JUST KIDDING? TWO ROLES FOR THE CONCEPT OF JOKING IN POLITICAL SPEECH

By Zoe Walker



In this paper, I discuss two roles for the concept of joking in political speech. First, I discuss how claiming to have been joking can provide speakers with a powerful form of deniability. I argue that the aesthetic dimension of jokes makes such a denial especially well placed to undermine both a hearer's evidence for an utterance having been sincere, and, separately, their belief that it was sincere—I call the latter 'aesthetic gaslighting'. Second, I discuss the use of jokes to influence hearers' thinking and behaviour under the radar. I show that not only does the fact that an utterance was a joke fail to prevent it from influencing hearers, but in some cases, the fact that it was a joke actually makes it more influential than a sincere utterance would have been.

Keywords: joking, deniability, gaslighting, dogwhistles, sexist humour, salience.

Ratings challenged @CNN reports so seriously that I call President Obama (and Clinton) "the founder" of ISIS, & MVP. THEY DON'T GET SARCASM?

Donald Trump, Twitter, 12th August 2016

I. INTRODUCTION

Donald Trump has employed a number of rhetorical strategies over the course of his political career, but one of the most common is to respond to criticism of his utterances by claiming that he was only joking; or, as he often puts it, 'being sarcastic'. In some cases, this may be an attempt to save face after making a mistake. For instance, after referring to living former president Jimmy Carter as 'the late, great Jimmy Carter' in a 2014 speech, he tweeted: 'Of course I don't think Jimmy Carter is dead—saw him today on T.V. Just being sarcastic, but never thought he was alive as President, stiff!'

(Woodward 2020). And after a press briefing in 2020 in which he asked whether injecting disinfectant could help to cure Covid, he told the press that he was 'asking a very sarcastic question to the reporters in the room' (*Ibid.*). In these two cases, it is not in Trump's interest for people to take these utterances seriously—to come to believe that Jimmy Carter has died, or that we should genuinely be considering injecting disinfectant to cure Covid. Therefore, let us call these 'Anti-Interest Disavowed Claims' (AIDCs).

In other cases, however, claims that Trump later disavows as jokes or sarcasm are of a rather different ilk. Take, for example, the 2016 press conference in which he discussed Hillary Clinton's email scandal and said he hoped that Russia were 'able to find the 30,000 emails that are missing'—before later claiming that he was 'being sarcastic' (*Ibid.*), and later still, that '[i]t was all said in a joke' (Bennett 2020). Or consider his repeated claims during his 2016 campaign that Barack Obama and Hillary Clinton were 'the founders of ISIS'—about which he eventually complained in a tweet that CNN 'DON'T GET SARCASM' (Woodward 2020). In these two cases, it was in Trump's interest for people to believe the relevant claims—that Clinton had connections to Russia, and that she and Obama both had connections to ISIS—because uptake of these assertions would have helped his presidential campaign. Call these 'Pro-Interest Disavowed Claims' (PIDCs).

In the case of both AIDCs and PIDCs, there is something puzzling about Trump's disavowal of them as 'being sarcastic'. In the case of the AIDCs, the disavowal is highly implausible. It is difficult to believe that he jokingly claimed that Jimmy Carter was dead, in a completely deadpan tone, without making any kind of point with the joke; or that he jokingly asked whether injecting disinfectant might help to cure Covid, in a completely deadpan tone, at a White House press briefing about a current global pandemic. So why bother pretending? Could anyone really be convinced that he was not being serious, and if so, how?

In the case of the PIDCs, the puzzle is a different one. With these claims, it is slightly more plausible that Trump was joking. After all, his base may well find it funny to entertain the thought that Clinton has connections with Russia, or Obama with ISIS—we can see the point of being insincere here, whereas there was no clear point to uttering the AIDCs insincerely. The trouble is that, at least on the face of it, it looks like Trump's disavowal will be found most plausible by the wrong people. Those hearers who would not have believed or been influenced by these claims about Clinton and Obama in the first place—Trump's critics—will also not be convinced that he was speaking entirely insincerely, and will insist that he did genuinely mean to suggest a connection between Clinton and Russia, and Obama and ISIS. It is only those who might have believed these claims in the first place—Trump's base—who will also believe him when he says he was joking or being sarcastic. So does claiming he was joking ultimately prevent these Pro-Interest Claims

from advancing his interests, or might these utterances still influence hearers in a manner favourable to him?

In this paper, I am interested in two different functions of the concept of joking in political speech, which are illustrated by these examples from Donald Trump. In Section II, I discuss joking as a way of building in the possibility of denial, even for cases where denial would otherwise seem implausible, as demonstrated by the AIDCs. I show that claiming to have been joking is a particularly powerful form of deniability, because the aesthetic aspect of a joke—its purported funniness—means such a denial can more easily undermine both a hearer's *evidence* for an utterance having been sincere, and, separately, their *belief* that it was sincere, than other forms of denial.

In Section III, I discuss joking as a way to influence hearers' thinking and action, as demonstrated by the PIDCs. I show that in fact, the disavowal of an utterance as 'only joking' or 'being sarcastic' not only fails to prevent it from influencing hearers, but can in some cases—those where the joke is sexist or racist—give the initial utterance greater influence than if it had been uttered sincerely.

Before I begin my argument in earnest, there are two further points worth briefly addressing. First, are these kinds of disavowal particular to Trump, or do they occur in political speech more generally? And second, is it justifiable to treat disavowals that invoke sarcasm as a sub-category of disavowals that invoke joking, as I am doing here? I discuss these two points in turn.

First, while Trump is an especially prolific user of this kind of strategy, these two roles for the concept of joking or sarcasm are by no means unique to him. Among Trump's allies, the strategy is a popular one. In 2018, his Senior Counsellor, Kellyanne Conway, claimed that former FBI Director James Comey 'swung an election' for Trump in 2016, but later disavowed the claim as 'sarcastic'—an AIDC, given that Conway's claim contradicted the narrative that the loss was solely down to Hillary Clinton's ineptitude (Stein & Arciga 2018). And Republican politician Marjorie Taylor Greene said in 2022 that she was making a 'sarcastic joke' when stating that if she had organised the 6th January Attack on the Capitol, 'it would've been armed' and 'we would have won'—a PIDC, given that it would benefit her for her audience, the New York Young Republican Club, to believe it (Steakin 2022).

