A Distortion or 'Our' Default?

Mikkola, M.

DOI
10.1093/arisup/akab004
Publication date
2021
Document Version
Final published version
Published in
Aristotelian Society supplementary volume

II—Mari Mikkola

A Distortion or 'Our' Default?

This paper considers Lynne Tirrell's analysis of toxic speech using examples epitomizing speech that is misleading, outright false, and without compelling justification. It is toxic in polluting and eroding democratic functioning. However, I argue that Tirrell's two epidemiological models (the common source model exemplified by poisons and the propagated transmission model that viruses exemplify) fail to make good sense of my examples, which are deeply insidious without being overtly invidious. The limitations of the epidemiological models suggest that toxicity is part of our default form of thinking and talking, rather than being an 'outside' pathology like a poison or a virus.

I

Introduction. In his 2015 examination of propaganda, *How Propaganda Works*, Jason Stanley writes:

> Speech that is silencing has the effect of restricting our free speech rights; this is why states that seem closer to embodying liberal democratic ideals, such as Canada and United States, have considerably less hate speech in the public domain than states further away from such ideals, such as Hungary. (Stanley 2015, p. 37)

Given the global events since, I was taken aback by this claim when recently reading Stanley's book. To me, Stanley seemed to be wrong about the United States more closely embodying liberal democratic ideals and containing considerably less hate speech than (say) Hungary. Or at the very least, his claim seemingly no longer obtains, and the change in the states of affairs has been both rapid and shocking.

Having said that, and perhaps with less pessimism than my initial reaction, Stanley may still be right if we employ a narrow legal conception of *hate speech*. An expression qualifies as hate speech (at least in the US legal system) if and only if:

> (a) it employs fighting words [like discriminatory epithets] or non-verbal symbols that insult or stigmatize persons on the basis of their [social kind membership]; (b) it is addressed to a captive audience [it is

Aristotelian Society Supplementary Volume xcv
doi: 10.1093/arisup/akab004

> suitably hard to avoid]; (c) the insult or stigma would be experienced by a reasonable person in those circumstances; and (d) it would be reasonable for the speaker to foresee that his words would have these effects. (Brink 2001, pp. 134–5)

Considering statistics from the legal realm, perhaps they do tell us that the United States contains considerably less of this sort of speech than states that are not as liberal and that are putatively less democratic. (But, of course, much more hate speech may go on than is recorded.)

Given Lynne Tirrell's understanding of toxic speech, however, this defence looks to be undermined. For her, toxic speech includes derogations, epithets and slurs. Standard examples include racist and sexist stereotypes. Toxic speech involves more though: it is

> a broad and mercurial category, encompassing speech that acutely and overtly harms, like racist epithets, but also speech that acts more chronically by gaslighting, undermining, threatening, and more. Some toxic speech surreptitiously reorients people away from their settled values and conceptions of the good. (Tirrell 2021, p. 000)

For Tirrell, toxic speech seemingly encompasses what we might more narrowly call 'hate speech' in the above sense, but it encompasses more diffuse kinds of oppressive and discriminatory speech too. My interest here (and elsewhere—see Mikkola 2020) is with a sort of toxic speech that is more amorphous and straightforwardly neither of the sort that Tirrell discusses nor of the kind that would fall under a narrow definition of 'hate speech'. It is akin to Tirrell's conception in that the sort of speech I am looking at certainly can (and does) undermine what is good for individuals. Given this characteristic, toxic speech in both Tirrell's and my sense is in some ways akin to propaganda, though I do not here wish to take issue with whether the sort of speech I have in mind counts as propaganda strictly speaking. Here are some examples:

- *The Brexit Bus*: Bright red Leave campaign bus stating 'We send the EU £350 million a week, let's fund our NHS [National Health Service] instead'.

- *Trump 'the Winner'*: Stating in an interview when facing electoral defeat in November 2020, 'This is a fraud on the American public. This is an embarrassment to our country. We were getting ready to win this election. Frankly, we did

Aristotelian Society Supplementary Volume XCV
doi: 10.1093/arisup/akab004

win this election. So our goal now is to ensure the integrity—for the good of this nation, this is a very big moment—this is a major fraud on our nation'. (The same sentiments of winning and the election being fraudulently stolen were, of course, expressed many times in the weeks to follow.)

- *Corona = Agenda* 21: During a 1992 UN Conference on Environment and Development, 177 national leaders (including George Bush Sr.) signed a non-binding statement of intent aiming to take action in order to ensure sustainability given population growth. This agreement, known as Agenda 21, has been dubbed by alt-right and political extremists as a secret plot to impose a totalitarian world order in a nefarious effort to use environmentalism as a means to crush freedom. In late 2020, groups protesting against restrictions brought on by the COVID-19 pandemic carrying signs stating 'Corona = Agenda 21' were seen (at least) in Germany, Switzerland, and the Netherlands. Their message is, in short, that the corona pandemic is used by global elites to annihilate people's freedoms and to reduce the world's population to advance the elite's iniquitous ends.[1]

These examples exemplify speech that is misleading, outright false, and without compelling justification, respectively. I take all of them to be toxic in that they pollute and erode our democratic functioning in a very material and real sense, for instance as was demonstrated by the US Capitol mob attack in January 2021.

