter certain ways of expressing bigoted ideas. They might deter slurs or vituperation without preventing people seeing how prevalent hateful attitudes are. This seems possible in principle, but doubtful in practice. If one deters certain expressions of ideas, people will encounter those ideas less often in their everyday life. That will affect how prevalent people believe them to be. And anyway hate speech laws often punish a wide range of non-vituperative speech. Many people have been convicted for dispassionate expressions of bigoted views (Weinstein 2017, 552–61). So I suspect hate speech laws affect our beliefs about other people's attitudes. An alternative way to resist this story is to claim that such laws don't foment false beliefs but rather prevent us forming beliefs about what others believe altogether. This also seems to me unlikely. We have dispositions that correspond to beliefs about the prevalence of bigoted attitudes. We will, for example, say what we think others believe if asked. We rarely simply plead ignorance. We will predict how strangers will respond to racist comments. These dispositions suggest we have a belief about the prevalence of racist attitudes.[18] So I suspect that hate speech laws do lead bigots to underestimate the prevalence of bigotry.

[18] In any case, I suspect it is inappropriate to wholly suspend judgement on issues of practical importance on which one has good evidence. So we could still run the argument that inducing inapt attitudes is manipulative, on the hypothesis that such suspensions are inapt.

Let's turn to an example of non-governmental censorship in the service of desirable ends. Consider how social media companies suppress misinformation. Take, for instance, the claim that vaccines cause autism. By 2019, Facebook was suppressing the distribution of such claims. It partly did so by downranking pages that contained them (Meta 2019). Such algorithmic censorship means fewer people are introduced to the offending claim, because fewer people see posts expressing it. In December 2020, during the pandemic, Facebook also announced that it would take down posts containing 'false claims about the safety, efficacy, ingredients, or side-effects' of COVID-19 vaccines (Meta 2020).[19] Facebook had embarked on a systematic campaign to suppress the spread of vaccine misinformation. I conjecture that this affected people's beliefs through the same epistemic mechanism that banning heresy affects beliefs. When people see less vaccine skepticism online, they underestimate how many other people doubt the safety and efficacy of vaccines. This leads them to the true belief that vaccines are safe and efficacious. The goal here is desirable. But, so I conjecture, Facebook gets there by misleading people about what others believe.[20]

[19] Facebook, separately, removes bot accounts spreading misinformation. This plausibly doesn't mislead people about what others believe, since comments from bots don't express sincere beliefs. So I'll identify no objection such removals. On the contrary, this paper provides an argument for them: removing bots protects users from manipulation. Thanks to a reviewer for raising this point.
[20] For some arguments in favor of suppressing such misinformation, see Ferretti (2023) and Millar (2024).

How plausible is this conjecture? Well, several experimental studies have found that, when we show people online comments on an issue, that affects what they think others think about the issue (Lee 2012; Soffer and Gordoni 2020; Zerback and Töpfl 2022). We do use user-generated online content to gauge others' attitudes. A meta-analysis of 44 studies finds that seeing such online comments also drives our views about the underlying issue: we align our view with the comments we read (Chen and Xia 2024). One experiment even provides evidence that this alignment is mediated by our perception of others' beliefs. We are most responsive to online comments when seeing them changes what we think other people think (Lewandowsky et al. 2019). So, in some form or other, every link in the conjectured causal chain has been supported. That redounds to its plausibility: likely, online misinformation suppression works partly by affecting what we think others believe.

Let's look at one final case. Mill, famously, said 'social tyranny' was 'more formidable than many kinds of political oppression' (Mill 1859, ch.1). He was worried about people imposing social sanctions on others, and undermining their livelihood, if they expressed certain views. From the mid-2010s, the U.S. especially saw attempts by decentralized activists to vigorously exercise such social sanctions. Consider one recent example. In early 2024 Joseph Morais left a Facebook comment on an old photograph of Woonsocket, Providence, saying it was a '[g]reat city for low crime because it was white' (NBC 10 News 2024). Black

Lives Matter Rhode Island (BLM RI) lobbied on Twitter for him to be fired from his job due to his comment. He was fired. I conjecture that such social sanctions work in the same way Waldron takes hate speech laws to work. They often (although not necessarily always) discourage people posting racist claims on social media.[21] This, I conjecture, leads others to underestimate the prevalence of racist attitudes in society, which reduces organization around those attitudes and probably their prevalence itself. Again: worthy ends. Still, such social sanctions likely lead people to have false beliefs about other people's attitudes.

In sum, I've claimed that a core mechanism by which censorship and propaganda work is by misleading people about others' attitudes. These attitudes can either be practical or doxastic: they can concern what other people support or what they believe. Misleading people about others' practical attitudes disrupts coordination. Misleading them about others' doxastic attitudes affects their beliefs. The regimes that utilize these mechanisms are extremely diverse. They involve very different entities: governments, churches, companies, decentralized activists. The key point though is the commonality. All work, in part, by misleading people about other people's attitudes. As I discuss later, I don't think this is the only way censorship and propaganda can affect us. But it is, I suggest, a core mechanism. My aim is to explain how such a mechanism wrongs listeners.

---

[21] For some evidence that relatively light-touch social sanction can do this, see Munger (2017).

The Current Literature

Now we have a good grasp of some real-world regimes of censorship and propaganda, let's see what the existing scholarly literature says about their wrongfulness. As I've said, such regimes – and censorship especially – plausibly wrong speakers. This has been the subject of much discussion.[22] But that doesn't elucidate how they wrong listeners. The most famous listener-focused theory is probably the Millian (indeed Miltonian) idea that free-speech restrictions stop the truth from winning out.[23] The thought is that, when presented with both true and false views, most people are more likely to believe the true ones. Free-speech protections ensure that people are often presented with opposing views. So, they mean the proportion of truths we each believe steadily increases over time. Censorship wrongs us by making us less likely to believe truths in the long-run.