Further afield, former Italian Prime Minister Silvio Berlusconi has used the concept of joking to disavow some Anti-Interest Claims, for instance his 2012 claim that if the European Central Bank did not print more money, 'we should have the strength to say 'ciao, ciao' and leave the euro', which he later said was 'a joke' (Day 2012). Meanwhile, former UK Prime Minister Boris Johnson has used joking to disavow Pro-Interest Claims, such as when he disavowed his 2002 description of 'flag-waving piccaninnies' and of black Africans with 'watermelon smiles' as 'wholly satirical', after being accused

of racism (Forrest 2019). And this strategy is not exclusive to right-wing political speech. In 1999, then US President Bill Clinton critiqued George W. Bush's self-described 'compassionate conservatism' as being against various policies such as 'Social Security and Medicare', but 'feel[ing] terrible about it' (Clymer 1999). Clinton later said these remarks were made to 'cut the atmosphere a little bit' and 'give us something to laugh about', and that those who complained 'ought to lighten up'—a PIDC, as it was in his interest for people to believe his critique of Bush (CBS News 1999).

The second question is whether it is right to treat all of these cases as uses of the concept of *joking*, given that many of Trump's disavowals, for instance, invoke sarcasm rather than joking. Though in some contexts the two can come apart, in this paper I will treat the claims invoking sarcasm as a sub-category of the claims invoking joking. In large part, this is because the speakers themselves often run the two together, and do not treat them as relevantly different. For instance, as quoted above, Trump disavowed his remark about Hillary Clinton's missing emails as 'being sarcastic' on one occasion and 'a joke' on the other. He likewise ran the two together in a 2019 tweet, disavowing an earlier claim that he was the chosen one: 'When I looked up to the sky and jokingly said 'I am the chosen one,'... I was kidding, being sarcastic' (Woodward 2020). In a similar vein, Marjorie Taylor Greene called her 6th January remarks a 'sarcastic joke', again suggesting no important distinction between sarcasm and joking.

What is more, for my purposes in this paper, sarcasm is relevantly similar to joking. The two central features of joking utterances that I will discuss are (1) that what is explicitly uttered is not meant sincerely and (2) that they have an aesthetic dimension, in that they aim to be humorous. The first of these features is clearly shared by sarcasm. As Elisabeth Camp has pointed out, sarcasm and joking have the same 'profile of deniability', in that '[w]ith sarcasm and jokes, the speaker can... deny having meant Q [– any risky message insinuated by their utterance –]...but they can also insist they were 'just kidding' about L [– what they explicitly said]' (Camp 2022: 231).

The second relevant feature of joking utterances—their aesthetic dimension—is perhaps not a necessary feature of sarcasm, but it is fairly typical: one often makes a point sarcastically rather than sincerely for comic effect, and a failure to 'get' sarcasm is often treated as a kind of aesthetic failing. Trump certainly draws a connection between his supposed sarcasm and this aesthetic dimension: He backed up his claim that he was '[o]bviously... being sarcastic' about Clinton's emails by stating that 'a lot of people really smiled and laughed' when he made the comments. Insofar as a sarcastic utterance shares—or purports to share—this aesthetic dimension with joking, my remarks about joking as a strategy for achieving deniability 'on the cheap' will apply to it as well.

II. AESTHETIC DENIABILITY: JOKING AND DENIABILITY ON THE CHEAP

When Donald Trump claimed that he was 'just being sarcastic' about Jimmy Carter being dead, what effects might he have achieved? Presumably, the most significant effect would be to stop his audience believing that he sincerely asserted that Jimmy Carter was dead, and, more crucially, to stop them acting on this belief (where acting on the belief might include mental acts such as lowering their opinion of him or being amused by the mistake).

Based on this intuitive picture of the goal of denial, Alexander Dinges, and Julia Zakkou have recently proposed a conception of deniability in terms of what a speaker's audience know: 'S has deniability relative to the proposition that she meant to Φ if and only if: if S denies that she meant to Φ , then S's audience does not know that she meant to Φ ' (Dinges & Zakkou 2023: 385). So on this definition, Trump has deniability relative to the proposition that he meant to assert that Carter was dead if and only if, after his denial that he meant to assert that Carter was dead, his audience did not know that he meant to assert that Carter was dead. Cashing it out in terms of knowledge rather than belief allows us to capture the thought that intuitively, justification also matters for deniability: If my audience has good evidence that I did not mean to Φ , then I may have deniability relative to Φ even if some audience members ignore the evidence and doggedly insist that I meant to Φ .

An important upshot of this knowledge-based conception of deniability, as Dinges and Zakkou point out, is that there are two ways to undermine an audience's knowledge, and so it follows that there are two ways for a speaker to achieve deniability relative to some proposition. First, the speaker could undermine their audience's *evidence* that she meant to Φ . When the audience lacks the requisite evidence to know that the speaker meant to Φ , Dinges and Zakkou call this 'evidential deniability' (*Ibid.*: 389). Alternatively, the speaker could undermine the audience's *belief* that she meant to Φ . When the audience lacks the requisite belief to know that the speaker meant to Φ , Dinges and Zakkou call this 'psychological deniability' (*Ibid.*).

While it will typically be the case that an audience's belief that the speaker meant to Φ is undermined *because* their evidence is undermined, it is also possible for these to come apart: One can achieve deniability merely by getting one's audience to stop believing that one meant to Φ , even if the evidence very strongly supports the proposition that one meant to Φ . Moreover, there is good evidence that making salient alternative possibilities can irrationally affect people's beliefs. In a 2014 study conducted by Joshua Alexander, Chad Gonnerman, and John Waterman (Alexander, Gonnerman and Waterman 2014), participants were given two vignettes, originally described by Jennifer Nagel. In the first, 'plain', story, John 'is looking at a bright red table under normal lighting conditions. He believes the table is red' (Nagel 2010: 287).

The second, 'more detailed', story, is the same, except that another possibility is made salient: 'a white table under red lighting would look exactly the same to him, and he has not checked whether the lighting is normal, or whether there might be a red spotlight shining on the table' (*Ibid.*) Alexander, Gonnerman, and Waterman asked participants whether, in each story, John *knows* that the table is red. Perhaps unsurprisingly, they found that there was a statistically significant difference between participants' responses to the two stories: they were significantly more likely to say that John knows that the table is red in response to the plain story than the more detailed story. It appears then that people's conviction in a justified belief wavers—to the point that they would no longer classify it as knowledge—when alternative possibilities are made salient, even when the justification for the belief has not been undermined.

