

Virtue Signalling and the Condorcet Jury Theorem

Scott Hill and Renaud-Philippe Garner

Abstract

One might think that if the majority of virtue signallers judge that a proposition is true, then there is significant evidence for the truth of that proposition. Given the Condorcet Jury Theorem, individual virtue signallers need not be very reliable for the majority judgment to be very likely to be correct. Thus, even people who are skeptical of the judgments of individual virtue signallers should think that if a majority of them judge that a proposition is true, then that provides significant evidence that the proposition is true. We argue that this is mistaken. Various empirical studies converge on the following point: humans are very conformist in the contexts in which virtue signalling occurs. And stereotypical virtue signallers are even more conformist in such contexts. So we should be skeptical of the claim that virtue signallers are sufficiently independent for the Condorcet Jury Theorem to apply. We do not seek to decisively rule out the relevant application of the Condorcet Jury Theorem. But we do show that careful consideration of the available evidence should make us very skeptical of that application. Consequently, a defense of virtue signalling would need to engage with these findings and show that despite our strong tendencies for conformism, our judgements are sufficiently independent for the Condorcet Jury Theorem to apply. This suggests new directions for the debate about the epistemology of virtue signalling.

Background

Tosi and Warmke¹ offer an account of paradigmatic moral grandstanding: they identify what is common to the central cases of grandstanding while allowing that peripheral cases may lack these commonalities. They employ three technical terms. First, *recognition desire* is the desire to be perceived as morally respectable. Second, *grandstanding expressions* are contributions to public discourse about morality. Third, they discuss an *in order to* relation. One does *x* in order to *y* if and only one would be disappointed if one did *x* and *y* did not occur. For them, one engages in paradigmatic grandstanding if and only if one contributes a grandstanding expression to public discourse in order to satisfy a recognition desire.

They also offer a typology. *Piling on* occurs when a moral judgment has already been made and one registers agreement. Suppose I notice Facebook friends condemning President Trump for irresponsibly speculating about using bleach to cure the coronavirus. To pile on, I would create a post seconding those condemnations. *Ramping up* occurs when others have made a moral judgment and one offers a revised and more extreme one. Suppose I notice Facebook friends condemning Speaker Pelosi for tearing up a copy of the President's speech. To ramp up, I agree and add that she deserves to be fired or imprisoned. *Trumping up* occurs when one claims that a blameless act is worthy of condemnation. For example, it seems harmless to dress as a little green alien for Halloween. Yet, if I object that it unfairly stereotypes and appropriates alien culture then I have engaged in trumping up. Two other instances are *excessive outrage* and appeals to *self-evidence*.

Note that for Tosi and Warmke, these acts only count as moral grandstanding if they are done for the sake of status. Motivation is the differentiating factor. And again, Tosi and Warmke, use 'in order to' in a

¹ Grubbs, J., Warmke, B., Tosi, J., Campbell, K. (2019) Moral Grandstanding in Public Discourse: Status-Seeking Motives As a Potential Explanatory Mechanism in Predicting Conflict *PLOS ONE* 1-31

Tosi, J. and Warmke, B. (2016) Moral Grandstanding *Philosophy and Public Affairs* 197-217

Tosi, J. and Warmke, B. (2020) *Grandstanding: The Use and Abuse of Moral Talk* Oxford University Press

technical sense to mean that if I believe that the relevant acts didn't raise my moral respectability in the eyes of others, then I would be disappointed. Grandstanders speak for the sake of status.

We stress that grandstanding is not about whether one's judgment is correct. Trump deserves condemnation for reckless medical speculation and Pelosi only added to polarization. Tosi and Warmke do not argue that one should never grandstand or that one should never act for the sake of one's status. In any case, mixed motives are incredibly common. And pure motives might be too much to ask.

Furthermore, it is important to be clear about exactly what Tosi and Warmke worry about in connection with grandstanding. They are mainly worried about its social consequences: they argue that using moral discourse for the sake of status leads to cynicism, outrage exhaustion, and group polarization.