With these sorts of examples in mind, I want to consider Tirrell's analysis of toxic speech with her two epidemiological models: the common source model exemplified by poisons, and the propagated transmission model that viruses epitomize. Even though I agree with Tirrell that some toxic speech can be so analysed, I argue in this paper that more diffuse forms of toxic speech work differently. The examples above are neither poisons nor viruses; this is because the epidemiological model suggests that in the absence of toxic speech qua poison or virus, its harms would be circumvented. I am less

[1] Whether the protesters think that the pandemic is a hoax used for these ends or that the real pandemic is being used for nefarious ends isn't entirely clear. But this does not make a substantive difference to my discussion of the example.

Aristotelian Society Supplementary Volume XCV
doi: 10.1093/arisup/akab004

convinced. Given the rapid rise and—importantly—*active endorsement* of toxic speech illustrated by my examples, the epidemiological model only tracks symptoms of (what I see as) the deeper disease. This is not to say that I find Tirrell's analysis unimportant or uncompelling. Quite to the contrary: the need to examine overtly hateful speech that causes toxic stress is both pressing and important. But I wish to explore how the epidemiological model fails to make good sense of the sorts of examples I note and which are deeply insidious without being overtly invidious. Seeing how and why some deeply problematic speech falls outside the epidemiological model is necessary in order to forge effective responses to such speech. Limitations of the epidemiological models, then, suggest to me that toxicity is our default form of thinking and talking, rather than being a distortion or an 'outside' pathology like a poison or a virus.

II

Toxic Speech: The Two Models. Let me start by outlining briefly the two epidemiological models. Tirrell employs epidemiological models to analyse the source and spread of toxic speech. First, there may be a *common source* of toxicity, as in the case of a water supply that is poisoned or contaminated by a pathogen. Second, there are networks of *transmission via propagation*, such as in the case of viruses spreading through a population. One of the examples Tirrell employs comes from Victor Klemperer's examination of speech employed by the Nazis: such speech was like adding millions of tiny doses of arsenic into the discursive realm, which ultimately polluted it entirely. Here we can see that the 'discursive poison' has a common source (the Nazi Party); but as this poison was repeated by ordinary Germans, small instances of the 'discursive poison' were transmitted via propagation—that is, the speech was contagious and behaved like a virus. As Tirrell puts this,

> Language can be contagious. Speech acts issue licences to others, and each licence issued and used participates in and may even reshape social norms. When others join the mode of discourse and add to the propagation of the speech, what may have begun as a common source transitions to a *propagated outbreak*. The discursive licence gets

Aristotelian Society Supplementary Volume XCV
doi: 10.1093/arisup/akab004

> passed along from speaker to hearer, again and again, each undertaking assertional commitments in using that speech and undertaking expressive commitment to that mode of discourse. This process strengthens the norm. (Tirrell 2021, p. 000)

Focusing on the connection between speech and health, Tirrell identifies toxic stress as a serious health hazard produced by toxic speech. Analysis of stress offered by epidemiologists and health psychologists is biopsychosocial. Positive stress is an 'ordinary' reaction to a challenging situation. Tolerable stress is a response to a challenging situation, like familial death. Social and emotional support systems help an individual to cope with the situation, hence making the damage caused tolerable. Toxic stress, however, 'overwhelms the immune system; it usually arises from traumatic events or prolonged adversity experienced *without mitigating interventions*' (Tirrell 2021, p. 000). Such stress can have serious health consequences and increase the likelihood of heart disease, diabetes, stroke and cancer, especially if toxic stress is experienced in childhood.

Prolonged toxic stress can come about through chronic exposure to discrimination, disparagement and exploitation, Tirrell notes. Some such discriminatory mechanisms are discursive:

> [T]he discursive delivery of identity-based aggressions—macro or micro—especially sexism, racism, xenophobia, anti-LGBTQ biases and others, are ongoing elements of allostatic load that demand attention or deflection. Those targeted are not only exhausted, but often damaged by the chronic strain. (Tirrell 2021, p. 000)

Although not an example Tirrell uses, being subject to stereotype threat has been linked to hypertension in a manner that fits Tirrell's analysis. Identity-based stereotypes can interfere with the performance of certain tasks by those negatively stereotyped relative to the task. Thus members of stigmatized groups may underperform because they are unconsciously distracted by worries about confirming the negatively valenced group stereotype. This phenomenon of *stereotype threat* is said to be a kind of self-evaluative threat or a self-stigmatizing anxiety that hampers performance (Steele 1997; Antony 2012). For instance, women tend to do worse in maths tests when they are primed for gender prior to the test (for example, when the examiner states that women usually perform worse in maths