This is surely part of the argument against censorship, and perhaps even against propaganda. But I suspect it is not the whole of it. Many of the issues are well-known, so I'll be brief.[24] A central problem is simply that censorship sometimes leads people to have more true beliefs. Consider Facebook's attempt to suppress vaccine misinformation. Vaccines really don't cause autism; in this case the Millian

---

[22] See e.g. n.2.
[23] See Mill (1859, ch.2) and Milton (1644).
[24] For more see, for example, Alexander (2005, 128–30), Baker (1989, ch.1), Brietzke (1997), Howard (2024, Sec. 2.1), and Schauer (1982, ch.2).

idea is probably just wrong. Suppressing vaccine skepticism, probably, leads to more true beliefs. I suspect that in many expert domains most of us benefit from similar gatekeeping. More interestingly, perhaps, consider China's attempt to convince people that protests against the regime are futile. This attempt is successful, which means anyone living in Beijing who protests the central government will not be joined by their fellows. They will, likely, end up being arrested. The success of the regime ensures the truth of the ultimate beliefs it aims to induce. I think there is still a listener-based objection to these regimes. But they do not ultimately make people believe fewer truths. So the Millian story doesn't capture this objection.

A natural alternative idea is that the problem with censorship and propaganda is that they mean someone else controls what people believe. We can find shades of this view in Shiffrin, when she says that each person has an interest in making up their mind through processes that 'do not follow a trajectory…scripted by forces external to that person' (Shiffrin 2014, 87). The thought is, in part, that I am wronged when someone else controls what I believe. One can understand the notion of control here counterfactually: someone controls what I believe when, if they want me to believe something, I will believe it, and if they don't, I won't. The idea is that when my beliefs are under someone else's control, I am wronged. This idea is clearly too crude. Consider, say, the British Broadcasting Corporation (BBC). I think the BBC is a reliable reporter of the news. If it tells me that

something has happened, I will believe it. So the BBC controls what I believe in the counterfactual sense. But this doesn't mean the BBC wrongs me. Any reliable testifier can control what I believe in this sense. So the objection to censorship regimes cannot be that they put listeners under others' control. There is nothing objectionable about control *per se*.

Shiffrin is, of course, too subtle to endorse the crude idea. She suggests that controlling a person's beliefs is objectionable only when the forces behind it are 'distinct from the reasons and other features of the world to which [the believer] is responding *qua* thinker' (Shiffrin 2014, 87). I am not entirely sure what she means by this. But I suspect she means that controlling someone's beliefs is unobjectionable when one is ultimately getting them to believe truths – the features of the world they are, after all, responding to as thinkers. But this, as I've said, seems false. Censorship can, ultimately, get people to believe truths. To reiterate, Chinese citizens probably are genuinely unwilling to protest CCP rule. If the CCP gets people to believe this through censoring critical voices, or through paying people to post supportively online, then it only gets people to believe the truth. It does not control people's beliefs to get them to believe falsehoods. Yet, still, I think CCP censorship and propaganda wrong Chinese citizens.

Let's turn to a different kind of listener-focused view. Scanlon (1972) influentially claimed that when speech causes harm solely due to inducing beliefs, those harms

cannot justify restricting such speech. He himself later recognized this was implausible: it means the government cannot restrict, say, false advertising (Scanlon 1979, 532–33). Additionally, this listener-focused theory doesn't try to explain what's wrong with propaganda. But it seems to me that our objection to censorship regimes extends beyond an objection to speech restrictions. The listener-based objection to CCP censorship is of the same sort as the objection to CCP propaganda. Yet no explanation of why the state should not use certain harms to justify speech restrictions will tell us why hiring millions of cheerleading commentators wrongs listeners. To capture this, we need an account of what is wrong not just with restricting but with propagating messages.

This brings out a larger point. Discussion of free speech focuses entirely on censorship. If the wrongs of censorship and propaganda are akin, this is a severe shortcoming. So let us see what has been said about propaganda. Probably the most prominent concern about propaganda is that it bypasses a subject's rational will. Jason Stanley, for instance, suggests that much propaganda involves affecting people by 'either emotional or nonrational means' (Stanley 2015, 53).[25] But this fails to identify the wrong of core cases of propaganda. Suppose I see only pro-CCP comments online and conclude that the vast majority of people strongly

[25] This view is endorsed by Godber and Origgi (2023) and Quaranto and Stanley (2021) and criticized by Hyska (2023) and Leydon-Hardy (2026). Note most of the focus here is on the *definitional* issue of what propaganda is, rather than the normative issue of why it is wrong.

support CCP rule. This might be a false belief, but my rational processes needn't have been bypassed. Not, at least, if we take the (common) view that the relevant rational processes are how we reason from our evidence to conclusions.[26] Expressions of support for the government are evidence that the government enjoys widespread support. It can be rational to update my beliefs in light of this evidence. It can thus also be rational to think other people won't protest the government and even, so I've suggested, become more confident myself that the government is doing a good job. Wrongful propaganda needn't involve bypassing a rational will.

Let's return to the literature on censorship. David Strauss (1991) also provides a theory of essentially Scanlonian form. He wants to explain why the 'government may not suppress speech on the ground that it is too persuasive' (1991, 334). Now again this doesn't explain what is wrong with propaganda: Strauss, like Scanlon, is focused on vindicating First Amendment protections. But I think Strauss has the germs of a more explanatory theory. He says that speech restrictions are like 'lies that are told for influencing behavior' (Strauss 1991, 354). I think that is exactly right. But I think Strauss has no good explanation of why the analogy holds. To begin with, he says that both speech restrictions and lies 'interfere with a person's control over her own reasoning process' (Strauss 1991, 354). But, taken literally,

[26] This is the view in, for example, Broome (2013) and Kolodny (2007). Stanley also seems to endorse it, which is why he denies that lying is propaganda (Stanley 2015, 48).

that seems false. The standard view is that the reasoning process is the process by which we go from premises to conclusions (e.g. Broome 2013). Premises can be false – the product of lies or censorship – while one still has control over the process. So censorship might convince us of false premises without doing anything to impair our control over our reasoning process.[27]

Strauss later says that 'lying forces the victim to pursue the speaker's objectives instead of the victim's own' (Strauss 1991, 355). I think this is his more considered view: speech restrictions ensure that victims pursue the censor's aims rather than their own. But this seems to me incorrect. It is clearly false about lying. I can lie to someone in order to get them to conform with their own objectives. I might have a student who wants to learn the material but procrastinates. I can motivate them to study by lying about the material being on the test. Here, I get them to pursue their own objectives. Still, I think I've wronged them by lying. And it also seems false about censorship. Consider the Church restricting the distribution of *De Revolutionibus*. This impacted many good Catholics. Such Catholics may not have wanted to be exposed to heretical material. They thought such exposure risked damnation. The Church's prohibition served their end. Yet the Church's restriction, in my view, still violated their rights. So Strauss's more considered view

[27] A reviewer suggests that control over our reasoning processes might include control over the premises we believe. But this takes us back to the view I initially associated with Shiffrin. The BBC has control over the premises I believe, but it doesn't thereby wrong me.

about the connection between lying and censorship does not seem tenable. He has no good account of the sense in which censorship is like lying.