Finally, we should note that we think the Tosi and Warmke account is too narrow in a certain way. We think that a recognition desire need not be present in all core stereotypical cases of virtue signalling. We think that the phenomenon of virtue signalling is better captured by allowing other psychological processes, conscious and unconscious, that push us to conform. This will come out in our discussion below. And we'll say more about it as we address objections at the end of the paper.

Levy on Virtue Signalling

Neil Levy² accepts Tosi and Warmke's analysis of moral grandstanding as equivalent to virtue signalling. In defense of virtue signalling, Levy offers an original argument: signalling contributes to public moral discourse by providing higher-order evidence. If we want to know something, it is useful to know what others think. Indeed, Levy reminds us that most of our knowledge depends on the testimony of others. He adds that there are features of testimony that enhance its credibility: confidence and numbers. On the value of confidence, consider:

Directions: You are looking for the railway station in an unfamiliar town and ask for directions. One person you ask hesitantly answers "I think it might be down the street?". Another person you ask confidently answers "straight ahead two blocks and then first right; you can't miss it".

In *Directions*, confidence strengthens the credibility of the answer. Moral grandstanding similarly makes use of confidence. The testimony of the virtue signaller is, *ceteris paribus*, weightier than the testimony of someone who appears less confident in their moral judgments. Displays of confidence add weight to testimony, whether the content is normative or descriptive.

On the value of numbers, consider:

Restaurant: You are at a restaurant with two friends. When dividing the check, you disagree. You come up with one number while your friends agree on another.

Assuming that all parties are equally skilled, we should believe the greater number of testimonies. The reason is that testimony *is evidence*. When the number of testimonies is unequal, the evidence favours one view rather than another. This idea is familiar. When people point out that ninety-seven percent of scientists believe in anthropogenic global warming, the underlying principle is that more testimony is more evidence.

This applies in the moral domain as it applies in the scientific one. To virtue signal is to publicly testify. If testifying is to offer evidence, then virtue signalling is a way of providing evidence for a view. The fact that virtue signalers are confident and high in number adds credibility to their testimony.

² Levy, N. (2021) Virtue Signaling is Virtuous *Synthese* 198 9545-62

Against Levy on Virtue Signalling

There is a general move Levy makes that we think is particularly insightful and that we agree with. In particular, Levy often points out that various belief forming processes that we typically think of as bad are instead good. In many cases, they seem bad when evaluating an individual. But they look much better when evaluating large groups. We think this general move is a fruitful one. And we agree that it applies in many contexts. We see why it is tempting to make the relevant move here in the context of virtue signalling. But we do not think it does. We think that in this one case, Levy's general insight does not apply. In particular, we do not think the credibility enhancing features he identifies are always credibility enhancing. In some cases, we have reason to reject testimony because of these features. For example, confidence might also undermine someone's credibility. A scholar who appears too confident might seem dismissive of competing views or insufficiently self-critical. Similarly, the most credible witness in a criminal trial is not always the most confident one. Confidence does not necessarily enhance credibility. And in some cases numbers do not enhance credibility. If Trump claims he will be reinstated to the presidency in the Fall and all of the Yes Men he surrounds himself with proclaim agreement, the lack of independence of Trump's Yes Men undermines any additional support the numbers would give to his claim.

Consider a real case and plausible variations. Before the 2003 invasion of Iraq, Americans debated the merits of war. Historically, both proponents and opponents seemed confident, but the former were far more numerous.³ We doubt that confidence and numbers in this case add to the credibility of testimony. Consider, for example, lack of independence due to conformism and self-interest. If people are echoing the judgments of others, if they are simply conforming, they are not making independent judgments. The lesson is that for the number of testimonies or judgments to matter, other conditions must be met. Namely, judgments must be arrived at independently. And this is a point that Levy seems to agree with.⁴

For Levy's defense of virtue signalling to be complete, he must show us that our cynicism is misplaced. This would require showing that the beliefs expressed in grandstanding are properly formed and not expressed merely or primarily to enhance or protect one's status and therefore undermining independence.

The Condorcet Jury Theorem

There is another way that numbers might vindicate grandstanding as a guide to truth. Suppose a group is deliberating about a proposition. Suppose each deliberator has a better than .5 chance of being right about a given proposition. Suppose the deliberators arrive at their judgments independently of one another. Then,

Condorcet Jury Theorem (CJT): As the size of the group of deliberators increases, the probability that the majority is right about the proposition approaches 1.