So when a speaker claims to have been 'being sarcastic', or 'only joking', are they primarily achieving evidential deniability, or (mere) psychological deniability? I believe that the aesthetic dimension of joking—that it purports to be *funny*—puts this kind of denial in a particularly strong position to undermine both an audience's evidence and, independently, their belief that the original utterance was sincere.

Let us start with evidential deniability. In typical cases, one has evidential deniability relevant to the proposition that one meant to Φ only if one speaks *indirectly*—only if one does not Φ explicitly. Take, for instance, one of Dinges and Zakkou's central cases:

Speeding Driver A driver stopped for speeding, to the police officer: 'I'm in a bit of a hurry. Is there any way we can settle this right now?'

Denial: 'I didn't mean to offer a bribe! I was just wondering if I could pay my fine right away.'

(Dinges & Zakkou 2023: 374, example originally from Lee & Pinker 2010: 790)

In this example, the driver insinuates that he is willing to bribe the police officer to avoid getting a speeding ticket. His ability to deny that he meant this insinuation comes from the fact that the bribe is not offered explicitly, giving him room to offer an alternative explanation for his explicit utterance. Moreover, the alternative explanation is plausible enough in the context that it does seem to undermine evidence that a bribe was being offered—wanting to pay a fine right away because one is in a hurry is somewhat reasonable, and asking to 'settle this right now' could plausibly be interpreted in this way.

However, it is sometimes possible to retain evidential deniability while speaking *directly*. Dinges and Zakkou do suggest some examples where one can retain deniability for direct speech, though they do not distinguish between evidential and psychological deniability for these cases:

think of situations in which we are prone to misspeak or where we can convincingly appeal to irony or metaphor, or where we can convincingly 'feign ignorance' (Stanley 2015: 156) of the literal meaning of a given word or sentence. Here, too, even semantically expressed contents can be deniable, not because the hearer does not know what the speaker said, but because she does not know whether the speaker meant what she said.

(Ibid.: 388)

So what would make *evidential* deniability possible in these direct cases, where the hearer does not know whether the speaker meant what she said? Presumably, the alternative explanation for their utterance must be objectively plausible, and equally or almost equally well supported by the evidence. If I explicitly Φ , but claim that I really meant to Ψ , my meaning to Ψ must fit well with the context of the utterance. If I claim I was being ironic, then it must be plausible that I meant the opposite of what I explicitly said; if I feign ignorance about the meaning of the word, it must be plausible that I was not aware of that meaning and took it to mean something else, which I could plausibly have meant in that situation.

Put more generally, the speaker's alternative explanation for what they meant must abide by what Paul Grice famously christened the 'Cooperative Principle' of conversation: 'Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged' (Grice 1991: 26). Whatever the speaker claims to have meant, this must be relevant to the question under discussion and appropriately informative, which places limits on the alternative explanations the speaker can plausibly offer.

What about claiming that one was joking? As I see it, there is a major difference between this and other forms of denial or disavowal. While in other cases, one's denial has to offer an alternative explanation for one's utterance that abides by the Cooperative Principle, in the case of joking, this is not so. For when one tells a joke, one's primary aim need not be the accepted purpose or direction of the talk exchange in which one is engaged; rather, one's primary aim can be to amuse one's interlocutors. In other words, the putatively aesthetic dimension of a joking utterance justifies it as an appropriate conversational move, even if it is not particularly relevant, informative, or even true. One has, as it were, poetic licence, and is not bound by the usual expectations of conversational contributions. Thus, one can get evidential deniability on the cheap, so long as one's original utterance can be claimed to have been an attempt at humour.

It might be objected that even when joking, one's utterance cannot be completely unrelated to the conversation at hand—we expect even humour to be somewhat topical to count as an appropriate conversational move. I think this

is not *always* the case, but by and large, it is true that a completely off-topic joke would be considered odd. However, this is not a problem for the sorts of denial at issue in my discussion here. For I am interested in cases where an utterance was originally uttered in a manner that appeared sincere, but is now being retracted. Such utterances will thus typically have been relevant to the conversation, which is why they were uttered in the first place.

A further feature of joking that makes this way of achieving deniability particularly devious is that the funniness of an utterance does not actually entail its insincerity. Rather, it is perfectly possible to speak amusingly yet sincerely. Take, for instance, this interaction George Bush had with reporters at a 2004 news conference:

Mr. Bush: Now that I've got the will of the people at my back, I'm going to start enforcing the one-question rule. That was three questions...

Q. Thank you, Mr. President. How will you go about bringing people together? Will you seek a consensus candidate for the Supreme Court if there's an opening? Will you bring some Democrats into your cabinet?

A. Again, you violated the one-question rule right off the bat. Obviously you didn't listen to the will of the people.

(Reuters 2004)1

Here Bush says something funny, but his chastisement of the speaker is likely to be perceived as sincere. This kind of sincere joking, sometimes called 'kidding on the square', demonstrates that funniness does not necessarily show that an utterance was insincere. More importantly, it opens the door to a uniquely powerful kind of deniability. One can make an utterance that is both explicit and sincere, and so long as it can pass as funny, this funniness can give one evidential deniability for it if one wants to disavow it at a later time, while allowing one to avoid the acrobatics usually required to provide an alternative explanation for one's utterance that abides by the Cooperative Principle.

Of course, to attain this kind of evidential deniability, one's utterance does still need to pass as funny—it needs to be reasonably clear what was supposed to be amusing about it. In many cases, such as Trump's gaffes about Jimmy Carter and disinfectant, this is simply not the case. It is difficult to see what the joke could have been, if Trump was indeed joking. So does this prevent him from attaining deniability for his utterances? Not necessarily. For it is here that psychological deniability comes into its own.

¹ My attention was drawn to this example by a talk given by Elisabeth Camp in 2021 entitled 'Just Kidding: Wilful Deniability in Speech'.

As we saw above, a speaker can still attain psychological deniability relative to the proposition that she meant to Φ so long as she can undermine her hearers' belief that she meant to Φ , even if her hearers would be justified in believing that she meant to Φ based on her evidence. Moreover, we have already seen that it is sometimes possible to achieve this just by introducing alternatives—that the speaker meant to Ψ instead — which make hearers irrationally give up on their beliefs.