Still, as I've said, I think Strauss provides the germs of the correct answer to how censorship regimes violate audience rights. One can think of my goal in this paper as spelling out a satisfactory version of Strauss's point. To do that, we have to think about the nature of autonomy.

What Is Manipulation?

I think of autonomy in Razian terms. The idea is that there is something valuable about authoring our own lives.[28] It is valuable for you to be the one who decides where you live, who you marry or what career you pursue. Autonomy is multifaceted: it has both internal and external components. Internally, to be an autonomous self-author your actions have to be hooked up to what you value in appropriate ways. So akrasia, paradigmatically, can undermine autonomy by preventing my doing what my values overall favor. Externally, to act autonomously one has to at least be free of certain kinds of interpersonal influence. As Raz puts it, 'if a person is to be maker or author of his own life then…his choice must be free of coercion and manipulation by others, he must be independent' (Raz 1986, 372–73). Distinctive forms of interpersonal influence

[28] The idea is central in Raz (1986).

can undermine autonomy. In this section I explore how non-coercive influence can do this.

Deception is one kind of non-coercive, but autonomy-undermining, influence. It is the kind Strauss points us towards. Suppose, for example, that a doctor suggests you join a clinical trial for a new drug. You ask what the risks are, and the doctor says there are none. This is a lie. The drug is in fact very risky. Misled, you agree to take part in the trial. Here your choice is not autonomous. The doctor's lying to you undermines the autonomy of your decision. There are many further ways the doctor could deceive you without baldly lying. Imagine they left confidential documents in your room suggesting that the drug was risk-free. They intend that you read them and form the belief that there is no risk. Or suppose they know you think anything a doctor recommends is risk-free. If they omit telling you about the risks, with the intention that you form a false belief, they also undermine your autonomy. In all such cases, the doctor undermines your autonomy. Deception generally wrongs people by undermining autonomy.

One might think that, in this case, it is the ignorance alone that undermines autonomy. It doesn't matter how the ignorance is induced; mere lack of knowledge makes a choice non-autonomous. But this is not so. We can make autonomous choices without knowing much about the results of our options. Consider medical drug trials. When deciding whether to take an experimental

drug, you don't know whether the drug will cure you or harm you. In some cases, you can't even come to a sensible probability estimate of specific harms happening: your evidence is too ambiguous. Yet you can still make an autonomous choice. The point generalizes. When you're deciding whether to have a child, take a job, move country, or marry someone, you might be radically ignorant of the results of your choice. But the choice can still be fairly autonomous. So it is how the doctor influenced you, rather than your ignorance itself, which undermines autonomy. The problem is not just that you falsely believe the drug is safe. It is that the doctor intentionally caused you to have false beliefs.

Many forms of interpersonal influence besides deception undermine autonomy. Consider, for instance, guilt tripping. Suppose you are going on a trip to New York City. Your father doesn't want you to go because he will be lonely. Over the course of a weekend he insists that you should feel guilty about his loneliness. You feel guilty and you cancel the trip.[29] Here I think your father has manipulated you and thereby undermined your autonomy. Likewise, flattery can be manipulative. Suppose a colleague is trying to get you to cover one of their lectures. They want to go on vacation. They have no particular views about your ability as a lecturer. But they tell you that you are an extremely perceptive speaker, one whose insight their students would find invaluable. In your pride you agree to cover them. Here

[29] The case is adapted from Barnhill (2014, 67).

again you have been manipulated. I think that your autonomy has also been undermined. Deception is one of a broader class of autonomy-undermining kinds of interpersonal influence.[30]

What unites all these cases? As I said at the start, they are all cases in which interpersonal influence undermines autonomy. So, really, we want to know what kinds of interpersonal influence undermine autonomy in the way that deception and manipulation do. To answer this, we start with an observation. In all these cases, the influencer causes their subject to exhibit an inapt attitude. When your father guilt trips you, he causes you to be inappropriately guilty. You shouldn't feel guilty about him being a little lonely. When your colleague flatters you, he causes you to be inappropriately prideful. You shouldn't think that you are so much better than your colleagues. When the doctor lies to you, he causes you to believe that the drug trial is not risky. This belief is inappropriate because truth is a norm for belief: you shouldn't have false beliefs. So in each of these cases, an influencer influences attitudes and actions via inducing, and then exploiting, an inappropriate attitude. Inducing such attitudes is the heart of manipulation.

Let me say more about (in)appropriateness. The background idea here is that our attitudes have fittingness conditions. Sometimes pride or guilt is excessive;

[30] For a long list of other examples, see Noggle (2025, 13–36).

sometimes anger unwarranted; sometimes a belief inapt.[31] These fittingness conditions vary with the attitude. Guilt is fitting only in proportion to you genuinely having done something wrong. You do nothing wrong by taking a trip to New York City, and so your guilt is inappropriate. Pride, meanwhile, is fitting only in proportion to your abilities. Anger is fitting only in proportion to you having been seriously wronged. And so on. For doxastic attitudes – beliefs and credences – fittingness hinges partly on justification and partly on truth. A belief is unfitting if you have insufficient evidence for it. But simple falsity also makes one's beliefs less fitting. Inaccuracy makes one's credences less fitting. Beliefs aim at the truth. Shortfalls from the truth fall short of the norms of belief.[32] All these cases involve failures to satisfy attitudinal norms. Manipulation involves inducing in people unfitting attitudes in this sense.

