



# Distorted Debates

Claudia Picazo<sup>1</sup>

Accepted: 2 November 2022 / Published online: 24 December 2022  
© The Author(s) 2022

## Abstract

One way to silence the powerless, Langton has taught us, is to pre-emptively disable their ability to do things with words. In this paper I argue that speakers can be silenced in a different way. You can let them speak, and obscure the meaning of their words afterwards. My aim is to investigate this form of silencing, that I call *retroactive distortion*. In a retroactive distortion, the meaning of the words of a speaker is distorted by the effect of a subsequent speech act by a different speaker. After introducing this notion, I explore some reasons why retroactive distortions can be difficult to challenge and argue that, besides constituting a communicative injustice, they can eliminate topics from public consideration and therefore erode public debate.

**Keywords** Retroactive distortion · Silencing · Hermeneutical injustice · Propaganda · Public deliberation

## 1 Introduction

In her ground-breaking paper on silencing, Langton wrote:

If you are powerful, you sometimes have the ability to silence the speech of the powerless. One way might be the ability to stop the powerless from speaking at all. Gag them, threaten them, condemn them to solitary confinement. But there is another, less dramatic but equally effective, way. Let them speak. Let them say whatever they like to whomever they like, but stop their speech from counting as an action. More precisely, stop it from being the action it was intended to be. (Langton 1993, p. 299)

One way to silence the powerless is, as Langton has shown, to pre-emptively disable their ability to do things with words, to create a communicative environment in which their words cannot have their intended illocutionary force. My aim here is to show that, if you are powerful, you can silence the powerless in a different way. You can let them speak, but distort the meaning of their words afterwards. Moreover, you can obscure the meaning of their words in ways that are particularly difficult to detect and challenge. In the examples I will discuss, the distortion is performed via an utterance that expresses praiseworthy ideals. If I am right, neutral sentences can be used to undermine the

communicative agency of oppressed groups, and to corrupt public deliberation.

I call the phenomenon of distorting the meaning of a previous utterance *retroactive distortion* (Sect. 2). In a retroactive distortion, the words of a speaker are distorted by a subsequent speech act. Her words, that could once be interpreted in the intended way, are obscured. Examples of retroactive distortions include the “All Lives Matter” response to “Black Lives Matter” and expressions of disapproval of all forms of violence offered as a correction to feminist protests against gender-based violence. In these examples, the meaning of the words of Black Lives Matter activists or feminist associations is obscured by a subsequent utterance. As I will show, making sense of the second speech act motivates a reinterpretation of the first. For instance, claiming that all lives matter as a reply to “Black Lives Matter” motivates an exclusive reading of “Black Lives Matter”, i.e., that *only* Black lives matters. The content that the first speaker ends up being attributed crucially depends on the second speaker’s words. Thus, the second speaker retroactively distorts the words of the first.

My interest is in those cases in which the distortion contributes to a system of oppression. This kind of distortion has two damaging effects (Sect. 3). Retroactive distortions can wrong the speaker and, in this sense, constitute a communicative injustice. But they can also have harmful consequences for society more broadly. As I see it, retroactive distortions harm public deliberation. By modifying the meaning of the words of an individual or group, the

---

✉ Claudia Picazo  
cpicazo@fsof.uned.es

<sup>1</sup> UNED, Paseo de Senda del rey 7, 28040 Madrid, Spain

distortion can remove the topic under discussion from public consideration. It can also exclude the perspective of the person whose words have been distorted. Retroactive distortions corrupt public deliberation. They can be, and have been, used as propaganda. I will argue that the linguistic mechanism is especially suitable to surreptitiously erode democratic ideals. One interesting feature of retroactive distortions is that they do not make use of explicitly harmful messages. The utterance that operates the distortion need not be an instance of hate speech, or derogatory or explicitly harmful in any sense. It might be neutral (at least from a semantic perspective), or even to embody praiseworthy ideals. Their damaging effect is the result of the interaction between the two utterances.

Moreover, retroactive distortions are difficult to reverse (Sect. 4). Because of the associations they make salient and the meta-discussion they force, together with the fact that they can undermine the standing of the speaker, their effects tend to survive attempts to block them. Moreover, challenging an apparently neutral utterance can itself be difficult. Hence their relevance and the need to properly understand their working.

## 2 Towards a Notion of *Retroactive Distortion*

Let us start with an example of retroactive distortion: the interaction between the slogan “Black Lives Matter” and the reply “All Lives Matter”. “Black Lives Matter” is the slogan of a social movement with the same name (BLM, for short) that began in the USA in 2013 as a response to police violence and other forms of racism.<sup>1</sup> BLM activists have been using the phrase “Black Lives Matter” as a slogan in different ways, including having it printed on posters shown in protests, used as a hashtag on social media, etc. Shortly after the BLM movement gained popularity, a second movement, this time claiming that “All Lives Matter” (ALM, for short), arose as a reply to the BLM message. Although there can be doubts about the extent to which the two groups are engaging in a conversation, or whether the activists of each group are better described as targeting different audiences, it seems that the ALM message emerged as a reply to the BLM one. In a context in which activists were trying to draw attention to the prevalence of racism by claiming that Black lives matter, some people opposed to this a second, allegedly more inclusive, message—that all lives matter. But how does exactly the interaction between the two work?

<sup>1</sup> According to their website (<https://blacklivesmatter.com/about/>), their mission is to “eradicate white supremacy and build local power to intervene in violence inflicted on Black communities by the state and vigilantes.”