This phenomenon, whereby one convinces hearers to give up on beliefs to which they are entitled, is curiously similar to Kate Manne's recent account of gaslighting, which she defines as 'the process of making someone feel defective in some of the most fundamental ways (for example, morally or rationally) for having (or for that matter lacking) mental states that she is in fact entitled to have (or lack)' (Manne 2023: 139). The central difference is, of course, that one could convince one's hearers to give up on their belief that one meant to Φ without making them feel defective in some fundamental way. However, it is clear that gaslighting is one effective way of achieving psychological deniability.

What is more, Manne's account of gaslighting offers a clue as to how claiming that one was joking might provide a particularly powerful way for speakers, including political speakers, to convince their audience that they did not mean what they said. In earlier accounts of gaslighting, the focus has tended to be on the way in which gaslighters convince their victims that they are *rationally* defective for having a mental state to which they are entitled (e.g. Abramson 2014). However, Manne's key insight is that some gaslighters instead convince their victims that they are *morally* defective for having a mental state to which they are entitled; for instance, the scammer known as 'Dirty John' convinced his wife to give up justified beliefs she had about his chequered past by making her out to be

a bad person when she challenged or withdrew from him, and a good one for believing him. He operated with both a moral stick—the prospect of his condemning [her]...—and a moral carrot—the prospect of him celebrating her as a wonderful wife, a forgiving person, the love of his life, and so on.

(Manne 2023: 128).

In both the rational and the moral case, the gaslighting is successful because the criticism—that the hearer is either irrational or immoral—has 'real bite: It is hard to credit the possibility that someone might be gaslit by the prospect of being envisaged (it does not seem right to say 'written off') as a little bit silly, just slightly ungenerous, or a tad over-cautious' (*Ibid.*: 138). The gaslighting can only succeed if the criticism stings enough to make the victim willing to give up a mental state to which they are entitled, which is no mean feat.

IO Z. WALKER

One reason why Manne's point about moral gaslighting is important for our purposes here is that is suggests the possibility for gaslighting to move out of the interpersonal context and onto the political stage. As she points out, when one focuses on rational gaslighting, it is implausible to suggest that a politician such as Donald Trump might be capable of gaslighting his base—though many people have claimed that he has (e.g. Woodward 2020). This would require Trump to convince big groups of people that they were irrational and out of touch with reality, which would be rather difficult to achieve. However, morally gaslighting big groups of people is much more plausible, given that, as Manne observes,

groups of agents can be made to feel guilty or ashamed for their beliefs with relative ease: If you inspire loyalty in a group, then a savvy political operator can weaponise that loyalty to make its members strongly inclined to stick to the party line, echo the claims of their leader, defend the leader, and so on.

(Manne 2023: 141).

Thus, the possibility of other forms of gaslighting, which do not require convincing hearers that they are rationally defective, makes this phenomenon more relevant to discussions of political speech, as we need not think of gaslighting as a solely interpersonal phenomenon.

However, the other reason Manne's point about moral gaslighting is instructive for our purposes is that it suggests the possibility of another form of gaslighting that could be performed by speakers who claim that they were joking; namely, speakers could perform *aesthetic* gaslighting, whereby they make hearers feel aesthetically defective—lacking in a good sense of humour—for believing that their original utterance was sincere.²

This, I propose, is what a speaker can do if they are not in a position to use joking to attain *evidential* deniability of an earlier utterance, because it was not sufficiently funny. In this situation, a speaker can insist that their utterance was funny, and that anyone who thinks otherwise is a humourless boor. For those hearers who wish to be held in high esteem by the speaker, such as, for example, Trump's base, a failure to share his aesthetic sensibilities, including his comic tastes, would, I contend, have sufficient bite to convince some of them to give up their belief that his initial utterance was sincere. In many circles, it seems to me, being aesthetically defective—unfunny, and thus more generally uncool—is considered a worse fate than being immoral, and it is noteworthy that Trump frequently calls his opponents 'losers' rather than morally condemning them.

² In her paper, Manne does not discuss the possibility of aesthetic gaslighting. However, when she delivered this paper at the 97th Joint Session Conference, she did suggest that aesthetic gaslighting may be a part of diet culture, whereby, for instance, people are made to feel ugly and slovenly if they want to eat in particular ways.

This aesthetic gaslighting is what Trump may be achieving when he claims that CNN 'DON'T GET SARCASM': There is a suggestion that reporters who take his claims seriously are lacking in some capacity, and other hearers of his must either share this defect, or share his good sense of humour, and thus see that he was joking. In this way, persistent moral and aesthetic gaslighting allows a political speaker to create a vast gulf of both moral and aesthetic value between their targets—their victims—and the people around them who do not buy into the gaslighting.

A similar gaslighting move appears in Silvio Berlusconi's disavowals of earlier claims as jokes. After receiving heavy criticism in 2022 for telling Monza footballer players that if they won, he'd 'get a bus full of whores to come to [their]... locker room', he responded that perhaps it was his critics' 'utter lack of humor that makes them so sad and yet so gratuitously mean' (Giordano 2022), painting them as defective for failing to appreciate that he was joking.

And in a 2019 profile on Boris Johnson for the *New York Review of Books*, Fintan O'Toole accuses Johnson of making use of this same strategy:

In England at least... knowingness is essential to being included. You have to be "in on the joke"—and Johnson has shown just how far some English people will go in order not to look like they are not getting it... Johnson has played on this to perfection—he knows that millions of his compatriots would rather go along with his outrageous fabrications than be accused of the ultimate sin of taking things too seriously.

(O'Toole 2019).

Johnson, like Trump and Berlusconi, uses his audience's fear of lacking a good sense of humour to gaslight them into believing that he was joking.

A final important feature of joking worth commenting on in relation to deniability is that whilst with most kinds of denial, repeated denials undermine the plausibility of the denial—one can only feign ignorance so many times—when it comes to joking, repeated claims that one was joking can actually increase the plausibility of the denial, if one gets the reputation for being a joker who frequently speaks insincerely. Thus, there can be value to insisting that one was joking even in the less plausible cases, as it can still be part of cultivating a more general insincere persona.