Aristotelian Society Supplementary Volume XCV
doi: 10.1093/arisup/akab004

tests than men). When the test participants were not primed for gender, women's test performance improved significantly (Spencer, Steele and Quinn 1999). African American students have been found to perform worse than white students of comparable intellectual ability (based on the participants' prior standardized test scores) in tests that were explicitly noted to measure the students' intellectual abilities. When the students were told that the test investigates different problem-solving strategies and does not measure intellectual ability, students from different racialized groups performed equally well (Steele and Aronson 1995). Stereotype threat differs importantly from internalized racial or gender stereotypes. It is *situationally specific* and experienced when some negative stereotype applies. In threat-provoking situations, those subject to the relevant negative stereotype tend to exhibit distraction, narrowed attention, anxiety, self-consciousness, withdrawal or over-effort, elevated blood pressure and increased heart rate (Steele and Aronson 1995; Blascovich et al. 2001). Such factors conceivably contribute to the subjects' poor performance by cognitively draining and distracting them.

Blascovich et al. (2001) suggest that stereotype threat may in fact do more: it may be one of the major causes of high blood pressure amongst African Americans, who as a group have higher rates of hypertension than white Americans. Blascovich et al. argue that this gap cannot be explained genetically or by appealing to lifestyle differences; rather, due to racial discrimination, African Americans experience more chronic episodes of stress, and this plausibly explains differences in hypertension rates: chronic hypertension might be partly explained by the higher frequency of stereotype-threatening situations that African Americans encounter relative to white Americans. Such situations are characterized by acute blood pressure increases, and, according to Blascovich et al., repeated acute increases over time are thought to cause chronic hypertension. If this is right, stereotype threat causally linked to toxic speech that stigmatizes groups via negatively valenced stereotypes combined with the cumulative macro-/micro-aggressions that Tirrell notes seems to generate chronic stress that is toxic, where such stress can concretely and seriously damage health.

Aristotelian Society Supplementary Volume XCV
doi: 10.1093/arisup/akab004

III

Limits of the Epidemiological Model. As I noted above, I think there is much that is right in Tirrell's analysis of toxic speech. I am also in full agreement with her about toxic speech causally contributing to toxic stress that has genuinely damaging effects on our physical health. Having said that, I wish to focus on a variety of toxic speech slightly different from that which Tirrell focuses on. Much of the philosophical literature on speech and toxicity has concentrated on overtly vilifying and hateful speech. Tirrell's analysis also does so. My interest here is on speech that is much more covert, non-vilifying, and not explicitly invidious. I contend that even though certain kinds of toxic speech that Tirrell deals with do generate toxic stress and function like a poison or a virus, my examples outlined above work differently—and still they are deeply toxic to us individually and societally. To adequately deal with the examples I have in mind, we need different frameworks for understanding different types of toxic speech in order to forge effective remedies.

Passivity of Tirrell's Models. When we are dealing with poisons or viruses, our power to avoid and resist them is limited. We can take precautions, wear protective gear, avoid contaminated water sources, and so on. But we have little control over whether we get poisoned or infected when we come into contact with arsenic or COVID-19. This fits the definition of hate speech I offered above: speech that employs fighting words, is hard to avoid, and is intentionally insulting or stigmatizing. We can try to avoid situations where we might face hate speech, just as we might try to avoid coming into contact with a poison or a virus. But if we unwittingly do come into contact with hate speech, our power to avoid being targets of such toxic speech goes only so far. This is considered to be one of the central harms of hate speech: its toxicity prevents further efforts to deliberatively engage with the speaker (Brink 2001)—it cuts off democratic deliberation and exchange. Propagating such speech, then, can lead to an outbreak due to endorsement of the speech. As Tirrell puts it, 'Generally, when we speak, endorsement is automatic, not something added; it takes special care to restrict endorsement when we must. Widespread uptake—a million tiny repetitions—refreshes and licenses every time' (Tirrell 2021, p. 000). And even though its targets can try to avoid the source of toxic speech, this does not erase the toxin: 'Removing the pump handle doesn't clean

Aristotelian Society Supplementary Volume XCV
doi: 10.1093/arisup/akab004

up the poisoned well. Donald Trump's Twitter behaviour has poisoned a rather large well, so even once Twitter closed that account, the poison keeps spreading through the licences he has already issued' (2021, p. 000).

However, I contend, looking at the sorts of examples I gave above show us that contamination and contagiousness are more active than in the case of poisons and viruses. In the cases I am looking at, endorsement prima facie isn't something automatic. It may be the case generally when we speak (as Tirrell says), and certainly not challenging some speech in some cases amounts to licensing the speech in a manner compatible with and conducive to Tirrell's analysis. After all, even though she holds that licensing hinges on update, uptake can be both active and passive. As Rae Langton has recently claimed,

> We can extend Austin's account of uptake to include a *default* or *tacit* uptake, which does not require an active state of the hearer's mind ... Omissions, failures to block, even unwitting or oblivious ones, function as default uptake, allowing what a speaker does to go through. (Langton 2018, p. 156)