In an insightful paper on how post-racialism results in hermeneutical injustice, Anderson writes:

For many, the meaning of the phrase “Black Lives Matter” is quite clear. There is an implicit “too” attached to the end of the phrase so that it should be read “Black lives matter, too!”. We can call this an inclusive reading. [N]ot everyone interprets the phrase in this way, however. A common retort to “Black lives matter” is “All lives matter”, expressing the idea that by singling out Black lives the former phrase represents a devaluing of non Black lives. Thus, the elliptical element is more like “Only Black lives matter”. We might call this an exclusive reading. (Anderson 2017, p. 109)

In this passage, Anderson distinguishes two readings of the BLM slogan. On the inclusive reading, the slogan means “Black lives matter, too”. By contrast, on the exclusive reading, the slogan would mean “Only Black lives matter”. Anderson’s quote further suggests what we can call a charitable view of the interaction between the BLM and the ALM slogan. In this view, some (many) people interpreted the BLM slogan in its inclusive reading, and some other people interpreted it in the exclusive way. The ALM slogan is the answer to the second interpretation. According to this option, the dynamics go as follows: some people understand “Black Lives Matter” as meaning “Only Black lives matter” and consequently reply “All Lives Matter”.

Here I want to defend a less charitable option. Anderson is right that the slogan is in principle open to the two readings. Moreover, it sounds plausible to assume that some people spontaneously interpreted it as inclusive, whereas other people was automatically inclined towards the exclusive reading. However, we know that interpretation depends on context, and in the context in which the slogan was coined, “the meaning of the phrase ‘Black Lives Matter’ is quite clear”, as Anderson points out. In a context of police violence against Black people, the inclusive reading sounds more natural. What, then, is the effect of claiming that “All Lives Matter”? According to the distortion view, answering “All Lives Matter” to “Black Lives Matter” distorts the meaning of “Black Lives Matter”. It is the ALM message that motivates the exclusive reading. In this second view, the relation between “All Lives Matter” and “Black Lives Matter” is reversed. “All Lives Matter” is not an answer to a pre-existing reading, but precisely what triggers the reading. As I see it, “All Lives Matter” promotes the exclusive reading of “Black Lives Matter” and has the capacity to lean interpreters who would have otherwise retained the exclusive reading, or no specific reading at all, towards the exclusive reading. Utterances of “All Lives Matter” retroactively distort the meaning of the BLM slogan.

The distortion view underlies Keiser’s discussion of discourse structure manipulation (Keiser 2021). According to Keiser, ALM activists manipulate the topic under discussion mid-discourse, which results in the misinterpretation

of “Black Lives Matter”. In order to explain how discourse manipulation can affect the interpretation of a previous utterance, Keiser uses formal pragmatics and, in particular, the Question Under Discussion framework (Roberts 2021). In this framework, discourses are structured around a set of questions (Questions Under Discussions, QUDs) that represent the topic and sub-topics under discussion. The goal of the conversation is to answer the main QUD. Moreover, QUDs affect interpretation. For instance, they play a crucial role in the generation of implicatures. Keiser uses as an example the sentence “Some pizzas were delivered”. Suppose that we are in a context where the QUD is *Were all pizzas delivered?* and someone says “Some pizzas were delivered”. This triggers the implicature that not all pizzas were delivered. The mechanism is broadly Gricean (Grice 1989). According to the Maxim of Quantity, speakers must make their contributions as informative as required for the purposes of the conversation. Here, the contribution fully answers the QUD, and therefore is informative enough, only on the assumption that the speaker means that some, but not all, pizzas were delivered. By contrast, in a conversation in which the QUD was *Was any of the pizzas delivered?* the implicature would not be derived.

Keiser’s explanation of the interaction between the BLM and the ALM slogans goes as follows. When BLM activists use the slogan “Black Lives Matter”, they address the QUD (topic under discussion), *Do Black lives matter?*, and provide an affirmative answer, *Black lives matter*. As Keiser notes, this choice of QUD is not arbitrary, there are good reasons to think that this is in fact the QUD BLM activists are addressing. First, there are linguistic reasons. QUD can be identified using focus. In this case, the absence of focus on the word “Black” suggests that the activists are addressing the polar question *Do Black lives matter?*, and not the wh-question *Which lives matter?* Second, considerations about the context support the polar QUD. In the context in which the BLM movement began, certain practices suggested that it was an open question whether Black lives mattered or not. So, let us conclude that BLM activist are addressing the QUD *Do Black lives matter?* Relative to this QUD, the content of the utterance is *that Black lives matter*.

Now, what happens when ALM activist utter their slogan? First thing to note is that “All Lives Matter” is uttered as a corrective of “Black Lives Matter”. It is used in opposition to this first slogan. Because of this, it only makes sense assuming that the question that both slogans target is *Which lives matter?* Otherwise, there would be nothing that it would be correcting. If it is a corrective, it must assume that BLM activists are answering the same question they are addressing. But this assumption is precisely what is problematic. If the slogan “Black Lives Matter” is an answer to the QUD *Which lives matter?*, and if we further make the plausible assumption that it is a complete answer, then it

triggers an implicature, namely *that only Black lives matter* (Anderson’s exclusive reading). Keiser’s point is that when they claim “All Lives Matter”, ALM activists introduce (by accommodation) a second QUD (*Which lives matter?*); relative to this QUD the slogan of BLM activists says *that only Black lives matter* (exclusive reading).

The misinterpretation of the BLM slogan exemplifies the phenomenon of *retroactive distortion*. By changing the QUD mid-discourse, ALM activists retroactively distort the content of the BLM slogan, a slogan that had been used many times before. After the QUD is modified, the interpretation of the slogan shifts from the inclusive to the exclusive reading.

Retroactive distortion consists in the modification of the content attributed to a speech act SA<sub>1</sub> because of the effect of a subsequent speech act SA<sub>2</sub>, performed by a different speaker:

Retroactive Distortion:

A speaker S<sub>2</sub> performs a speech act SA<sub>2</sub> that motivates a new interpretation of a previous speech act SA<sub>1</sub> (semantic content, explicature, implicature) by a different speaker S<sub>1</sub>, where SA<sub>2</sub> obscures S<sub>1</sub>’s intended meaning.