In some of the above cases, the inappropriate attitude is induced intentionally. Is intentionality required for manipulation? Most writers think not (Barnhill 2014; Baron 2014; Manne 2014; Noggle 2025, 82–101). To see why, consider the flatterer. Suppose they are completely indifferent to whether their flattery makes you excessively prideful. This might be because they don't have any views about your lecturing ability: they've never seen you lecture. They might not have a view

[31] For a discussion of the general point, see D'Arms and Jacobson (2000). For a detailed exploration of relevant cases, see Noggle (2025, 120–39).
[32] Some think that beliefs aim at knowledge (Williamson 2000). Still, knowledge requires truth and so falsity will be an epistemic failure.

about what counts as reasonable pride and what counts as excessive pride. They don't care. Still, if they induce excessive pride in you due to their indifference, they seem to me to undermine your autonomy. Marcia Baron plausibly suggests that the best way to accommodate such cases is to appeal to recklessness (Baron 2014). The idea is that we should care about not inducing in people inappropriate attitudes. When we do induce in them such attitudes because we lack a sufficient level of such care, we act recklessly.[33] In this view, recklessness, in the form of an indifference to the appropriateness of other people's attitudes, is the *mens rea* of autonomy-infringing interpersonal influence.

To bring things together, our discussion supports the following principle:

> **Manipulation:** If person A affects person B's acts or attitudes by recklessly causing them to have inappropriate attitudes, then person A infringes person B's autonomy.

As the label suggests, I see the kind of interpersonal influence this principle identifies as 'manipulation': manipulation is inducing and then exploiting other people's inappropriate attitudes.[34] So this principle simply says that manipulation

---

[33] How much must we care about avoiding inducing inapt attitudes? I think: a substantial amount. But I don't have any crisp characterization of the threshold. Obviously, the more one thinks one should care the more actions one will regard as reckless.

[34] Noggle (1996) is an early version of this kind of principle as an account of manipulation. He has recently spelt it out in detail (Noggle 2025). For other views of manipulation of this general thought, see Barnhill (2014), Gorin (2014), Klenk (2022), Lovett (2021), Mills (2014), and Renzo (2025a, 2025b). See Gibert (2023) for perhaps the most compelling critique (to which Renzo 2025a responds – I think convincingly).

infringes autonomy. We have a pro tanto duty not to infringe people's autonomy. So we have a pro tanto duty not to manipulate them. In the next section, we will see how this furnishes an objection to censorship and propaganda.

Before this, let me make clear what this principle says. This principle identifies a pro tanto duty. A pro tanto duty is one that, absent countervailing considerations, yields an obligation.[35] Still, there can be countervailing considerations. If the only way to save a life is by manipulating someone, perhaps one should sometimes manipulate the person. Some instances of manipulation are all-things-considered justified. Accordingly, any objection to censorship and propaganda generated by **Manipulation** will also be pro tanto: it can be outweighed by countervailing considerations. Sometimes, as we've seen, censorship and propaganda have worthwhile goals. They may sometimes effectively achieve these goals. Whether an instance of such censorship or propaganda is all-things-considered justified will depend on how well it achieves worthwhile goals weighed against how manipulative it is. Little in general can be said about the relative weight of these different considerations: it depends on the case. But what can be said, I think, is that there is a weighty objection to many instances of censorship and propaganda. We now turn to laying out that objection.

---

[35] The notion is from Ross (1930).

Censorship as Manipulation

I've suggested that a core mechanism by which censorship and propaganda work is via misleading people about others' attitudes. They mislead people about what others support in order to disrupt their coordination. They mislead people about what others believe in order to mold their further beliefs. It is now straightforward to see how censorship and propaganda can wrong listeners. When you mislead people, you cause them to have false beliefs. You cause them to have inappropriate attitudes they otherwise would not have had. By **Manipulation**, that means you infringe their autonomy. Now, as I've said, some autonomy-infringements are all-things-considered justified. We can imprison people to stop them harming others. But there is a strong prima facie objection to autonomy infringements. This is the objection to censorship and propaganda. Succinctly, censorship and propaganda manipulate people by causing them to hold false beliefs about other people's attitudes. This infringes their autonomy.

Let us see how this works in some of our real-world cases. Start with China. When China suppresses information about, say, discontent with the government, it causes people to falsely believe that almost everyone else supports the government. This false belief affects both their actions and attitudes: it leads them to think that any anti-government activity is doomed to failure, and so it means they are much less likely to engage with such activity. The same point goes for

Chinese propaganda. When the 50 Cent Army links Qingming to communist martyrs in thousands of online posts, it leads people to (falsely) think that many others see Qingming through this frame. And that leads them to think this is a reasonable frame and one that they themselves should sometimes use. China induces downstream beliefs by causing people to believe falsehoods about others' attitudes. **Manipulation** neatly explains why all this is wrongful. China manipulates its citizens.

The concern extends to censorship regimes with much more sympathetic goals. Hate speech laws, at least according to Waldron, suppress bigotry by preventing bigots recognizing other bigots. They mislead bigots into underestimating how widely shared their attitudes are. I have suggested that decentralized campaigns of social sanction do the same thing. When BLM RI gets someone fired because they posted something racist on Facebook, it discourages other people from posting racist things. That means everyone underestimates the extent to which people hold explicitly racist attitudes. When Facebook suppresses the distribution of vaccine misinformation, it leads people to underestimate how worried others are about vaccine safety. That leads them to themselves be less worried about such safety. To reiterate, I think the ends of these censorship regimes are good. Racism and vaccine skepticism are bad. But both regimes plausibly work in part by inducing false beliefs about other people's attitudes. If **Manipulation** is correct, they thereby infringe people's autonomy.