With my examples, I am not thinking about cases where uptake that licenses is of this passive sort. I am thinking about cases where speakers and hearers alike actively endorse the sorts of toxic speech examples I have outlined. In other words, I am thinking about cases where people are actively keen to endorse misleading, false and unjustified claims, even when such claims are vehemently and openly challenged in public discussions. This sort of actively keen endorsement may look like an automatic reaction to toxic speech, but I think that the situation isn't quite so straightforward. Endorsement may be immediate in that it happens 'on the spot', so to speak. But the sort of endorsement I am looking at isn't spontaneous in the sense of wholly lacking control and being without conscious thought or attention. Contra Tirrell, I am not concerned with toxic speech that targets understandably try to avoid (see Tirrell 2021, p. 000); I am thinking about the far too many contemporary instances of toxic speech that targets run toward with open arms. As noted before, this is more akin to how propaganda works. Stanley discusses how ideological beliefs that act as driving forces of propaganda are cherished beliefs: 'A cherished belief is one that *we will be reluctant to give up*.' (2015, p. 196) It is something that we want to retain, and so

Aristotelian Society Supplementary Volume XCV
doi: 10.1093/arisup/akab004

safeguarding that belief requires being on one's guard should any counterevidence threaten the belief. The same looks to be true with my example cases.

Take the Brexit Bus. It is misleading in not taking into account how much the UK gains from the EU each week (whether the UK in fact is a net beneficiary), and it misleadingly suggests that the £350 million *will* be used to fund the NHS in the case of Brexit. These (and other) misleading claims were publicly challenged, and so the Remain campaign did make a concerted effort to undercut any automatic endorsement and licensing of the Brexit Bus claims. Hence I contend that the Leave campaign's toxic messages were not like tiny doses of poison with a common source, or a virus that was transmitted via discursive licensing. Rather, the campaign was toxic because people *wanted* to believe plainly misleading claims. Endorsement was active and wilful—this is what undergirds licensing in the Brexit Bus example. In so being, the toxic message of the Brexit Bus safeguarded something cherished: an overblown and chauvinistic commitment in and to British Greatness, likely combined with unduly negative cherished beliefs about Europe and the EU. Moreover, in this case (as with other misleading claims) it seemingly takes effort and care to endorse the claims made in order to succumb to infelicities, falsehoods, partial explanations, and misleading assertions. It involves actively ignoring critics, experts, and those who point out how much Britain depends on and benefit from the EU in various ways. Endorsement also required actively ignoring challenges to negative depictions of Europe and the EU, which were misleading if not downright false. In this sense, I hold, endorsement isn't automatic in being devoid of control and conscious attention. Toxic speech may be immediate in that it 'strikes a chord' in hearers; but this is because of hearers' prior willingness to succumb to and embrace the speech.

This aspect of my examples of toxic speech isn't well captured by the epidemiological models. If we suspect that a well is poisoned, we rarely take active steps to ignore that or even embrace the poison. Of course, this may happen in instances of extreme scarcity where the poisoned well is our only water source; but those forced to drink water that they strongly suspect is contaminated are hardly embracing the poison. Now, one might think that Tirrell's discussion accommodates this when she writes, 'To the white supremacist, such speech may be nourishment, fuelling their sense of inherent superiority, and grounding their self-esteem' (2021, p. 000). I agree that this is what

Aristotelian Society Supplementary Volume XCV
doi: 10.1093/arisup/akab004

is happening in the case of a white supremacist. But the sort of active stance to toxic speech that I have in mind isn't overt in this sense. Rather, the active endorsement that I am looking at echoes Klemperer's claim, which Tirrell quotes, of how astonished he was about otherwise harmless individuals adapting to and adopting the Nazi regime's discursive changes with ease—how easily 'ordinary' citizens endorse and embrace the toxic messages without being (say) overt white supremacists. As Iris Marion Young notes, oppression refers to

> the vast and deep injustices some groups suffer as a consequence of often unconscious assumptions and reactions of well-meaning people in ordinary interactions, media and cultural stereotypes, and structural features of bureaucratic hierarchies and market mechanisms—in short, the normal processes of everyday life. (Young 1990, p. 41)

Tirrell's epidemiological models can capture toxicity of some type of speech; but, I contend, toxic speech that is part of 'our' normal processes of everyday life can remain untouched. In other words, being a white supremacist isn't (I hope) part of 'our' normal processes of everyday life. It is, rather, something extreme, and an ideologically driven lifestyle that typically diverges from 'mainstream' ways of life. My concern with speech that is part of 'our' everyday life is with the willingness of those we never thought would succumb to toxic speech endorsing such speech; seeing how people we know endorse toxic speech of my sort with ease—something that is and was unimaginable for many on the Remain side, for instance. This sort of 'ordinariness' that Young alludes to gives me great cause for concern, since we are clearly very prone to actively endorsing toxicity. Of course, one might say in response to my concerns that in examples like the Brexit Bus there are elements of both passive and active endorsement. Hence pointing out that endorsement is active does not *eo ipso* count against Tirrell's analysis. I am happy to accept that there may be and probably are mixed cases of toxic speech in this sense. But nonetheless, in the sorts of examples I am looking at, it is importantly the active part that seems to be doing the work. Endorsement does not happen to people; people make it happen. This is something that the epidemiological model cannot sufficiently capture.