Retroactive distortion has two features: it is a phenomenon of retroactive character and it consists in a distortion. What I mean by *retroactive* is that it changes the (current) status of a past event. When SA<sub>2</sub> is performed, the content of SA<sub>1</sub> is re-interpreted. Something that happens in the future affects the impact of the utterance. The analogy here is with *ex post facto* laws. An *ex post facto* law changes the legal status of actions that happened in the past. For example, an amnesty law may retroactively decriminalize some acts: acts that constituted a crime cease to be criminal once the law is approved. My suggestion is that something similar happens with some speech acts. A retroactive distortion retroactively modifies the interpretation of a speech act: the speech act is now interpreted in a new way and, in this sense, it acquires a new status. As for the distortion, it operates a *modification* of the content attributed to the original speech act.

The retroactive character of speech acts has been studied by Langton, who, following Austin, holds that it is possible to retroactively undo a speech act (Langton 2018). In her view, this is what happens when a presupposition on which the felicity conditions of the speech act depend is blocked.<sup>2</sup> The manoeuvre of blocking, for instance, the presupposed authority, can retroactively undo the speech act. Thus, Langton’s idea is that one can retroactively disable a speech act.

<sup>2</sup> In conversational dynamics, blocking consists in a resistance to accommodate the presuppositions of an utterance. By contrast, rejection would target the whole utterance, and crucially its semantic content.

There are similarities between my notion of retroactive distortion and Langton's analysis of blocking as retroactive undoing. In both cases, an ulterior speech act affects a previous one. The common idea is that one can use speech to counter previous speech. However, our analyses are different in three important respects. First, on the linguistic side, my notion of retroaction is weaker. In my analysis, retroactive distortions motivate new interpretations, but I do not claim that they undo an act in any strong sense. Second, and also concerning the linguistic properties of the act, a retroactive distortion does not target the felicity conditions of the speech act, but its meaning. It does not block the speech act. Rather, it obscures its significance. Third, our focus is on very different forms of countering speech. Langton's proposal is about blocking as counterspeech, i.e., as a way to counter hate or oppressive speech. By contrast, my analysis focuses on how speech can be used to deprive speakers of their communicative agency. The retroactive distortions that I focus on do not promote social justice. They undermine it.

Retroactive distortions are also similar to other phenomena such as interpretive injustice (Peet 2017) and discursive injustice (Kukla 2014). Interpretive injustice occurs in those cases in which the intended meaning of a speaker is misinterpreted because of the stereotypes associated with her social identity. Peet imagines a woman giving an entrepreneurial pitch who says "We aim to achieve x". Despite the fact that she means something like "We will achieve x", and this is what the audience actually grasps when it is a man who uses the same sentence, in this case the audience interprets the woman's words as meaning something equivalent to "We would like to achieve x", that is, as weaker than intended, simply because she is a woman. Here, there is a misinterpretation. We can say that existing stereotypes distort the woman's intended message. Analogously, in a retroactive distortion an utterance distorts a previous message. However, note that I am using *distortion* in a stronger sense. In Peet's proposal, there are two distinct contents: the intended meaning and the received meaning. These two contents remain unmodified. The distortion consists in attributing the speaker a meaning that she did not mean, regardless of the fact that the audience was in a position to retrieve this speaker meaning. By contrast, what I am suggesting is that the received meaning itself (or the available meaning, or the content-in-context) changes as a consequence of the second speech act.

What about the content that is distorted? I include here what we can call the total content of the speech act, that is, the semantic content, together with the explicatures (or modulated content) and implicatures, if there are any. For our current purposes, we can understand semantic content as the content determined by the conventional meaning of the utterance in context. This content departs from the conventional meaning whenever there is an indexical

or some other context-sensitive expression. The explicature or modulated content includes the pragmatic adjustments of the semantic content that are not mandated by linguistic meaning, and implicatures are those non-literal meaning that speakers might mean. In the example above, the distortion concerns the triggering of an implicature. However, other cases could involve the other kinds of content mentioned, as I show below.

Whenever the content departs from conventional meaning, including indexicality, it has to be fixed in context. In deciding how exactly it is fixed, one can appeal to speaker intentions, abductive reasoning given the available contextual information on the part of the interpreters, discourse structure... Here I am assuming that content is somehow fixed, without discussing how. This assumption, however, can be challenged by those who think that meaning is negotiated in conversation or the product of the interaction between speaker and audience.<sup>34</sup> From this point of view, one could reply that there is no distortion of a previous meaning, but an on-going negotiation of what the content of the first speech act is. Against this objection, I think that the previous example intuitively involves the modification of a previous content. Note that, in the case of "Black Lives Matter", this message had been used well before some people started replying that "All Lives Matter". Presumably, it already had a content. Moreover, this is not a case in which there is a disagreement about how to interpret an utterance, or where the interlocutors engage in negotiation, but a case in which the context has been modified in order to suggest a new interpretation of a past utterance. Retroactive distortion occurs via context manipulation, not negotiation.<sup>5</sup>

It is worth mentioning that attempts to retroactively distort a speech act are not necessarily successful. It can happen that, after coming across the "All Lives Matter" slogan many people take BLM activists to care only for Black lives (exclusive reading), but it could also happen that most people just take ALM activists to be confused about the BLM movement. Moreover, there need not be a homogeneous reaction. It can happen that the distortion is successful for some audiences—those who share the beliefs, attitudes or goals of the distorter, or those who see the distorter as trustworthy—but not for others—those who concede more authority or trustworthiness to the original speaker than to

<sup>3</sup> As an example, Carassa and Colombetti distinguish speaker meaning from what they call *joint meaning*, i.e. meaning that 'is formed every time a speaker and a hearer jointly commit to the fact that a specific communicative act has been performed' (Carassa and Colombetti 2009, p. 1849).

<sup>4</sup> Thanks to Marcin Lewinski, Dima Mohammed and Lilian Bermejo Luque for discussion on this point.