Let me look at two ways to resist this conclusion. One response hinges on intentionality. Perhaps some censors don't intend to mislead people about what others believe. Consider, say, decentralized activists calling for the firing of racists. They may not intend to deceive others about the prevalence of racism. Their opprobrium may be entirely non-strategic. Does this exonerate them? In the previous section, I argued that intentionality wasn't necessary to undermine autonomy. So lack of intentionality does not get one off the moral hook. Instead, the key question is whether such censors act recklessly. To avoid acting recklessly, would-be censors have to give due weight to the aim of avoiding inducing in others inapt attitudes. Practically, that suggests they must take some measures to avoid misleading others about the prevalence of the attitudes they censor. This risk needs to temper their proposed sanctions. Perhaps they need to examine the actual prevalence of racist attitudes and communicate this prevalence amid their criticisms. Sometimes, decentralized activists do this; often, they do not. When they don't, I suspect that they act recklessly. So a lack of *mens rea* might exonerate some such censors, but not all of them.

Let's consider a second, more general, way to resist my argument. A salient feature of certain sympathetic instances of censorship and propaganda is that they improve people's attitudes overall. Facebook's misinformation suppression might cause would-be vaccine skeptics to have false beliefs about others' attitudes. But it also causes them to have true beliefs about vaccine safety and efficacy. This, one

might think, enhances their autonomy overall. After all, autonomy partly involves connecting one's actions to one's core values. By giving people more accurate beliefs overall we might engineer such a connection, even if we do so via inducing some false beliefs about others' attitudes. So, when censorship and propaganda enhance autonomy overall, is there still a weighty objection to them?

I think there is.[36] We can see this by considering cases in which manipulation enhances overall autonomy. Imagine, for example, that you suffer from severe, autonomy-degrading anxiety. The anxiety regularly prevents you from doing what you judge you ought to do. A doctor wants you to take a new drug. They (correctly, justifiably) believe that the drug will greatly alleviate your anxiety, enhancing your autonomy overall. But you are much too anxious to take a new drug. So the doctor lies to you; they tell you the drug is just a sugar-pill. You take it and your autonomy is enhanced. Here I think the doctor has seriously wronged you, despite overall enhancing your autonomy. Similar cases are easy to come by for non-deceptive manipulation. Suppose your father finds out you are going to New York City to join a church. He (correctly, justifiably) thinks that they will induce all kinds of false beliefs in you and exploit your irrationalities. This is why he lays on the guilt. Still, guilt-tripping you out of going to New York is pro tanto

[36] For a related discussion, see Noggle (2025, 173–75). Noggle's ultimate view is that manipulation that enhances autonomy overall is wrongful, but not autonomy-infringing.

wrong, even when it leaves you with more capacity for autonomous choice overall. It is wrong to manipulate someone, even when it overall enhances their autonomy.

To understand these cases, remember that the Razian notion of autonomy, from the prior section, is multifaceted. There are both internal and external components to autonomy. In these cases, degrading an external component of someone's autonomy improves an internal component. This may even, as hypothesized in these cases, lead to an improvement in the person's overall autonomy. The key normative question is whether this is objectionable: is it objectionable to infringe someone's autonomy in one respect in order to improve it overall? Our intuitions about the above cases suggest that it is. It is seriously objectionable to deceive, manipulate or coerce someone even if by so doing one makes them more capable of autonomous choice. Of course, in some cases it may be objectionable but all-things-considered permitted. Perhaps I am sometimes permitted to very mildly manipulate someone in order to enormously improve their autonomy long-term. But doing so is still pro tanto wrong. **Manipulation** is designed to capture this verdict. It says that any time we induce an inapt attitude and exploit it, we infringe autonomy rights, even if doing so enhances their autonomy overall. So even when censorship and propaganda enhance people's autonomy overall, they are objectionable. This is a weighty objection; our judgements about the above cases suggest that it is often seriously wrong to deceive and manipulate people.

I suspect, then, that when censorship and propaganda induce false beliefs about other people's attitudes, they often seriously wrong them by infringing their autonomy. This is a very general way in which censorship and propaganda wrong listeners. Still, that may not be the only way in which censorship and propaganda work. In the next section, we evaluate some alternative mechanisms.

Suppressing Evidence, Demonstrating Dominance

I've claimed censorship and propaganda often work by inducing false beliefs about other people's attitudes. Are there manipulation-based objections to other mechanisms by which they mold our beliefs? To answer that question, we need to identify some other potential such mechanisms.

Censorship and propaganda might sometimes mold our beliefs simply by suppressing evidence. If we come across less evidence for a claim, we will be less likely to believe it. The elegance of Copernicus's heliocentric model, for instance, is evidence that the earth does indeed orbit the sun. By preventing people encountering that evidence, Church censorship might make them less likely to believe in heliocentrism. Propaganda, meanwhile, can serve to distract people from evidence (Roberts 2018, 80–92). Consider the August 3, 2014, earthquake in Yunnan. The damage looked like evidence of shoddy construction, which would be evidence of government incompetence or corruption. On August 4, 2014, the CCP flooded the news with the sensational confession of gambling and

prostitution from a young internet celebrity (Roberts 2018, 191). The aim seemed to be to ensure people read about the celebrity confession rather than the earthquake damage; to stop them encountering this evidence of government shortcomings. Suppressing evidence in these ways might sometimes mold our beliefs without inducing false beliefs about others' attitudes.

I think, usually, there is a manipulation-based objection to censorship and propaganda even when they work solely by suppressing evidence. To begin with, such instances of censorship and propaganda often induce false beliefs. The Catholic Church induced false beliefs about astronomy. The CCP induces false beliefs about local government competency. One way to affect people's attitudes by inducing inapt beliefs is simply by inducing false beliefs about such issues. So these cases are covered by the limit instance of **Manipulation.** The more difficult cases are those in which censorship or propaganda increases overall accuracy. Think about Facebook suppressing reports of adverse responses to vaccinations. It might do this because people overreact to such reports and thus come to overestimate the risks of vaccination. So such suppression might lead people to have more accurate beliefs overall. Can we identify a manipulation-based objection to these cases too?