Of course, once one has developed this kind of reputation, it can be difficult to then return to being taken seriously. Journalist Edward Docx, writing for *The Guardian*, suggests that this is the fate of Boris Johnson, who spent his entire career cultivating a persona as 'the archetypal clown, with his antic posturing and his refusal to take anything seriously', but later attempted to 'creep away from his clownish past altogether':

I2 Z. WALKER

The difficulty for the clown is that once truth and seriousness have been merrily shattered, they cannot be put back together and served up anew. Or, to put it another way, the buffoon who has just entertained the audience by smashing all the plates cannot now say that he proposes to use them to serve up a banquet in honour of himself becoming a wise and honest king. Everyone can see: the plates are all in pieces on the floor.

(Docx 2021).

The 'I was joking' defence, then, though it may appear to be a risk-free way of attaining deniability, may in fact involve taking a risk after all: the risk of becoming an eternal clown, trapped in perpetual insincerity—the buffoon that cried wolf.

III. THE MASK OF LEVITY: JOKING AND UNDER-THE-RADAR INFLUENCE

So far, we have focused on the AIDCs, and seen why it might be effective to disavow them as jokes, even though the denial is—on the face of it—highly implausible. However, what about the PIDCs, whose disavowal will only be found plausible by the very people they are meant to influence? In this section, I will argue that getting one's target audience to believe that one was joking about these Pro-Interest Claims does not prevent the claims from being influential—in fact, in some circumstances it could be a more effective way of influencing hearers than simply making the claims sincerely.

To see this, it will be helpful to start by introducing a distinction from J. L. Austin between three things one does with an utterance. Take, for example, the utterance 'Jimmy Carter is dead'. The literal meaning of what one says—ascribing the property of being dead to the person Jimmy Carter—is one's *locutionary* act. The force of the utterance—as an assertion, a threat, or a promise—is one's *illocutionary* act. And the effects the utterance has—making people believe that Jimmy Carter is dead, upsetting them, making them laugh scornfully at one's mistake—are one's *perlocutionary* acts (Austin 1975: Lecture VIII).

When one claims to be 'only joking' or 'being sarcastic', which of these acts is being disavowed? It would be difficult to disavow the locutionary act, unless one could plausibly deny have uttered those specific words—perhaps if one had mumbled, or was speaking in a loud room. Certainly, 'I was only joking' is no denial that one uttered those words—in fact, it acknowledges that they were indeed uttered. Rather, what one denies in claiming to have been joking is that one performed the *illocutionary* act one appeared to have performed—one did not sincerely *assert* that Jimmy Carter was dead, or did not sincerely *ask* whether injecting disinfectant could cure Covid. Indeed, Austin himself says

that for an illocutionary act to be successfully performed, 'the words must be spoken "seriously" and so as to be taken "seriously"... I must not be joking, for example' (*Ibid.*: 9).

What follows from the successful disavowal of an illocutionary act? Primarily, one undoes the normative changes that would have occurred if one's illocutionary act had been felicitous. For instance, a felicitous assertion entitles hearers to believe that the speaker believes the contents of the assertion, but they lose this entitlement if the illocutionary act of assertion is successfully disavowed. Similarly, a felicitous question makes it appropriate for hearers to offer a response, whereas this is not appropriate if the question was infelicitous.

Are the perlocutionary effects of the utterance likewise undone if one successfully disavows one's utterance as 'only joking' or 'being sarcastic'? In many cases, if a hearer accepts the disavowal, perlocutionary effects—such as being amused or offended by the utterance, or forming new beliefs based on it—will cease, as the speech act that justifies them turns out not to have happened. However, this is not the same as these effects never having happened—it is still true that the initial utterance caused offence or amusement or the temporary adoption of new beliefs.

Moreover, not *all* perlocutionary effects will necessarily cease after the illocutionary act that caused them has been disavowed. One such effect, I contend, is the raising in salience of a particular perspective. When a joke is made about Hilary Clinton having connections with Russia, or about Barack Obama having connections with ISIS, the salience of this possibility is raised in people's minds. And once someone's attention has been drawn to this possibility, the fact that the connection was raised in jest does nothing to lower this salience. Indeed, it is remarkably difficult to lower the salience of something that has been raised, even if one wants to—as it is often said, one cannot unring a bell.

Nonetheless, one might think that we need not worry about this raising of salience. Merely being aware of, and attending to, a supposed connection is not the same as actually believing it, and so long as people do not genuinely believe that Clinton and Obama are connected to Russia and ISIS in this way, their subsequent political decisions and actions will not be affected. Given that the original illocution was claimed to be a joke, hearers are not entitled to form any new beliefs on its basis, so no harm ought to be done.

Unfortunately, this optimistic view of human behaviour and decision-making underestimates the extent to which these are influenced by what is made salient to the person in question. First of all, it is well documented that adopting a fictional perspective that makes certain concepts salient to someone can have a lingering effect on their behaviour (see e.g. Anderson and Pichert 1978; Banfield *et al.* 2003; Bargh *et al.* 1996). Thus, the fact that a

³ In making this point, I follow Camp (2017).

I4 Z. WALKER

perspective is not actually endorsed—as is the case when it is said as a joke—does not prevent it from influencing people's actions.

Moreover, when it comes to conscious political decision-making, salience plays a significant role. A large part of our thinking involves making inferences from one proposition to another: We might infer from the fact that a certain politician voted against gay marriage to the proposition that they will not support LBGTQ rights in the future, or from the fact that a politician chose a fundamentalist Christian as their running mate to the proposition that they are sympathetic to fundamentalist Christianity. Sometimes, the inferences we make will preserve the truth, and sometimes they will not. However, a crucial factor in whether or not we make them is how cognitively accessible they are: how much mental effort it takes us to move from the input proposition to the output proposition. The less mental effort it takes, the more likely we are to make a particular inference, and vice versa.

So, what determines the cognitive accessibility of a particular inference? As Rachel Fraser has argued in relation to metaphor, one factor that increases the cognitive accessibility of an inference is if a perspective according to which the inference is licensed has been made salient (Fraser 2018). If you have been entertaining a perspective that foregrounds connections between two things, then it will require less imaginative effort for you to make inferences based on said connections. For instance, if you have been entertaining a perspective that foregrounds connections between Barack Obama and ISIS, then the inference from voting for Obama to supporting ISIS will be more cognitively accessible to you than it otherwise would be.