Immunity to Toxic Stress. As I see it, those who endorse the toxic speech in my examples are not suffering from 'mere' ignorance of

Aristotelian Society Supplementary Volume XCV
doi: 10.1093/arisup/akab004

the facts. Moreover, the endorsement involved does not first and foremost strike me as being automatic. Instead, it seems to involve serious cognitive work in the form of active ignorance—ignorance that wilfully forecloses certain representational and conceptual realities. For instance, those who endorse (and I would say, embrace) 'COVID = Agenda 21'-type conspiracy theories are not ignorant in the sense of simply lacking some conceptual, representational and justificatory tools and resources. It is not that they are straightforwardly merely ignorant of some facts or states of affairs in the sense of just lacking relevant information. Rather, endorsement requires that one actively forecloses and supresses alternative (what we might call) 'lines of justification' and challenges. That endorsement seemingly requires active ignorance points to another limitation in the epidemiological models Tirrell presents. I agree that overt toxic speech that is racist, sexist, homophobic, ableist or transphobic causes toxic stress in its targets, and that this is a serious harm to be acknowledged. But different types of toxic speech affect their targets differently with some targets being immune to toxicity. Or to put the point differently and somewhat bluntly: the targets of toxic speech are not always 'the opponents', as with (say) racist hate speech. In my example cases, the messages are not directed at those who disagree in an effort to convince them otherwise. They rather target those who already are likely to believe. And this makes a difference to the harms generated.

Consider the aftermath of the Trump 'the Winner' example. Trump's supporters were targets of his plainly false allegations, and succumbed to and endorsed his seriously toxic speech with remarkable ease. They were clearly not the ones to suffer from toxic stress. The kind of fatigue that stress typically brings on was arguably suffered by non-supporters and opponents who witnessed with incredulity the ease with which Trump supporters actively ignored the falsity of his claims. But Trump's claims, it seems to me, were not primarily targeted at his opponents. So toxic speech of the kind that I am looking at infects and contaminates people differently. Of course, Tirrell notes that people respond to toxins differently, and some may suffer more than others. I agree, but I am not talking about cases where some are especially susceptible to a toxin and hence suffer more than others—this is something Tirrell's model can easily accommodate. My point is slightly different. Those infected by COVID-19 are all infected, irrespective of the severity of their

Aristotelian Society Supplementary Volume XCV
doi: 10.1093/arisup/akab004

subsequent symptoms; those who drink from the poisoned well are all poisoned, regardless of how sick this may make them. By contrast, Trump supporters who endorse his toxic speech look to be immune to infection or illness. It is not that they are harmed differently by toxic stress; rather, they are impervious to it. In other words, Tirrell holds that discursive epidemiology tells us how toxic speech can literally make us ill, just like a poison or a viral disease. Taking the epidemiological approach to speech seriously, then, helps us track 'how speech norms enact changes to health' (Tirrel 2021, p. 000). My point is precisely that this epidemiological approach does not track the toxicity of the sort of insidious but not invidious speech I am focusing on. This sort of toxic speech does not causally interact with health in Tirrell's manner for those who endorse it. In this sense, the proponents of my sort of toxic speech—both speakers and hearers—are immune (at least to toxic stress). This sort of immunity should give us deep cause for concern, though.

With this in mind, we might hold that even though those who endorse more diffuse, non-invidious toxic speech do not suffer from health-related harms like toxic stress, they do suffer from some other kinds of harm. Most importantly, it looks as though they suffer from some cognitive harms or pathologies; for instance, their epistemic agency must be corrupted in a manner similar to a computer virus corrupting a hard drive, or air pollution corroding a metal. Such harms are conceivably generated by toxic speech, and can sit alongside toxic stress. I agree that those who succumb to plainly misleading, false and unjustified toxic speech are in a sense harmed: their cognitive agency is perverted in a manner that does undermine what is good for the individuals. Having our cognitive agency and abilities undermined by forces of structural injustice looks to be something that always undermines human good owing to the source of the perversion. However, in order to understand this type of harm, and to account for why some allow themselves to succumb to toxic speech, we need a different framework to the epidemiological one. In short, the epidemiological model is too focused on the symptoms rather than the underlying illness. It allows us to explain how certain discursive toxins are spread and propagated; but it does not sufficiently address what enables such spread and propagation. To put the point differently: we can explain the outbreak of an infectious disease by realizing that the water source was contaminated by a pathogen dangerous to humans. We can further explain why we are susceptible to

Aristotelian Society Supplementary Volume XCV
doi: 10.1093/arisup/akab004

the pathogen by microbiological means; in other words, there is something about our microbiological constitution that explains the toxicity of certain pathogens, which then cause outbreaks. In the case of toxic speech, though, what is the underlying mechanism akin to our microbiological constitution that explains endorsement and willingness to succumb to toxicity? This is something that the epidemiological models leave open.