I am inclined to think so. Often, such instances of censorship and propaganda induce false beliefs about auxiliary issues that bear on the ultimate question. For

example, the rate of adverse responses to vaccines is evidence of their safety. Users might underestimate this rate if they never read about adverse responses. So Facebook's suppression may foster a false belief on an auxiliary issue. Equally, we often form beliefs about other people's evidence. We form beliefs about what experts' evidence supports or how well the best publicly accessible evidence supports certain claims. If we are blind to some of the evidence, these beliefs will likely be false. We will think, for example, that experts' evidence more strongly supports certain conclusions than it does (even when the conclusions are true). So censorship that works through suppressing evidence will often affect our ultimate attitudes via inducing in us some sort of inappropriate attitude. It needn't always do so. But in most cases, I suspect, **Manipulation** provides an objection even to censorship and propaganda that works wholly by suppressing evidence.

Let us turn to a final way censorship and propaganda can mold our attitudes. They may simply demonstrate the power of the government (Carter and Carter 2023). Getting people to repeat absurd claims might prove that the government has a lot of power. It can punish dissenters. This makes people less likely to dissent. When this is all censorship and propaganda do, I doubt there is a manipulation-based objection to them. If the government is powerful, demonstrating such power just induces true beliefs. The natural objection to this form of censorship and propaganda is instead an egalitarian one. I think subjecting people to power asymmetries is objectionable (Kolodny 2023; Pettit 1997). Plausibly, exercising or

exacerbating those power asymmetries is also objectionable. This kind of censorship and propaganda exercises and exacerbates power asymmetries. But this is not a distinctive objection to censorship or propaganda. Much of what authoritarian states do is subject to this kind of objection. So **Manipulation** may not explain the objection to censorship and propaganda in all instances, but it illuminates what makes them wrongful in a way more quotidian exercises of state power are not. It illuminates the distinctive way they wrong listeners.

Conclusion

Let me conclude. Censorship and propaganda wrong people not just in their capacity as speakers but also in their capacity as listeners. My goal has been to limn the nature of this wrong. I argue it is the wrong of manipulation. Censorship and propaganda often induce false beliefs about others' attitudes. This is a core mechanism by which they affect what we do and think. Affecting people's actions and attitudes via inducing inappropriate attitudes infringes on their autonomy. So censorship and propaganda wrong listeners by infringing on their autonomy. This helps explain why censorship and propaganda by authoritarian states, churches, companies and decentralized activists can be wrongful. When one intervenes in the information environment in a way likely to mislead people about others' attitudes, one infringes those people's autonomy.

Can entities intervene in our information environment without autonomy-infringements? Clearly, they can. Good public broadcasters do so. BBC journalists, for example, only report things for which they have good evidence. They do not recklessly mislead anyone, even about what others believe. Likewise, when statistical agencies carefully gather and publish demographic or economic data, they do not induce any inappropriate attitudes. Similarly, when Facebook labels vaccine skepticism as misinformation but doesn't downrank it, I doubt it misleads people about what others believe. Fact-checking isn't autonomy-infringing. Most generally, when communicators take due care to avoid inducing inappropriate attitudes, they do not infringe ordinary citizens' autonomy. So the lesson in this paper is not that all information operations are impermissible. It is that, when we mold people's beliefs, we should take care to convey solely truths.

Acknowledgements. I would like to thank three anonymous reviewers for this journal for helpful comments on the manuscript.

Financial support. This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

Competent interests. None to disclose.

References

Alexander L (2005) *Is There a Right of Freedom of Expression?* Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511614668.

Asch SE (1956) Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied* 70(9), 1–70. https://doi.org/10.1037/h0093718.

Bailey PE, Leon T, Ebner NC, Moustafa AA and Weidemann G (2023) A meta-analysis of the weight of advice in decision-making. *Current Psychology* 42(28), 24516–41. https://doi.org/10.1007/s12144-022-03573-2.

Baker CE (1989) *Human Liberty and Freedom of Speech*. New York: Oxford University Press.

Baker CE (2011) Autonomy and free speech. *Constitutional Commentary* 27(2), 251–82.

Barnhill A (2014) What is manipulation? In Coons C and Weber M (eds.), *Manipulation: Theory and Practice*. New York: Oxford University Press, pp. 51–72. https://doi.org/10.1093/acprof:oso/9780199338207.003.0003.

Baron M (2014) The mens rea and moral status of manipulation. In Coons C and Weber M (eds.), *Manipulation: Theory and Practice*. New York: Oxford

University Press, pp. 98–120. https://doi.org/10.1093/acprof:oso/9780199338207.003.0005.

Bonard C, Contesi F and Marques T (2024) The defectiveness of propaganda. *The Philosophical Quarterly* 74(4), 1080–1102. https://doi.org/10.1093/pq/pqae048.

Brietzke PH (1997) How and why the marketplace of ideas fails. *Valparaiso University Law Review* 31(3), 951–69.

Broome J (2013) *Rationality Through Reasoning*. Chichester: Wiley-Blackwell.

Brown É (2023) Free speech and the legal prohibition of fake news. *Social Theory and Practice* 49(1), 29–55. https://doi.org/10.5840/soctheorpract202333179.

Buckley N, Marquardt KL, Reuter OJ and Tertytchnaya K (2024) Endogenous popularity: How perceptions of support affect the popularity of authoritarian regimes. *American Political Science Review* 118(2), 1046–52. https://doi.org/10.1017/S0003055423000618.

Capuano C and Chekroun P (2024) A systematic review of research on conformity. *International Review of Social Psychology* 37(1), 13. https://doi.org/10.5334/irsp.874.

Carter EB and Carter BL (2023) *Propaganda in Autocracies: Institutions, Information, and the Politics of Belief*. Cambridge: Cambridge University Press.

Cheadle JE and Schwadel P (2012) The 'friendship dynamics of religion', or the 'religious dynamics of friendship'? A social network analysis of adolescents who attend small schools. *Social Science Research* 41(5), 1198–212. https://doi.org/10.1016/j.ssresearch.2012.03.014.

Chen J and Xia S (2024) Are online users influenced by what other users say? Meta-analyzing the cognitive, emotional, and behavioral impact of online comment valence. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace* 18(5), Article 8. https://doi.org/10.5817/CP2024-5-8.