<sup>5</sup> Nonetheless, I acknowledge that the borders between retroactive distortions and negotiations over meaning are blurred.

the distorter, or who are more aware of the original context of utterance, for instance.<sup>6</sup>

In order to show that the phenomenon is widespread, I will mention three other examples of retroactive distortion. The first follows the same pattern exhibited by the “All Lives Matter” case. For years, feminist movements have been urging to stop violence against women. In Spain, these claims include slogans such as “Contra la violencia de género”, “Contra la violencia a las mujeres”, “stop violencia de género”,<sup>7</sup> etc., used by feminists and also public institutions, as well as in politician’s discourses. One of the aims of the feminist movement has been, and still is, to call attention to the specific problem of gender-based violence and to demand solutions. The slogans have been repeatedly used in protests organized in response to women’s murders. Arguably, the context in which such phrases are used makes it clear that the topic under discussion is violence against women. However, in the last years, the Spanish right-wing party Vox has distorted the debate and, with it, it has broken the consensus on gender-based violence. This party is against the current Spanish legislation on gender-based violence, and claims that all forms of violence should be treated equally, regardless of the historical or sociological context in which they occur. As part of their effort to challenge feminist policies, Vox has mimicked the feminist protest organized after a murder, but this time organizing protests against what they call intra-family violence, which includes any crime in which victim and perpetrator belong to a family unit. In this context, members of Vox have exhibited the slogan “Contra todo tipo de violencia familiar”<sup>8</sup> in protests organized as a reply to non-gendered-based crimes. But here, as in the BLM/ALM case, there is implicit reference to the feminist slogans. It is a correction of the feminist focus on gender-based violence. Vox’s slogan suggests that feminists (as well as the public institutions who have implemented policies against gender-based violence) are against some forms of violence only (i.e., those in which the victim is a woman, exclusive reading) and do not care about other forms of violence. With their words, they distort the previous message of feminists and misrepresent their goals.

The second example is different.<sup>9</sup> In her divorce deposition, from the 90s, Ivana Trump stated that Donald Trump had raped her. Years later, when the accusation gained public attention, Donald Trump’s lawyer, Michael Cohen,

explained that Ivana Trump “had felt raped emotionally... She was not referring to it [as] a criminal matter, and not in [the] literal sense”. Here we have a speech act that was presumably understood in the literal sense and, years later, a second speech act that explains the first, motivating a non-literal interpretation were “raped” would be equivalent to “emotionally raped”. This exemplifies another pattern of retroactive distortion. Suppose a speech act is understood in a certain way, that coincides with the meaning that the speaker intended. Later, another speaker comes in and “explains” what the first meant, by giving information about the speaker or the content (“She was not speaking literally”, and so on...), so that the audience ends up interpreting the first speech act in a different way, namely, in the way suggested by the second speaker. As I see it, the second speech act retroactively distorts the first.<sup>10</sup>

In these two cases the distorter’s goal is to manipulate the conversation. In other cases, the distortion can be merely accidental. Mohammed’s analysis of the anti-#MeToo Manifesto includes what can be considered an accidental retroactive distortion (Mohammed 2019). The anti-#MeToo Manifesto is an open letter published in the French newspaper *Le Monde* in 2018, one year after the #MeToo movement emerged. This letter was signed by actress Catherine Deneuve and radio host Brigitte Lahaie, among others. As Mohammed explains, even though the manifesto is clearly not an explicit rape apology, some of its claims have the potential to be misinterpreted as such. One of these claims is “woman can [...] enjoy being the sexual object of a man”, a sentence that can be re-interpreted as a form of saying that women can get something good out of harassment. What is interesting here is that this potential can be activated by subsequent claims of any of the signatories. This seems to have happened with the quoted statement. A few days after the manifesto was released, Lahaie said in a TV debate that “One can have pleasure during a rape”. Mohammed notes that Lahaie’s words support an interpretation of the manifesto according to which women can get something good out of harassment. Moreover, this meaning was attributed not only to Lahaie and the manifesto, but also to each of the signatories, to the extent that Deneuve felt compelled to apologize. Because they had become, in Mohammed’s terminology, argumentative associates, the meaning of the collective utterance of the sentence “woman can [...] enjoy being the sexual object of a man” can be distorted by Lahaie’s statement, regardless of whether Lahaie’s intended

<sup>6</sup> In the BLM case, as well as in other cases that I present below, the distortion is not entirely successful. Nonetheless, it is successful for certain audiences.

<sup>7</sup> These slogans mean “Against gender-based violence”, “Against violence against women” and “stop gender-based violence”.

<sup>8</sup> “Against all forms of family violence”.

<sup>9</sup> Here I follow Manne (2017).

<sup>10</sup> There can be doubts that this is what happens with Ivana Trump’s words, as she has denied her previous testimony. Here I use Cohen’s words to exemplify a possible way to retroactively distort an utterance.



to manipulate the meaning of the manifesto or not.<sup>11</sup> This example shows that the distorter can distort the message either intentionally or accidentally.

In what follows I focus on intentional retroactive distortions and, more specifically, I will pay special attention to those that distort a previous message via the manipulation of the topic under discussion. Moreover, I focus on distortions that contribute to perpetuate an oppressive system—what we can call *oppressive retroactive distortions*. However, let me note that the same linguistic mechanism<sup>12</sup> could in principle be used in less problematic cases, or even in order to counter hate speech or toxic discourse.<sup>13</sup>

### 3 Harmful Effects

Retroactive distortions as the ones exemplified in the BLM/ALM case can be harmful. The aim of this section is to explore two forms of harm. First, retroactive distortions can constitute a communicative injustice. Second, I will argue that retroactive distortions not only harm the individuals or groups whose words are distorted, but also undermine political deliberation. By obscuring the messages of some groups, the distortion might preclude deliberation about the topics that the target group intended to introduce to public consideration. These two consequences are of course related and can be considered two sides of the same coin: individuals and groups are wronged in that their communicative agency is diminished (and even their hermeneutical resources are diminished, as we will see) and, by the same token, political deliberation is impoverished by the exclusion of their perspectives and topics of interest.