It turns out, then, that just because one was joking, one's utterance can still influence hearers' decision-making. Disavowing an utterance as a joke may undo the illocutionary act, but it will not necessarily prevent the perlocutionary effects of the utterance, and these effects can be influential. What I now want to show is an even more surprising effect of joking: that successfully disavowing an utterance as a joke can in some circumstances make the influence on hearers even more powerful than if the utterance had not been disavowed. In particular, when a joke relies on a sexist or racist ideology without subverting it, this can more strongly influence people to act on pre-existing sexist or racist attitudes than hearing non-humorous sexist or racist statements would.

To see how this might be so, it will be helpful to consider another kind of political speech act that has received more attention than joking: dogwhistling. In the typical case, a dogwhistle is

a speech act designed, with intent, to allow two plausible interpretations, with one interpretation being a private, coded message targeted for a subset of the general audience, and concealed in such a way that this general audience is unaware of the existence of the second, coded interpretation.

(Witten forthcoming: 2, as discussed in Saul 2018: 362).

One example of a dogwhistle is the term 'wonder-working power', a fundamentalist Christian term that refers to the power of Christ. Politicians use this term to signal to fundamentalist Christian hearers that they are sympathetic to them, without putting off non-fundamentalist hearers who would not pick up on this connotation (Saul 2018: 362). Another example is the term 'inner city', which ostensibly describes a geographical location, but has in the USA come to have racialised connotations: The term 'inner city criminals', for instance, has come to signal 'black criminals' to many hearers, and allows politicians to appeal to racist voters without saying anything explicitly racist (*Ibid.*: 367).

While fundamentalist Christians are well aware of the hidden connotations of 'wonder-working power', most people are not consciously aware of the racial connotations of 'inner city': it is, in Jennifer Saul's terminology, a 'covert' dogwhistle (*Ibid.*: 361). This makes its potential use as a political tool rather puzzling. If no one consciously picks up on these racial connotations, how could these connotations influence them?

The first thing to say is that experimental data shows that the racial connotations of the term *do* influence people's thinking. In a 2005 study conducted by Jon Hurwitz and Mark Peffley, 'a random half of respondents was asked whether they support spending money for prisons (versus antipoverty programmes) to lock up "violent criminals," while the other half was asked about "violent inner city criminals" (Hurwitz and Peffley 2005: 99). The study found that 'whites' racial attitudes (e.g. racial stereotypes) were much more important in shaping preferences for punitive policies when they receive the racially coded, 'inner city' question' (*Ibid.*). In other words, subjects with pre-existing racist attitudes were much more likely to support spending money for prisons to lock up violent criminals when the criminals in question were described as 'inner city' criminals than when this dogwhistle was not used, suggesting that the term somehow subconsciously influenced their thinking. Likewise, subjects with pre-existing anti-racist attitudes were more likely to oppose this spending when the term 'inner city' was used.

What is more, there is evidence that this effect is far reduced when racist attitudes are cued explicitly rather than implicitly. In a 1993 experiment, Tali Mendelberg investigated the difference in effects between explicitly racist, implicitly racist, and counter-stereotypic priming on subjects with and without pre-existing implicitly racist attitudes. She found that implicitly racist priming had a significantly greater effect on subjects with pre-existing implicit racist attitudes support for policies than either explicitly racist or counter-stereotypic priming, suggesting that the covert nature of racist dogwhistles is key to their effectiveness (Mendelberg 2001: chapter 7).

Mendelberg's findings are important for our purposes because they indicate that raising the salience of a particular ideology without endorsing it influences behaviour more than explicitly endorsing it does—which is exactly what happens when a particular ideology features in a joke. We might

therefore expect that, just as covert racist dogwhistles influence behaviour more than explicit racism does, racist jokes influence behaviour more than explicit racism does—and likewise for other ideologies.

And indeed, a 2008 study into sexist humour supports this hypothesis. Thomas Ford and other researchers began by testing participants for preexisting sexist attitudes, in a manner similar to how Mendelberg tested her participants for pre-existing racist attitudes, by testing their agreement with implicitly sexist statements (Ford et al. 2008). They then gave these participants vignettes containing either sexist statements, sexist jokes, or non-sexist jokes, and asked them to imagine that they were in the situations described in the vignettes, hearing the statements or jokes as they occurred. Finally, participants were asked to imagine that they were in a newspaper group being asked for donations by the National Council of Women, and had to report how much money they would be willing to donate. The study found that participants with pre-existing sexist attitudes who had read the sexist jokes were significantly less willing to donate money to the National Council of Women than those who had read either the non-sexist jokes or the sexist statements a result that closely matches Mendelberg's findings regarding implicit racist priming. Thus, the surprising takeaway is that not only can sexist humour influence behaviour, it can have a greater influence over behaviour than nonhumourous sexism—and, judging by Mendelberg's results, this is likely to be true of racist humour as well.

At this point, we might well wonder how this odd phenomenon is to be explained, whereby covert raising of salience of sexist or racist ideology has a greater influence on behaviour than explicit sexism or racism. An explanation proposed by Mendelberg and later developed by Saul is that although many people in the USA still have implicit racist attitudes, there is an increasingly prevalent 'Norm of Racial Equality' in the USA, according to which it is unacceptable to explicitly express racist attitudes. (It is worth mentioning, as Saul does, that '[t]he only kind of racial equality this commits one to is an extremely thin sort of formal equality... [which] seems to preclude the use of obvious pejoratives, assertions of *genetic* (though not cultural) inferiority, and support for obviously discriminatory behaviour (legally enforced segregation, rules against hiring black people, etc.)' (Saul 2018: 365.) This means that politicians who want to appeal to voters who have implicitly racist attitudes cannot make explicitly racist claims, as this will put them in violation of the norm, rendering what they say widely unacceptable, and preventing it from influencing hearers' thinking and behaviour. However, covertly raising the salience of racist perspectives, using dogwhistles like 'inner city', allows them to appeal to the implicit racist attitudes of these voters without violating the norm (Mendelberg 2001; Saul 2018).

Ford *et al.* also hypothesise that norms explain their surprising result, although on their account it is norms of humour that play a central role. Ford's proposal, originally put forward in an earlier paper, is that humourous

communication 'activates a conversational rule of levity', whereby the audience 'tacitly consents to an implicit normative standard communicated by the humour that, in this context, one need not be critical of discrimination against the target group' (Ford and Ferguson 2004: 81–2). The audience then uses this norm to self-regulate, and if they have existing prejudicial attitudes towards the target group, go on to act on these, as the new norm deems this to be acceptable.