Dahlum S, Hanson T, Johnsen Å, Kotsadam A and Wuttke A (2026) Is support for authoritarian rule contagious? Evidence from field and survey experiments. *American Journal of Political Science*. https://doi.org/10.1111/ajps.70067, accessed 23 July 2026.

D'Arms J and Jacobson D (2000) The moralistic fallacy: On the 'appropriateness' of emotions. *Philosophy and Phenomenological Research* 61(1), 65–90. https://doi.org/10.2307/2653403.

Darnton R (2015) *Censors at Work – How States Shaped Literature*. New York: W. W. Norton & Company.

Deutsch M and Gerard HB (1955) A study of normative and informational social influences upon individual judgment. *The Journal of Abnormal and Social Psychology* 51(3), 629–36. https://doi.org/10.1037/h0046408.

Edelman Trust Institute (2025) *2025 Edelman Trust Barometer: Global Report*. Available at: https://www.edelman.com/sites/g/files/aatuss191/files/2025-01/2025%20Edelman%20Trust%20Barometer%20Global%20Report_01.23.25.pdf, accessed 9 November 2026.

Fedasiuk R (2021) A different kind of army: The militarization of China's internet trolls. *China Brief* 21(7), 8–18.

Ferretti MP (2023) Fake news and the responsibilities of citizens. *Social Theory and Practice* 49(4), 629–51. https://doi.org/10.5840/soctheorpract2022726166.

Frances B and Matheson J (2024) Disagreement. In Zalta EN and Nodelman U (eds.), *The Stanford Encyclopedia of Philosophy*, Summer 2026 edn. Stanford, CA: Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2026/entries/disagreement/, accessed 27 July 2026.

Franzen A and Mader S (2023) The power of social influence: A replication and extension of the Asch experiment. *PLOS ONE* 18(11), e0294325. https://doi.org/10.1371/journal.pone.0294325.

Gervais WM and Najle MB (2015) Learned faith: The influences of evolved cultural learning mechanisms on belief in gods. *Psychology of Religion and Spirituality* 7(4), 327–35. https://doi.org/10.1037/rel0000044.

Gibert S (2023) The wrong of wrongful manipulation. *Philosophy & Public Affairs* 51(4), 333–72. https://doi.org/10.1111/papa.12247.

Godber A and Origgi G (2023) Telling propaganda from legitimate political persuasion. *Episteme* 20(3), 778–97. https://doi.org/10.1017/epi.2023.10.

Gorin M (2014) Towards a theory of interpersonal manipulation. In Coons C and Weber M (eds.), *Manipulation: Theory and Practice*. New York: Oxford University Press, pp. 73–97. https://doi.org/10.1093/acprof:oso/9780199338207.003.0004.

Grendler P (1977) *The Roman Inquisition and the Venetian Press, 1540–1605*. Princeton, NJ: Princeton University Press. https://press.princeton.edu/books/hardcover/9780691638539/the-roman-inquisition-and-the-venetian-press-1540-1605.

Griffiths J (2021) *The Great Firewall of China: How to Build and Control an Alternative Version of the Internet*. London: Zed Books.

Henrich J (2009) The evolution of costly displays, cooperation and religion: Credibility enhancing displays and their implications for cultural evolution. *Evolution and Human Behavior* 30(4), 244–60. https://doi.org/10.1016/j.evolhumbehav.2009.03.005.

Howard JW (2024) Freedom of speech. In Zalta EN and Nodelman U (eds.), *The Stanford Encyclopedia of Philosophy*, Spring 2024 edn. Stanford, CA:

Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2024/entries/freedom-speech/, accessed 3 November 2025.

Huang H and Cruz N (2022) Propaganda, presumed influence, and collective protest. *Political Behavior* 44(4), 1789–812. https://doi.org/10.1007/s11109-021-09683-0.

Hyska M (2023) Against irrationalism in the theory of propaganda. *Journal of the American Philosophical Association* 9(2), 303–17. https://doi.org/10.1017/apa.2022.4.

Ishii T and Watanabe K (2024) Cultural transmission and religious belief: An extended replication of Gervais and Najle (2015) using data from the International Social Survey Programme. *PLOS ONE* 19(6), e0305635. https://doi.org/10.1371/journal.pone.0305635.

Klenk M (2022) (Online) manipulation: Sometimes hidden, always careless. *Review of Social Economy* 80(1), 85–105. https://doi.org/10.1080/00346764.2021.1894350.

Kolodny N (2007) State or process requirements? *Mind* 116(462), 371–85. https://doi.org/10.1093/mind/fzm371.

Kolodny N (2023) *The Pecking Order: Social Hierarchy as a Philosophical Problem*. Cambridge, MA: Harvard University Press.

Kuran T (1989) Sparks and prairie fires: A theory of unanticipated political revolution. *Public Choice* 61(1), 41–74. https://doi.org/10.1007/BF00116762.

Kuran T (1995) *Private Truths, Public Lies: The Social Consequences of Preference Falsification*. Cambridge, MA: Harvard University Press.

Kurtulmus F (2025) Learning from the enemies of freedom: Freedom of expression and collective power. *British Journal of Political Science* 55, e5. https://doi.org/10.1017/S000712342400067X.

Lau J (2016) Who are the Chinese trolls of the '50 Cent Army'? *Voice of America*, 7 October. Available at: https://www.voanews.com/a/who-is-that-chinese-troll/3540663.html, accessed 6 November 2025.

Lee EJ (2012) That's not the way it is: How user-generated comments on the news affect perceived media bias. *Journal of Computer-Mediated Communication* 18(1), 32–45. https://doi.org/10.1111/j.1083-6101.2012.01597.x.

Lewandowsky S, Cook J, Fay N and Gignac GE (2019) Science by social media: Attitudes towards climate change are mediated by perceived social consensus. *Memory & Cognition* 47(8), 1445–56. https://doi.org/10.3758/s13421-019-00948-y.

Leydon-Hardy L (2026) Epistemic infringement and the wrong of propagandizing. *The Philosophical Quarterly* 76(1), 259–79. https://doi.org/10.1093/pq/pqae076.