This suggests an argument:

The Argument from CJT

1. CJT is true.
2. If CJT is true, then a consensus among grandstanders that P is significant evidence that P.

³ Nearly every Republican Representative and Senator voted for the war, with a plurality of Democratic Representatives and most Democratic Senators. These can be consulted at <https://clerk.house.gov/evs/2002/roll455.xml> and https://www.senate.gov/legislative/LIS/roll_call_lists/roll_call_vote_cfm.cfm?congress=107&session=2&vote=00237.

⁴ Levy, pp.6,8

3. So, a consensus among grandstanders that P is significant evidence that P.

We take it that (1) is well-established and needs no defense. We can motivate (2) as follows: it does not matter whether the deliberators are mostly or entirely virtue signallers. If each deliberator has a probability of being right that is better than .5 *and* if each judgment is sufficiently independent, then as their number increases, the probability that the majority holds the correct view approaches 1.

We stress how modest a requirement this at first seems. The Condorcet Jury Theorem only requires that the deliberators be barely better than chance and that each judgment is independently formed. For example, imagine that people are generally cynical about people who engage in volunteer work abroad, sometimes referred to as voluntourism. Yet, our cynicism about motives does not entail that we must consider that the likelihood of them making a correct moral judgment is less than .5. Virtue signalling might be both dishonourably motivated and epistemically useful.

The question therefore is whether or not virtue signalling, or its usual manifestations, meets the above criteria and can be defended as an application of the Condorcet Jury Theorem.

Empirically Informed Skepticism About Independence

We are skeptical of this argument. We think premise (2) is false. Or, more cautiously, we think that the preponderance of evidence tells against premise (2). In particular, we doubt that the testimony of virtue signallers is sufficiently independent for the Condorcet Jury Theorem to be relevant. Our skepticism rests on empirical evidence that suggests, first, that preference falsification is very common among humans; that conformism is stronger amongst stereotypical virtue signalers than typical humans; and that displays of confidence are often diversionary tactics; and that in general people are very conformist. So although we agree that testimony is evidence and virtue signalling is a form of testimony, we think that evidence has a defeater.

First, there is considerable evidence that people will lie about their beliefs to either promote or protect their reputations. *Socially desirable reporting* (SDR) is the form of bias that describes the tendency of people to answer survey questions with answers that are perceived as socially desirable. For instance, if frequent masturbation is perceived as socially undesirable, people will underreport the frequency at which they masturbate when asked.⁵ Preference falsification (PF) occurs when people report a false preference to avoid reputational costs.⁶ For instance, if people publicly claim to support a speech code despite privately abhorring it, they are falsifying their preferences. While SDR essentially concerns responding to surveys or similar exercises, PF does not. Additionally, SDR does not need to concern a preference while PF is specifically about one's preferences.

In other words, the general tendency to lie to protect or promote one's reputation is well-established.⁷ Once we are aware of this fact, we should be more prudent in treating the number of testimonies as significant evidence. We add two further points to expand on the threat to independent judgments. First, this tendency is not limited to authoritarian or repressive societies. Although preference falsification was common in the Soviet Union and is surely very prevalent today in countries like China or Iran, this does not mean that it does not occur elsewhere or that it is rare outside of such societies. As the foremost theorist of preference

⁵ For key readings see Allen Edwards, *The social desirability variable in personality assessment and research*, (The Dryden Press, 1957), D.L. Paulhus, "Measurement and control of response biases", in *Measures of personality and social psychological attitudes*, (Academic Press, 1991)

⁶ Timur Kuran, *Private Truths, Public Lies*, (Harvard University Press, 1997)

⁷ It is worth noting that neither SDR nor PF are focused exclusively on improving or promoting one's reputation; they also serve to protect one's existing reputation. Thus we should understand a recognition desire as a disjunctive: either one improves or one protects one's reputation.

falsification states, the phenomenon is widespread in democratic societies too.⁸ We might ask ourselves if the prevalence of it has not increased with the advent of social media that makes it possible to tarnish someone's reputation like never before. Or to put the point otherwise, if PF is about avoiding reputational costs and social media enables new threats or greater threats to one's reputation, then it is reasonable to assume that PF will be prevalent when reputational threats are enhanced. Social media and online character assassination aside, we can all observe the phenomenon at play: people commonly falsify their preferences to protect their status or to enhance it. Think of every job candidate who fibs during an interview, every person who makes up a preference to create a false connection, every 'yes-man' who unflinchingly approves of his superior's decision, every time a family member denies preferring a political candidate to avoid losing face. Examples abound.