Bringing both of these explanations together will, I believe, give us a promising theory as to why sexist humour sometimes influences behaviour more strongly that non-humourous sexism. Following Saul and Mendelberg's proposal of an increasingly prevalent 'Norm of Racial Equality', I propose that there is also an increasingly prevalent 'Norm of Gender Equality', according to which it is unacceptable to explicitly express sexist attitudes. Explicit non-humourous sexist statements violate this norm, and so are rejected by hearers and do not influence their subsequent thinking and behaviour. However, there are two ways around this norm. One is to raise sexist ideology to salience inexplicitly, so that the norm is not violated and people are not aware that they might be being influenced. This is what happens with covert dogwhistles. Another is to raise it explicitly, but at the same time as introducing a norm of levity: a localised norm that temporarily relaxes a number of widespread social norms concerning politeness and acceptability, among them the 'Norm of Gender Equality'. This relaxing of the norm thereby renders gender discrimination acceptable in that context, and may then have a lingering effect, as demonstrated by Ford's (2008) study, whereby hearers go on to think and act in ways that would not be permitted by the Norm of Gender Equality after they have engaged with the humour.

In the introduction to this paper, I mentioned a 2002 column written by Boris Johnson in which he used racist language about black Africans, but disavowed it as 'wholly satirical'. If my argument in this section is correct, then the satirical, comedic tone of Johnson's allusions to racist ideologies in that column will not have prevented their harmful perlocutionary effects—rather, the satirical tone will have made the column *more* likely to have a racist influence on the behaviour of readers with pre-existing implicit racist attitudes because this tone introduced a norm of levity that relaxed the Norm of Racial Equality for those readers.

At this point, one might wonder whether we can expect *all* jokes that rely on sexist or racist ideology, in *all* contexts, to have this kind of negative effect. Many people think that jokes that subvert sexist ideology, or sexist jokes made amongst feminists, are more acceptable than unsubverted sexist humour that appears in non-feminist contexts, so can we vindicate this thought on my account?

I would suggest that in these problem cases, while there is a sense in which sexist—or, similarly, racist—ideology is being made salient, the main thing that is being made salient is the absurdity of these ideologies. When one uses

humour to subvert sexist or racist stereotypes, one is performing a sort of *reductio ad absurdum*, demonstrating that the stereotypes are false, or not fitting. This, I suspect, will prevent the humour from activating pre-existing sexist or racist attitudes, as the perspective it offers is not the same one as the one of which those attitudes are a part. Similarly, in a feminist context, the absurdity of the sexism in question will typically be salient to all interlocutors, and thus disrupt the perspective needed to activate the pre-existing sexist attitudes. Moreover, interlocutors will tend to have fewer and weaker pre-existing attitudes in such a context, meaning that a prerequisite for the harmful behaviour is absent.

In most cases of political speech, however, when a speaker has a diverse audience with different views, there will be some—often, many—hearers who have the requisite pre-existing sexist or racist attitudes. This means that absent any subversion of sexist or racist ideology, telling sexist or racist jokes will influence those hearers to act on those sexist or racist attitudes—not just *despite* the fact that they are jokes, but in part *because* they are jokes.

IV. CONCLUSION

In this paper, I have considered two functions of the concept of joking in political speech: claiming to be joking as a way to attain deniability for earlier utterances on the cheap, and claiming to be joking as a way to influence attitudes under the radar. In Section II, I showed that claiming that an earlier utterance was a joke can be a way to secure evidential deniability for that utterance 'on the cheap', as the aesthetic dimension of a joke gives a speaker poetic licence, meaning that they do not need to provide an alternative explanation for their utterance that abides by the Cooperative Principle. What is more, if a joke lacks this aesthetic dimension, the speaker can engage in aesthetic gaslighting, whereby the prospect of lacking a good sense of humour is used to convince hearers to give up their belief that the original utterance was sincere, thus allowing the speaker to achieve psychological deniability.

In Section III, I showed that claiming that one's utterance was a joke does not necessarily prevent it from influencing hearers' behaviour and decision-making, as the perlocutionary effect of raising things to salience is not undone by disavowing the illocutionary act of assertion. What is more, it looks as though when humour raises the salience of sexist or racist ideology in particular, this can in some contexts have a greater negative effect on an audience's behaviour than non-humourous sexism or racism, because it introduces a norm of levity that relaxes widespread norms of gender and racial equality.⁴

⁴ Many thanks to two anonymous referees of this journal for their helpful comments on this paper, to Rae Langton, Jessie Munton and Anne Eaton for discussions of earlier work that inspired this paper, and to my colleague Stephen Fisher for suggesting useful political examples.