The answer to me seems to be our underlying cognitive architecture. I agree with Tirrell that discursive toxicity isn't a matter of evil or malicious intent per se on the part of the speakers, although such intent too can be present. Rather, as I see it, our cognitive architecture is like untreated steel: without protection, and if exposed to both oxygen and water, steel will rust. To prevent this, we treat and coat steel. When the coating is damaged, we repair and recoat it to prevent rusting. In a parallel fashion, without a protective coating our cognitive architecture will rust—or as I put it above, will be perverted. Basic educational efforts are hopefully aimed at creating that first layer of protection, so that our cognitive and epistemic agency can reach maturity. The examples of toxic speech I have looked at, though, suggest that this type of prevention is lacking. The willingness to endorse and to embrace half-truths, falsehoods and unjustified claims in the example cases is suggestive of an underlying problem at the core of epistemic agency: it has been left exposed to the elements, so to speak, without sufficient protections. The epidemiological models, however, only come in at a later stage—when we presume that the steel has already been coated, but it is damaged and hence begins to rust. My concern is that our underlying cognitive mechanism is what enables rusting, which is why I think that the epidemiological models target symptoms rather than the foundations of toxicity. In order to understand how and why toxic speech is so compelling, we need to look deeper than the epidemiological models do.

Ideal Picture of Language Basics. Above I claimed that seeing how some targets of toxic speech seem to be immune to the forms of toxic stress Tirrell identifies tells us that the epidemiological models are focusing on symptoms rather than root causes of toxicity—what it is about our microbiological or cognitive constitution that makes something toxic to us. If we notice that some people are immune to a virus while others are not, and wish to understand why, we do not just examine the virus—instead, we examine what is different about

Aristotelian Society Supplementary Volume XCV
doi: 10.1093/arisup/akab004

the microbiological constitutions of those who are immune and those who get infected. As I see it, this deeper analysis is needed to provide a fuller explanation of how toxic speech works. I am not able to undertake such an analysis here. However, I think there is something about Tirrell's methodology that lends itself to the epidemiological analysis, but which ultimately undermines a deeper analysis. This has to do with her 'language basics'.

Following Brandom, Tirrell holds:

> The basic tools of social-practice inferentialism are discursive *commitments* undertaken by speakers and *entitlements* issued to hearers. On this view, a speaker undertakes—and must be ready to defend—a complex set of commitments when they say something. Hearers are licensed to rely on what the speaker said, and defer justification for that back onto the speaker. This is the basic deontic structure of asserting. (Tirrell 2021, p. 000)

This picture of language basics, as I see it, is an idealized one, which isn't unproblematic for analyses of non-ideal toxic speech. Although Tirrell does not explicitly maintain this, her language basics look to function in a similar fashion to ideal theory in John Rawls's political philosophy. According to Rawls, the ideal theory of justice 'develops the conception of a perfectly just basic structure and the corresponding duties and obligations of persons' (1971, pp. 245–6). In being idealizations, the principles that Rawls develops and articulates in *A Theory of Justice* are meant to cover societies that are well-ordered: where 'everyone accepts and knows that the others accept the same principles of justice' and 'the basic social institutions generally satisfy and are generally known to satisfy these principles' (1971, p. 5). Moreover, Rawls assumes that these principles of justice will be strictly adhered to in well-ordered societies. In so doing, Rawls is methodologically wedded to (what I call in Mikkola 2021) 'the normative priority thesis': that we need to articulate an ideal theory of justice before we are able to tackle non-ideal circumstances. The thought is that we need to know what perfect justice consists of in order to examine injustices through its lens. Rawls writes, 'the ideal part presents a conception of a just society that we are to achieve if we can. Existing institutions are to be judged in the light of this conception and held to be unjust to the extent that they depart from it without sufficient reason' (1971, p. 246). Moreover, principles that define the perfectly just arrangement of basic

Aristotelian Society Supplementary Volume XCV
doi: 10.1093/arisup/akab004

institutions 'set up the aim to guide the course of social reform' (Rawls 1971, p. 245).

Tirrell's basic tools of language are reminiscent of such ideal theory in the realm of philosophy of language. Language involves a Brandomian game of giving and asking for reasons, and Lewisian language games, where 'the deontic score keeps track of the range of allowed next moves and who can make them. It imposes obligations on the speaker and it licenses uptake and responses by the hearer(s)' (Tirrell 2021, p. 000). Although focusing on language practices, which is taken to be an aspect of non-ideal philosophy of language by Jason Stanley and David Beaver in their forthcoming book *Hustle: The Politics of Language*, I hold that this focus is still too much associated with ideal conditions. Tirrell's language basics look like those that govern well-ordered language use and speech, where the toxicity of speech is to be assessed through the lens of idealized language use. But in line with doing non-ideal philosophical theorizing, I want to ask what happens to our analysis of toxic speech if we take toxicity as our starting point and as normatively prior. The idea here is similar to that in political philosophy. To put the point metaphorically: when starting with the ideal case, we (perhaps implicitly) assume that on day one, perfect justice reigned; then on day two, some irrational procedures and false beliefs distorted the well-ordered societal arrangements, resulting in injustices and non-ideal social conditions; the task now is to forge strategies that will restore that ideal situation and starting point. Some formulations of non-ideal theory (my own included) challenge this: what if there never were an ideal first day of justice to be recovered? If we think that the non-ideal situation is our default and starting point, we need different ways to conceptualize how to escape that situation. In this sense, non-ideal theory advocates that we start from deeply unjust non-ideal conditions when thinking about principles of justice—otherwise we end up formulating principles of justice that are ineffective to fight injustices in that our theorizing has misconceived the sources of injustice.