Let us start with the harm towards the speaker. Retroactively distorting the words of a speaker is a way of making

unavailable her intended meaning. The result is a form of silencing: the message intended by the speaker is lost. We can say, somewhat metaphorically, that retroactive distortions disable speaker meaning. By manipulating the context in which they are interpreted, the words of the original speaker lose their intended import. Thus, her communicative agency, understood as her ability to communicate with others, is undermined.<sup>14</sup>

Anderson goes even further and argues that the misinterpretation of “Black Lives Matter” constitutes a hermeneutical injustice. Hermeneutical injustice is “the injustice of having some significant area of one’s social experience obscured from collective understanding owing to hermeneutical marginalization” (Fricker 2007, p. 158). An example of hermeneutical injustice would be the injustice that women victim of sexual harassment suffered before the concept of sexual harassment was coined. Because of the conceptual lacuna, their experiences were obscured, difficult to understand for them and also difficult to share with others in an understandable way.

Why does retroactive distortion in the BLM case constitute hermeneutical injustice? Anderson’s point is that “the pursuit of the post-racial ideal results in hermeneutic injustices” (2017, p. 145). Proponents of post-racialism sometimes support the elimination of race talk. Now, Anderson argues that nothing guarantees that eliminating race talk will result in a society in which race has no importance. On the contrary, racism could survive, even if we ban it from discourse. The problem, then, is that, if we eliminate race terms but racism persists, it will be difficult to detect and theorize about—it will be obscured, in Fricker’s term. The “All Lives Matter” response to “Black Lives Matter” is precisely a way of eliminating talk about race. As such, it can contribute to the elimination of race talk, which would constitute a hermeneutical injustice. In Anderson’s words, “it has the effect of blocking marginalized speakers’ contribution to making sense of social reality.” (2017, p. 146).

Now, in my view, the contribution to the elimination of race talk is not the only way in which the ALM message can be damaging. The distortion could have more local or temporal effects. It could, for instance, have the effect of stopping the discussion about racist police violence against Black people—for some time, for some people. Although this would not constitute hermeneutical injustice in any strong sense, it would nonetheless be a communicative harm. Thus the second harmful effect of the distortion is that it can make

<sup>11</sup> This example is in fact more complex. Given that Lahaie was one of the signatories, her words can be understood as clarifying what she (or all of them) meant. However, I think that the manifesto is a collective statement. Even though Lahaie might intend to clarify what she meant, she is nonetheless distorting the words of a collective speaker. Also, one could also argue that the interpretation “Women can get something good out of harassment” was already active and Lahaie’s statement merely reinforces it (rather than distort the message), or even that her words are part of the meaning-fixing process.

<sup>12</sup> The mechanism is content modification via context manipulation, i.e., the context is modified and the new context motivates a new interpretation of a previous speech act. There can be more specific mechanisms: one can manipulate context by shifting QUD mid-discourse, by saying something about the speaker, etc.

<sup>13</sup> I have already mentioned Langton’s view on blocking as counterspeech. In a similar vein, Caponetto & Cepollaro (ms.) introduce a form of distortion (called *bending*) that can be used in counterspeech. In their view, one can counter hate speech by intentionally and visibly attributing the speech act a non-hateful interpretation. One can perhaps use retroactive distortions in a similar way.

<sup>14</sup> Keiser (2021) considers the loss of communicative agency as a form of epistemic injustice. I prefer to see it as silencing. The reason is that, in the examples that I have examined, the distortion is used to, using Langton’s words, “stop [the utterance] from being the action it was intended to be” (Langton 1993, p. 299) (in particular, by preventing it from having the meaning it was intended to have).

certain topics, such as racist police violence, unavailable or less available, not a matter of public discussion or political debate, e.g. less discussed in media. This loss would harm both those groups and individuals who have introduced them (communicative or hermeneutical injustice), and those who could potentially benefit from the discussion, for instance, those who could benefit from its resulting policies. For clarity's sake, I will call this harmful effect *topic-elimination*.

Topic-elimination is particularly relevant given the intimate connection between public debate about policy and democracy.<sup>15</sup> According to deliberative accounts of democracy, democratic legitimacy depends on the process of public deliberation. On this view, a given policy is legitimate only to the extent that it is the result of joint deliberation. If this is correct, then public deliberation is something valuable. Now, not any kind of deliberation will do. When we speak of public deliberation, what we mean by that is rational, informed deliberation. It is clear that this kind of deliberation is damaged when the information and arguments presented and about which citizens are supposed to deliberate are somehow defective, for example, when lies are presented as truths, rational debate is replaced by emotional responses or fallacies occupy the place of good arguments.<sup>16</sup> My point here is that deliberation can also be damaged in a different way. Sometimes, there simply is no deliberation about certain topics—topics that is in the interest of oppressed groups to discuss—because the efforts to introduce them to public consideration have been disabled. In these cases, the deliberation excludes those groups. This kind of deliberation, of course, falls short of being democratic. Some people's perspectives have no place in it. Retroactive distortion is precisely a mechanism whereby topics under discussion or perspectives can be eliminated.

Retroactive distortions can be considered a form of propaganda. According to a classical view, propaganda is speech that manipulates reason in order to close off debate.<sup>17</sup> As we saw, some retroactive distortions close off debate about certain topics and, furthermore, they do it sneakily. The participants in the debate can fail to realize that the topic has change and that some messages have been distorted, which is a way of manipulating reason. When a distortion occurs, the audience reasons using false premises. For instance, that BLM activists believe that only Black lives matter. Thus, some retroactive distortions count as propaganda in the classic sense. Nonetheless, the classical definition is too broad

to capture the specific features of the cases I am interested in here, and in what follows I will use Stanley's view of propaganda in liberal democracies (Stanley 2015).