Lovett A (2021) Democratic autonomy and the shortcomings of citizens. *Journal of Moral Philosophy* 18(4), 363–86. https://doi.org/10.1163/17455243-22220001.

Manne K (2014) Non-Machiavellian manipulation and the opacity of motive. In Coons C and Weber M (eds.), *Manipulation: Theory and Practice*. New York: Oxford University Press, pp. 221–46. https://doi.org/10.1093/acprof:oso/9780199338207.003.0011.

Masterson M (2025) Regime comments as ballot stuffing: How astroturfing can deter protest without persuasion or distraction. Unpublished manuscript, 30 August. https://michaelrmasterson.com/static/docs/masterson_astroturfing.pdf .

Meiklejohn A (1948) *Free Speech and Its Relation to Self-Government*. New York: Harper & Brothers.

Meta (2019) Combatting vaccine misinformation. *Meta Newsroom*, 7 March. Available at: https://about.fb.com/news/2019/03/combatting-vaccine-misinformation/, accessed 4 November 2025.

Meta (2020) Keeping people safe and informed about the coronavirus. *Meta Newsroom*, 3 December update. Available at: https://about.fb.com/news/2020/12/coronavirus/, accessed 4 November 2025.

Mill JS (1859) *On Liberty*. London: John W Parker and Son.

Millar B (2024) Epistemic obligations and free speech. *Analytic Philosophy* 65(2), 203–22. https://doi.org/10.1111/phib.12279.

Mills C (2014) Manipulation as an aesthetic flaw. In Coons C and Weber M (eds.), *Manipulation: Theory and Practice*. New York: Oxford University Press, pp. 135–50. https://doi.org/10.1093/acprof:oso/9780199338207.003.0007.

Milton J (1644) *Areopagitica*. Kingdom of England.

Munger K (2017) Tweetment effects on the tweeted: Experimentally reducing racist harassment. *Political Behavior* 39(3), 629–49. https://doi.org/10.1007/s11109-016-9373-5.

NBC 10 News (2024) BLM RI demands Pawtucket official's firing over alleged racist post. *WJAR*, 3 May. Available at: https://turnto10.com/news/local/blm-ri-demands-pawtucket-officials-firing-over-alleged-racist-post-southern-new-england-rhode-island-may-3-2024, accessed 4 November 2025.

Noggle R (1996) Manipulative actions: A conceptual and moral analysis. *American Philosophical Quarterly* 33(1), 43–55.

Noggle R (2025) *Manipulation: Its Nature, Mechanisms, and Moral Status*. Oxford: Oxford University Press.

Oktar K, Lombrozo T and Griffiths TL (2024) Learning from aggregated opinion. *Psychological Science* 35(9), 1010–24. https://doi.org/10.1177/09567976241251741.

Orticio E, Martí L and Kidd C (2022) Social prevalence is rationally integrated in belief updating. *Open Mind* 6, 77–87. https://doi.org/10.1162/opmi_a_00056.

Pan J, Shao Z and Xu Y (2022) How government-controlled media shifts policy attitudes through framing. *Political Science Research and Methods* 10(2), 317–32. https://doi.org/10.1017/psrm.2021.35.

Pan Y and Shu Z (2022) How can pro-regime online comments influence Chinese netizens' attitudes and opinion expressions? Normative or informational social influence? *International Journal of Public Opinion Research* 34(4), edac036. https://doi.org/10.1093/ijpor/edac036.

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

Post R (2011) Participatory democracy and free speech. *Virginia Law Review* 97(3), 477–89.

Quaranto A and Stanley J (2021) Propaganda. In Khoo J and Sterken RK (eds.), *The Routledge Handbook of Social and Political Philosophy of Language*. Abingdon: Routledge, pp. 125–46. http://ebookcentral.proquest.com/lib/durham/detail.action?docID=6529867.

Raz J (1986) *The Morality of Freedom*. Oxford: Oxford University Press.

Renzo M (2025a) Manipulation and practical agency. *Free & Equal: A Journal of Ethics and Public Affairs* 1(1), 108–38. https://doi.org/10.16995/fe.17722.

Renzo M (2025b) Why manipulation is wrong. *Political Philosophy* 2(1), 176–203. https://doi.org/10.16995/pp.17395.

Roberts ME (2018) *Censored: Distraction and Diversion Inside China's Great Firewall*. Princeton, NJ: Princeton University Press.

Ross WD (1930) *The Right and the Good*. Oxford: Clarendon Press.

Scanlon TM (1972) A theory of freedom of expression. *Philosophy & Public Affairs* 1(2), 204–26. https://www.jstor.org/stable/2264971.

Scanlon TM (1979) Freedom of expression and categories of expression. *University of Pittsburgh Law Review* 40(4), 519–50.

Schauer FF (1982) *Free Speech: A Philosophical Enquiry*. Cambridge: Cambridge University Press.

Shiffrin SV (2014) *Speech Matters: On Lying, Morality, and the Law*. Princeton, NJ: Princeton University Press.

Soffer O and Gordoni G (2020) The role of user comments in estimation of the public opinion climate and perceived support for one's opinion. *International Journal of Public Opinion Research* 32(3), 569–85. https://doi.org/10.1093/ijpor/edz036.

Stanley J (2015) *How Propaganda Works*. Princeton, NJ: Princeton University Press.

Strauss DA (1991) Persuasion, autonomy, and freedom of expression. *Columbia Law Review* 91(2), 334–71. https://doi.org/10.2307/1122761.

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

Weinstein J (2011) Participatory democracy as the central value of American free speech doctrine. *Virginia Law Review* 97(3), 491–514.

Weinstein J (2017) Hate speech bans, democracy, and political legitimacy. *Constitutional Commentary* 32(3), 527–83. https://doi.org/10.24926/26390742.1674.

Williamson T (2000) *Knowledge and Its Limits*. Oxford: Oxford University Press.

Zerback T and Töpfl F (2022) Forged examples as disinformation: The biasing effects of political astroturfing comments on public opinion perceptions and how to prevent them. *Political Psychology* 43(3), 399–418. https://doi.org/10.1111/pops.12767.