Perhaps owing to pessimism about recent events, I am rather more convinced that toxic speech is not a distortion of ideal speech situations; it is the norm and default. And so there is a need for non-ideal philosophy of language, which reverses the order of normative priority. To put the point somewhat differently and polemically: Tirrell's treatment of language basics looks to take civility,

Aristotelian Society Supplementary Volume XCV
doi: 10.1093/arisup/akab004

well-orderedness, discursive care, and supportive speech environments as the 'uncontaminated' foundation, where poisons and viruses can attack and distort that foundation in ways that generate toxic speech. I tend to think that we are first and foremost 'contaminated' by our cognitive architecture: toxicity is in the foundations of our ways of thinking and talking. Civility, well-orderedness, discursive care, and supportive speech environments are the distortions—but of course distortions that are desirable and to be striven for.

To make this clearer, think back to the analogy above about steel. The most basic 'kind' of steel is untreated and uncoated steel. To prevent it from rusting, we take steps to protect the steel. Protected steel isn't the foundation; so when it becomes damaged and is in need of repair, we are dealing with symptoms that give rise to possible rusting, and not the underlying causes of rusting. Starting from the level of already protected steel—starting from the more ideal situation—is where the epidemiological models enter the picture. My concerns above about 'our' willingness to endorse and embrace toxic speech focus on the non-ideal situation that is akin to uncoated steel: where civility, well-orderedness, discursive care, and supportive speech environments are lacking, in the sense of not yet even being on the scene.

Healthy eating offers another parallel to demonstrate my point. Given our evolutionary past, the default is to consume large amounts of calorific food. In some distant past when food was scarce and work for many required physical strain, we were largely unable to consume beyond our needs, and the 'toxicity' of our default nutritional desires remained dormant. However, this all has changed dramatically with easy access to worthless calories, sedentary lifestyles, lack of healthy and nutritious foods, and so on. The ideal conditions of having good and affordable access to healthy foods, exercise opportunities, and a supportive environment that fosters these conditions are correctives to our default non-ideal conditions when it comes to health and nutrition. But as these non-ideal conditions are our default, the ideal ones are the distortions. Bluntly put: left to our own devices and without public health measures, many of us would be living deeply unhealthy lives. The same is seemingly true of the sort of toxic speech that flourishes on Twitter, social media, and discussions boards, exemplified by the case of 'COVID = Agenda 21' (and depressingly, many others): this is what happens to our speech when we are left to our own devices. Civility, well-orderedness,

Aristotelian Society Supplementary Volume XCV
doi: 10.1093/arisup/akab004

discursive care, and a supportive speech environment are not the default, but rather something akin to discursive coating we need in order to protect ourselves from ourselves.

Now, one might wonder whether my view of toxicity being more the default than a distortion goes against what I said earlier about endorsement and automaticity. Contra Tirrell, I claimed that endorsement is not automatic, but active. The idea is that it takes effort to endorse the claims made in my example cases in order to succumb to infelicities, falsehoods, partial explanations and misleading assertions. Does this not, however, go against the view that non-ideal conditions ground the default deontic structure of language? If the active endorsement of the sort of toxic speech I am looking at hinges on our cognitive architecture, this lends itself to the idea that something automatic is grounding endorsement. After all, our cognitive processes are not always (or even often) immediately under our conscious control, and they are typically intransparent to us. So, prima facie, it looks as though there is a tension in my view.

I think that there isn't, though. Let me clarify. Think again about nutrition. We may be evolutionarily primed to consume calorific foods high in sugars and fat. But eating a diet that consists of such foods isn't automatic in the sense that we have no control over our diets. Endorsement in this case too takes effort: it involves following certain consumption patterns, buying certain foods instead of others, refusing certain foods (like fruit and vegetables) point blank, refusing to cook for oneself from scratch, and so on. Clearly, if the structural and material conditions are such that breaking this type of putative endorsement is difficult or impossible, there is much to be concerned about. But in this case, I contend, we are no longer speaking about endorsement; we are speaking about coercion. In a similar fashion, the examples of toxic speech I have used involve endorsement that is active: one seeks out materials one finds in the media or online that confirm some prior views one holds. If there is nothing but toxic speech, we are faced with a different pathology that is more akin to brainwashing than endorsement. However, as things stand, we are hopefully not there yet. Endorsing toxic speech of my sort, then, still involves an active component and a choice, despite ultimately being enabled by our all-too-easily-rustable cognitive architecture.