According to Stanley, the central characteristic of propaganda in liberal democracies is that it usually occurs masked and thus it is not always easy to identify. In his view, political propaganda has to do with the employment of political ideals for political purposes. Stanley distinguishes two kinds of propaganda. Supporting propaganda is speech that is presented as an embodiment of certain ideas and supports those ideals by nonrational means. Undermining propaganda includes discourses presented as an embodiment of certain ideals but that nonetheless tend to undermine those very ideals. In democratic societies, undermining propaganda often takes the form of speech that superficially seems to support democratic ideals, but that in fact undermines them. Because of this use of democratic ideals, it can be specially difficult to detect.

Retroactive distortions can be, and have been, used as undermining propaganda. In particular, "All Lives Matter" and "Contra todo tipo de violencia familiar" have been used as propagandistic claims that erode the ideals they appear to support. Note, first, that they clearly embody democratic ideals. "All Lives Matter" is an expression of the ideal of equality, a vindication that all lives are equally valuable. "Contra todo tipo de violencia familiar" plausibly vindicates dignity and respect for all people. It says that we (the utterers) are against all forms of family violence, presumably because all members of a family deserve equal respect. But both claims vindicate these democratic ideals only on the surface. In fact, the effect that using them brings about, in the specific context in which they are uttered, is exactly the opposite. "All Lives Matter" counters the effort to make racist police violence visible. The "all" in "all lives" has the effect of making Black lives invisible to public debate. Something equivalent happens with "Contra todo tipo de violencia familiar". By subsuming sexist violence under family violence, it makes it invisible and closes off debate about its specific mechanisms and prevalence in our societies. The words "all" and "todo" might seem to embody democratic ideals, but in fact their effect is to exclude certain groups, which goes against the spirit of democracy. We can say that these messages semantically express democratic ideals but that they (or some uses thereof) pragmatically erode those very ideals. What they do is the opposite to what they say.

Something interesting about Stanley's account is that in it propaganda is neither necessarily false<sup>18</sup> nor insincerely delivered. This captures well some important forms of propaganda used in liberal societies, such as the ones that

<sup>15</sup> See Cohen (1989).

<sup>16</sup> See Brown (2018) for the epistemic effects of propaganda and misinformation in general.

<sup>17</sup> This view, recognizable in Klemperer's work on the language of the Third Reich, is what Stanley (2015) calls the classical sense of propaganda.

<sup>18</sup> See also Tuttle Ross (2002). In Tuttle Ross' view, propaganda need not be false, but it is epistemically defective.

I am discussing. Although their use involves some kind of manipulation and erosion of democratic ideals, slogans as “All Lives Matter” are true. They can also be used sincerely. Moreover, they can be used by people who (wrongly but honestly) believe to be supporting a democratic ideal. In fact, the slogans I am using as examples have probably been used by people who were unaware of its silencing effects—together with people whose goal is to maintain the status quo or to avoid discussion about certain topics. Truth, sincerity, and good will, however, do not cancel out the harmful consequences of the distortion.

To sum up, retroactive distortions such as the BLM/ALM example can be considered a form of propaganda. Their effect is the silencing of the speaker or group whose words have been distorted, and they contribute to topic-elimination, which impoverishes public debate.

At this point, one could object that retroactive distortions alone cannot possibly have such harmful effects. The distortion is just a mechanism to change topic and obscure meaning, not something that makes the original topic or the intended meaning unavailable. In principle, one can introduce the topic again, or make clear what one meant. This is precisely what happens in many cooperative conversations. One of the participants introduces a topic, another introduces a new topic, etc. Or, one says something, and another says something that suggests a misinterpretation of the first. I start talking about my plans for the weekend but you are not interested and subtly change the subject; I say something ironically but you take it seriously and answer with a serious tone that makes the other participants in the conversation think that I was also being serious, etc. According to this objection, retroactive distortions are just like these ordinary ways in which conversations evolve, and can be challenged on similar grounds. Usually, I can complain “I was talking about my plans for weekend, let me finish!”, or “I was being ironic!” and redirect the conversation. But then, what could be so damaging about the distortion?

Against this line of thought, in the next section I will argue that retroactive distortions have some features that make them difficult to challenge. They are, using Lepoutre’s expression for hate speech and toxic discourse, sticky (Lepoutre 2019). Their effects tend to remain. Although some of them can be blocked, the kind of manipulative distortions that I analyse here easily survive attempts to deactivate them. It is partly because of this stickiness that they have the harmful effects explored in this section. Once the distortion has occurred, it is difficult to go back to the previous state of the conversation and continue to talk about Black lives or gender-based violence as if nothing had happened. Often, the topic becomes unavailable or available to

a significantly lesser degree<sup>19</sup> and the intended message is difficult to grasp.

## 4 Stickiness

In the previous section I argued that the harmful effects of retroactive distortion are at least partly due to the fact that retroactive distortions are difficult to challenge. But why is that so? In principle, it should be possible to block the distortion of a previous speech act. In a BLM protest after ALM activists started talking about “all lives”, a woman showed a poster saying “Yes, all lives matter, but we are focused on the black ones”. This utterance constitutes an attempt to block the QUD-shift. It does not deny what ALM activists say (namely, that all lives matter). However, by explicitly saying that they are focused on Black lives, it tries to block the change of topic. Similar strategies are in principle available: one can explain what one meant, claim that the distorter has misinterpreted her words, denounce the change of topic, etc. Unfortunately, reversing a retroactive distortion is not always easy.