REFERENCES

- Abramson, K. (2014) 'Turning Up the Lights on Gaslighting', *Philosophical Perspectives*, 28: 1–30. https://doi.org/10.1111/phpe.12046
- Alexander, J., Gonnerman, C., and Waterman, J. (2014) 'Salience and Epistemic Egocentrism: An Empirical Study', in J. R. Beebe (ed.) *Advances in Experimental Epistemology*, 97–117. New York, NY: Bloomsbury Academic.
- Austin, J. L. (1975) *How to Do Things with Words*, 2nd edn. J. O. Urmson and M. Sbisà (ed.). Oxford: Clarendon.
- Anderson, R. and Pichert, J. (1978) 'Recall of Previously Unrecallable Information Following a Shift in Perspective', *Journal of Verbal Learning and Verbal Behavior*, 17: 1–12. https://doi.org/10.1016/S0022-5371(78)00485-1
- Banfield, J. et al. (2003), 'The Effects of an Elderly Stereotype Prime on Reaching and Grasping Actions', Social Cognition, 21/4: 299–319. https://doi.org/10.1521/soco.21.4.299. 27002
- Bargh, J., Chen, M., and Burrows, L. (1996) 'The Automaticity of Social Behavior: Direct Effects of Trait Construct and Stereotype Activation on Action', *Journal of Personality and Social Psychology*, 71/2: 230–44. https://doi.org/10.1037/0022-3514.71.2.230
- Bennett, J. T. (2020) 'Trump Makes False Claims About His 'Russia if you're listening' Plea During Pennsylvania Rally', *The Independent*, 27 September (https://www.independent.co.uk/news/world/americas/us-election/trump-rally-pennsylvania-lies-fact-check-russia-hillary-clinton-emails-wikileaks-b633765.html), accessed October 2023.
- Camp, E. (2021) 'Just Kidding: Wilful Deniability in Speech', at HaLO Workshop: Oppressive Speech, Societies & Norms theme two: "Silencing, Speaking up & Free Speech"
- (2022) 'Just Saying, Just Kidding: Liability for Accountability-Avoiding Speech in Ordinary Conversation, Politics and Law', in L. R. Horn (ed.) From Lying to Perjury: Linguistic and Legal Perspectives on Lies and Other Falsehoods, 227–58, Berlin: De Gruyter Mouton.
- —— (2017) 'Why Metaphors Make Good İnsults: Perspectives, Presupposition, and Pragmatics', *Philos Stud*, 174/1: 47–64. https://doi.org/10.1007/s11098-015-0525-y
- CBS News Staff (1999) 'Clinton: Just Joking!', CBS News, 14 July https://www.cbsnews.com/news/clinton-just-joking/) accessed November 2023.
- Clymer, A. (1999) 'Clinton Joins in Gore's Call for Licensing of Gun Owners', New York Times, 15 July (https://www.nytimes.com/1999/07/15/us/clinton-joins-in-gore-s-call-for-licensing-of-gun-owners.html) accessed November 2023.
- Day, M. (2012) 'Leave the Euro, Silvio Berlusconi Tells Italy—Then Says He was Only Joking', *The Independent*, 4 June (https://www.independent.co.uk/news/world/europe/leave-the-euro-silvio-berlusconi-tells-italy-then-says-he-was-only-joking-7814813.html) accessed November 2023.
- Dinges, A. and Zakkou, J., (2023) 'On Deniability', *Mind*, 132/526: 372–401. https://doi.org/10.1093/mind/fzaco56
- Docx, E. (2021) 'The Clown King: How Boris Johnson Made It By Playing the Fool', *The Guardian*, 18 March (https://www.theguardian.com/news/2021/mar/18/all-hail-the-clown-king-how-boris-johnson-made-it-by-playing-the-fool) accessed June 2021
- Ford, T. E. et al. (2008) 'More Than "Just a Joke": The Prejudice-Releasing Function of Sexist Humor', *Personality and Social Psychology Bulletin*, 34/2: 159–70. https://doi.org/10.1177/0146167207310022
- Ford, T. E. and Ferguson, M. A. (2004) 'Social Consequences of Disparagement Humor: A Prejudiced Norm Theory', *Personality and Social Psychology Review*, 8/1: 79–94. https://doi.org/10.1207/S15327957PSPR0801_4
- Forrest, A. (2019) 'Boris Johnson Says Describing Black People as Having "watermelon smiles" was "wholly satirical", *The Independent*, 30 June (https://www.independent.co.uk/news/uk/politics/boris-johnson-conservative-leadership-latest-racism-watermelon-smiles-satirical-a8981166.html) accessed November 2023.
- Fraser, R. (2018) 'The Ethics of Metaphor', *Ethics*, 124/4: 728–55. https://doi.org/10.1086/697448

- Giordano, E. (2022) 'Berlusconi Promises Footballers a "bus full of whores", *Politico*, 14 December (https://www.politico.eu/article/silvio-berlusconi-italy-football-promise-bus-whores/) accessed November 2023.
- Grice, H. P. (1991) Studies in the Ways of Words: Paul Grice. Cambridge, MA: Harvard University Press.
- Hurwitz, J. and Peffley, M. (2005) 'Playing the Race Card in the Post-Willie Horton Era: The Impact of Racialized Code Words on Support for Punitive Prison Policy', *Public Opinion Quarterly*, 69/1: 99–112. https://doi.org/10.1093/poq/nfi004
- Lee, J. J. and Pinker, S. (2010) 'Rationales for Indirect Speech: The Theory of the Strategic Speaker', *Psychological Review*, 117/3: 785–807. https://doi.org/10.1037/a0019688
- Manne, K. (2023) 'Moral Gaslighting', Aristotelian Society Supplementary Volume, 97/1: 122–45. https://doi.org/10.1093/arisup/akadoo6
- Mendelberg, T. (2001) The Race Card: Campaign Strategy, Implicit Messages, and the Norm of Equality. Princeton, NJ: Princeton University Press.
- Nagel, J. (2010) 'Knowledge Ascriptions and the Psychological Consequences of Thinking about Error', *Philosophical Quarterly*, 6o/239: 286–306. https://doi.org/10.1111/j.1467-9213. 2009.624.x
- O'Toole, F. (2019) 'The Ham of Fate', The New York Review, 15 August. (https://www.nybooks.com/articles/2019/08/15/boris-johnson-ham-of-fate/) accessed July 2023.
- Reuters. (2004) 'Transcript of President Bush's News Conference', *The New York Times*, 4 November (https://www.nytimes.com/2004/11/04/politics/transcript-of-president-bushs-news-conference.html) accessed July 2023.
- Saul, J. (2018) 'Dogwhistles, Political Manipulation, and Philosophy of Language', in D. Fogal, D.W. Harris and M. Moss (eds.) New Work on Speech Acts, 360–83. Oxford: OUP.
- Stanley, J. (2015) How Propaganda Works. Princeton, NJ: Princeton University Press.
- Steakin, W. (2022) 'Marjorie Taylor Greene Says If She Ran Jan. 6 Capitol Attack, "We would have won", ABC News, 12 December (https://abcnews.go.com/Politics/marjorie-taylor-greene-jan-6-capitol-attack-wouldve/story?id=95065444) accessed November 2023.
- Stein, S. and Arciga, J. (2018) 'Kellyanne Conway: I Was Just Joking When I Said James Comey "Swung an Election", *The Daily Beast*, 16 April (https://www.thedailybeast.com/kellyanne-conway-james-comey-interview-was-revisionist-version-of-history) accessed November 2023.
- Woodward, A. (2020) 'Six Times Donald Trump Used "Sarcasm" to Gaslight America', *The Independent*, 27 April (https://www.independent.co.uk/news/world/americas/us-politics/donald-trump-sarcasm-twitter-jokes-gaslight-inject-disinfectant-a9486966.html) accessed luly 2023.

Trinity College, Oxford, UK

[©] The Author(s) 2023. Published by Oxford University Press on behalf of The Scots Philosophical Association and the University of St Andrews. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.