Aristotelian Society Supplementary Volume XCV
doi: 10.1093/arisup/akab004

IV

Final Remarks: Philosophy as Vaccine? In an editorial of *The Scotsman* in January 2021, philosophy was compared to a vaccine:

> [W]e are heading towards a new world in which philosophy and the ability to think logically become increasingly important. To use a current metaphor, we need to vaccinate ourselves against the virulent lies of people like Trump and the best way to do that is to teach the wisdom of Socrates and co to our children. (*The Scotsman* 2021)

The idea that there is an antidote to toxic speech fits Tirrell's epidemiological models nicely. If I am right, however, that there are widespread and deeply insidious forms of toxic speech which the epidemiological models fail to cover, such a vaccine will only go so far, and a single dose often won't do. Consider a parallel to flu shots: those in the risk groups need to get yearly flu jabs to protect themselves. We all need to periodically renew various other important vaccines, such as getting regular tetanus shots. This once more highlights how the epidemiological models, especially when thinking about viral outbreaks, focus on intermediate rather than ultimate causes of toxicity. We should, of course, also pay attention to the intermediate level—just because the flu jab only works for a limited amount of time is not a reason not to get the jab. In this sense, I wish to stress again that seeing how toxic speech works in Tirrell's sense is valuable and compelling. I too used to think of philosophy and the ability to think logically functioning as a sort of vaccine (or rather, a bullshit filter) that inoculates us against the sorts of examples I have focused on here. I no longer share my earlier optimism. Given how easily our cognitive abilities and faculties can be perverted—or given how rustable they are—different remedies are needed. I cannot here discuss in detail what those remedies are, though I think that both a more supportive speech environment (especially in social media) and some cognitive therapy, perhaps in the form of virtue epistemology, are needed. But I anticipate that thinking about speech and cognition non-ideally by taking our limitations and flaws as the starting point and default, rather than thinking of them as distortions, will figure in the solution.[2]

[2] I am grateful to Guy Longworth for comments on an earlier version of the paper.

Aristotelian Society Supplementary Volume XCV
doi: 10.1093/arisup/akab004

Department of Philosophy
University of Amsterdam
Oude Turfmarkt 141–147
Amsterdam 1012 *gc*
Netherlands
Email: m.mikkola@uva.nl

References

Antony, Louise 2012: 'Different Voices or Perfect Storm: Why Are There So Few Women in Philosophy?' *Journal of Social Philosophy*, 43(3), pp. 227–55.

Blascovich, Jim, Steven J. Spencer, Diane Quinn, and Claude Steele 2001: 'African Americans and High Blood Pressure: The Role of Stereotype Threat'. *Psychological Science*, 12(3), pp. 225–9.

Brink, David O. 2001: 'Millian Principles, Freedom of Expression, and Hate Speech'. *Legal Theory*, 7(2), pp. 119–57.

Langton, Rae 2018: 'Blocking as Counter-Speech'. In Daniel Fogal, Daniel W. Harris, and Matt Moss (eds.), *New Work on Speech Acts*, pp. 144–62. Oxford: Oxford University Press.

Mikkola, Mari 2020: 'Self-Trust and Discriminatory Speech'. In Katherine Dormandy (ed.), *Trust in Epistemology*, pp. 265–90. New York: Routledge.

——2021: 'Ideal Theory, Ideology, and the Epistemology of Recognition'. In Paul Giladi and Nicola McMillan (eds.), *Epistemic Injustice and the Philosophy of Recognition*. New York: Routledge.

Rawls, John 1971: *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Spencer, Steven J., Claude M. Steele, and Diane M. Quinn 1999: 'Stereotype Threat and Women's Math Performance'. *Journal of Experimental Social Psychology*, 35(1), pp. 4–28.

Stanley, Jason 2015: *How Propaganda Works*. Princeton, NJ: Princeton University Press.

——and David Beaver forthcoming: *Hustle: The Politics of Language*. Forthcoming from Princeton University Press.

Steele, Claude M. 1997: 'A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance'. *American Psychologist*, 52(6), pp. 613–29.

——and Joshua Aronson 1995: 'Stereotype Threat and the Intellectual Test Performance of African Americans'. *Journal of Personality and Social Psychology*, 69(5), pp. 797–811.

Aristotelian Society Supplementary Volume XCV
doi: 10.1093/arisup/akab004

The Scotsman 2021: 'Donald Trump: Philosophy is the antidote to dangerous liars like the US President'. *The Scotsman*, 10 January 2021. https://www.scotsman.com/news/opinion/columnists/donald-trump-philosophy-antidote-dangerous-liars-us-president-scotsman-comment-3090608.

Tirrell, Lynne 2021: 'Discursive Epidemiology: Two Models'. *Proceedings of the Aristotelian Society Supplementary Volume* 95, pp. 000–000.

Young, Iris Marion 1990: *Justice and the Politics of Difference*. Princeton, NJ: Princeton University Press.

Aristotelian Society Supplementary Volume XCV
doi: 10.1093/arisup/akab004