A first reason is that, unlike hate speech, retroactive distortions can be performed via sentences that are neutral, or even via sentences that express praiseworthy messages. Because of this, the distorted speaker finds herself in a situation in which she needs to reject the effect of the distortion without rejecting its message. But this can be difficult to do: someone who tries to counter the effect of the ALM slogan can be asked ‘But don’t you agree that all lives matter?’. This derails the conversation and puts an unjust burden on the one who tried to counter the distortion, since she might now need to justify herself.<sup>20</sup> Moreover, the neutrality of the distorting utterance can preclude some audiences from identifying the harm.

Second, speakers cannot simply repeat the claim—just insisting that “Black Lives Matter!” is out of the table once the new QUD has been introduced. Instead, they are forced into a meta-discussion.<sup>21</sup> The most straightforward way in which a BLM activist can block or reverse the distortion is by replying that they were talking about Black lives. But by doing so they abandon the ground-level discussion and

<sup>19</sup> The topic is not unavailable for everybody. Society is not homogenous, as Medina notes in a discussion about hermeneutical injustice (Medina 2012). It will still be available for many people—crucially, for BLM activists and feminists. However, it can become less available as a topic for public consideration, as part of the political agenda.

<sup>20</sup> My impression is that attempts to counter apparently neutral forms of harmful speech are met with particularly high standards about what counts as a justified claim.

<sup>21</sup> Thanks to Álvaro Domínguez Armas for drawing my attention to meta-discussions.



move to a meta-discussion, i.e., a discussion about what the conversation was about. This is a problem because the detour can be costly. Moreover, the strategy can easily misfire. The attempt to reverse the distortion by a meta-discussion can give rise to a new debate at the metalevel. For example, there is a risk that the conversation ends up being about whether *Black lives*, as opposed to *all lives*, is a legitimate topic for public discussion. Given that there are practical limits to the conversations one can have and their duration, this can be a way to silence in practice the ground-level discussion.

The third reason why retroactive distortions are difficult to challenge is that, borrowing McGowan's expression, trying to reverse a distortion is like trying to un-ring a bell. This is a phenomenon that philosophers have identified in hate speech. Here I will follow McGowan's view, as it has been developed by Simpson and Lepoutre. In her work on oppressive speech, McGowan has put forward a linguistic mechanism, conversational exercitives, that explains how ordinary instances of hate speech by non-authoritative speakers, such as sexist remarks, can constitute oppression (McGowan 2009). In her view, the contributions that the interlocutors make to the conversation enact norms for the conversation. For example, the use of a slur for women makes it appropriate to rank women as inferior in the conversation. In this sense, most utterances, if not all, are exercitives—unless they are rejected, they enact norms for the conversation, and they do so in virtue of how conversations work. In principle, these norms can be rejected or reversed. Just as one enacts norms by speaking, one can cancel previous norms with more speech. Suppose that the norm that I have enacted is that the conversation is about our plans for the weekend (I have said something like “This weekend I’ll go camping”, thus accommodating a new topic). My interlocutors can easily change the topic, thus enacting a new norm that cancels the previous one (one can say, “What about your plans for summer?” and, if everybody go along, successfully change the topic). Now, according to McGowan, not all norms are easy to reserve. Some are easy, some are not. For example, Lewis thought that it is easier to raise standards than to lower them. Similarly, McGowan thinks that the norms enacted by hate speech are difficult to reverse. This is called the *asymmetric pliability of norms*.

Why are conversational norms asymmetrically pliable? Simpson provides a particularly convincing explanation (Simpson 2013). His proposal is that hate speech activates associations that tend to remain “because the discrediting of the association subtly perpetuates the association” (Simpson 2013, p. 571). For example, when one says something that establishes an association between women and sexual objectification, the usual ways to reject this claim (“Don’t use the word ‘bitch’”, etc.) still makes salient the relation between women and sexual objectification, which survives the audience’s attempt to reject the harmful utterance. The norms

enacted by hate speech are difficult to reverse because they make salient harmful associations, and it is easier to make an association salient than unsalient.

Now, how does this apply to retroactive distortions? Above I have characterized them as propaganda, not hate speech. Is propaganda also one of those case in which pliability is not guaranteed? Interestingly, Lepoutre (2019) has argued that the linguistic phenomenon that Simpson identifies is quite general. The reason is that, in Simpson’s explanation, the difficulty of reversing hate speech is not specifically related to hate, but to the properties of associations and salience. Now, if this is so, then any association that is made salient is potentially difficult to challenge. Let us go back to BLM/ALM. When the exclusive reading is suggested, this reading activates an association between BLM activist (the utterer) and a devaluing of non-Black lives (a view that BLM do not actually endorse but that is nonetheless suggested by the distortion). The problem is precisely that cancelling this association is not an easy task. BLM activists can try to clarify their position, but there is a risk that the doubt whether the association is true will remain. Even worse, if BLM activists explain that their views do not entail a devaluing of non-Black lives, their saying this re-activates the association. Rejecting the distortion can reactivate the problematic association.

Finally, retroactive distortion can undermine the status (or standing) of the speaker, and erode her moral or epistemic authority. Since it misrepresents the message, the distortion can misrepresent the speaker. This happens when the distorted message is flawed or inappropriate in some sense. Let us go back to the ALM/BLM case. According to the exclusive reading, BLM activists claim that only Black lives matters. Thus, the distortion suggests that BLM activists are indifferent towards non-Black lives (or even worse). Now, this is especially problematic in the context in which the distortion takes place, a context in which it is supposed to be a shared principle that all lives are valuable. In this context, people who are thought to be indifferent towards certain lives can be judged negatively on moral grounds. Thus, the ALM message has the power to undermine BLM activists’ moral authority. This, in turn, can undermine their status as speakers and have an impact on how their words are received. The reason is that there seems to be a correlation between the standing of a speaker and her conversational success.<sup>22</sup> If the audience distrusts the moral character of a speaker, this can be used as a reason for not trusting what she says, or simply disregarding her words. In some cases, this will mean that the distorted speaker will have to devote some efforts to regaining her status before she can become

<sup>22</sup> I have explored this correlation in Picazo (2021).

a conversational peer. In worse scenarios, the distorted speaker can be excluded from the conversation.

Moreover, the distortion can take advantage of pre-existing power imbalances and exacerbate them. Take again the BLM/ALM example. If those who claim that all lives matter are in a position of relative power, then it is quite likely that they will manage to shift the topic from *Black lives* to *all lives*. For example, they might have access to media that are difficult to reach for BLM activists and that might make their words appear more authoritative than they actually are. Conversely, if BLM activists are less powerful and therefore less influential, their attempts to block the distortion might sound unconvincing.

To sum up, I have presented several reasons that explain why retroactive distortions are difficult to reverse. These reasons work together: retroactive distortions make alternative interpretations and certain associations salient, which are difficult to make unsalient, and, on top of that, they undermine the position of the speaker, thus making her seem less reliable than she is. It raises doubts, about the speaker and the message, that persist attempts to reverse the distortion. Moreover, they force a meta-discussion, which implies a shift of focus. Consequently, undoing a distortion is more difficult than distorting a (non-distorted) message.

## 5 Concluding Remarks

Speech can be used to retroactively silence a speaker or group of speakers. The meaning of a speech act can be distorted by the effect of a subsequent speech act. When this happens, the speaker whose words are distorted is wronged. Her communicative agency is undermined. Furthermore, the quality of public deliberation is also affected. Through retroactive distortion, topics under discussion can be eliminated.

I will finish with some remarks on two different purposes for which retroactive distortions can be used. First, retroactive distortions can be used to prevent certain groups from having an influence on the political agenda. Imagine an oppressed group who tries to force discussion on a certain subject matter that has so far been ignored. Distorting their words is a way to counter and deactivate their efforts. But retroactive distortions can be put to a different use. They can be used to break existing consensus and to eliminate topics and perspectives that are already part of the political agenda. Arguably, it is this second use that underlies Vox's retroactive distortions. There is a consensus in Spain about how to fight gender-based violence, a consensus that includes specific legislation. However, one of Vox's aims is to break this consensus and to replace the current legislation on gender-based violence by a new legislation on family violence. The thesis that I have defended here is that speech can be used, together with other things, to achieve this aim.

Devaluing feminism by misrepresenting its commitments and goals can contribute to breaking the consensus. And this can be done by retroactively distorting the words of feminist associations and institutions.

**Acknowledgements** I would like to thank the participants at the Argumentation and Politics Conference in Granada for very helpful comments.

**Funding** Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. The funding was provided by Ministerio de Ciencia e Innovación, Government of Spain (Grants No. PID2019-105728GB-I00 and PID2021-123938NB-I00) (MINECO/FEDER, EU).

## Declarations

**Conflict of interest** The author has no conflict of interest to declare.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Anderson L (2017) Epistemic injustice and the philosophy of race. In: Medina J, Kidd I, Pohlhaus Jr G (eds) *The Routledge handbook of epistemic injustice*. Routledge, London
- Brown É (2018) Propaganda, misinformation, and the epistemic value of democracy. *Crit Rev* 30(3–4):194–218
- Caponetto L, Cepollaro B (ms.) *Bending as Counterspeech*
- Carassa A, Colombetti M (2009) Joint Meaning. *J Pragmat* 41(9):1837–1854
- Cohen J (1989) Deliberation and democratic legitimacy. In: Hamlin A, Pettit P (eds) *The good polity: normative analysis of the state*. Blackwell, Oxford, pp 17–34
- Fricker M (2007) *Epistemic injustice: power and the ethics of knowing*. Oxford University Press, Oxford
- Grice P (1989) Logic and conversation. In: Grice P (ed) *Studies in the ways of words*. Harvard University Press, Cambridge, pp 22–40
- Keiser J (2021) The “All Lives Matter” response: QUD-shifting as epistemic injustice. *Synthese* 199(3–4):8465–8483
- Kukla R (2014) Performative force, convention, and discursive injustice. *Hypatia* 29(2):440–457
- Langton R (1993) Speech acts and unspeakable acts. *Philos Public Affairs* 22(4):293–330
- Langton R (2018) Blocking as Counter-Speech. In: Fogal D, Harris D, Moss M (eds) *New Work on Speech Acts*. Oxford University Press, Oxford, pp 144–164
- Lepoutre M (2019) Can ‘more speech’ counter ignorant speech? *Journal of Ethics and Social Philosophy*. <https://doi.org/10.26556/jesp.v16i3.682>
- Manne K (2017) *Down girl: The logic of misogyny*. Oxford University Press, Oxford

- McGowan MK (2009) Oppressive speech. *Australasian J Philos* 87(3):389–407
- Medina J (2012) The Epistemology of resistance: gender and racial oppression, epistemic injustice, and the social imagination. Oxford University Press, Oxford
- Mohammed D (2019) Managing argumentative potential in the networked public sphere: The anti-#MeToo manifesto as a case in point. In B. Garssen, D. Godden, G. Mitchell, & J. Wagemans (Eds), *Proceedings of the 9th conference of the International Society for the Study of Argumentation*, (pp.&nbsp;813–822)
- Peet A (2017) Epistemic injustice in utterance interpretation. *Synthese* 194(9):3421–3443
- Picazo C (2021) Not all speakers are equal: harm and conversational standing. *Daimon* 84:81–95
- Roberts C (2021) Information structure: towards an integrated theory of pragmatics. *Semant Pragmat* 5(6):1–69
- Simpson R (2013) Un-ringing the bell: McGowan on oppressive speech and the asymmetric pliability of conversations. *Australasian J Philos* 91(3):555–575
- Stanley J (2015) *How propaganda works*. Princeton University Press, Princeton
- Tuttle Ross S (2002) Understanding propaganda: the epistemic merit model and its application to art. *J Aesthetic Educ* 36(1):16–30

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.