In addition to the findings about SDR and PF, we also have studies that have specifically targeted people's willingness to lie or mask their judgments concerning ethically and politically sensitive topics.⁹ In a widely cited paper, Gilens, Sniderman and Kuklinski demonstrated that many Americans falsify their preferences and mask their anger over affirmative action. By using obtrusive and unobtrusive questions, interviews can tease out what respondents do not want to reveal. The basic lesson is that responses to obtrusive questions and unobtrusive questions differ which is indicative of respondents masking their beliefs in the face of the former and revealing them in the face of latter. Unacknowledged opposition to affirmative action - for all groups - was as high as **30.4**.¹⁰

More recent work on immigration has revealed a nearly identical pattern. When respondents are directly asked if they support restricting immigration, many falsify their preferences. Yet, when unobtrusive measures are used, we discover that **31%** of Americans hide their desire to restrict immigration.¹¹

For different research teams, working on different sensitive topics, the result is essentially the same: people will mask their true emotions or judgments on sensitive topics. Still, one might respond that these numbers, though perhaps surprisingly high, do not show that the majority of people are consciously falsifying their testimony. Are not upwards of 70% of respondents providing sincere testimony?

We do not think so. It is important to examine the details of the dissimulation, such as the distribution of this conformist behaviour. Let us return to Gilens' study. Overall, 30.4 percent of Americans masked their opposition to affirmative action, but among liberals the unacknowledged anger was **41.4%** while among conservatives it was **13.9%**.¹² In other words, among liberals a very sizable proportion was willing to dissimulate their true views. In Janus's study on immigration, we find very similar numbers. In response to obtrusive questions, only **26%** of liberals admit to wanting to restrict immigration, while **71%** admit to as much when the question is posed unobtrusively.¹³ That means that well over **40%** of liberals are presenting a false judgment or belief in order to conform while only **14%** of conservatives falsify their views. Similarly, when respondents are sorted according to their affect - how they feel towards Hispanic or Asian immigrants -

⁸ Kuran, op. cit., pp.ix-xi

⁹ Admittedly, the evidence primarily stems from studies conducted on Americans. However, we do not consider this to necessarily be an issue. It appears that the unstated focus of the conversation is behaviour that is particularly prominent in the United States or influenced by it. Moreover, the general issues of SDR and PF have been observed elsewhere; indeed, Kuran's book cites examples from all over the world. I see little reason to accept the general claim that people lie to protect or promote their reputations and yet claim that this behaviour is somehow uniquely American if and when it concerns certain moral judgments.

¹⁰ Martin Gilens et al., "Affirmative action and the politics of realignment", *British Journal of Political Science*, 1998, Vol. 28, No. 1, p.171

¹¹ Alexander Janus, "The Influence of Social Desirability Pressures on Expressed Immigration Attitudes", *Social Science Quarterly*, Vol. 91, No. 4, pp.928-946

¹² Gilens et al., p.173

¹³ Janus, op. cit., p. 939

we find a similar pattern. Those who claim to feel ‘warm’ respond very differently to obtrusive and unobtrusive questions: **30%** admit to wanting to restrict immigration in response to the former and **69%** admit as much in response to the latter.¹⁴

While about one third of the general American population is caught outright masking their views on sensitive topics like affirmative action or immigration, this is not the whole story. When we consider the breakdown by ideology, we discover two important facts. One, there is a strong asymmetry between liberals and conservatives, the latter falsifying their beliefs far less. Two, among liberals, false or conformist testimony is about two fifths or higher (40-44%), which although it is not a majority is strikingly close.

These findings are troubling. They reveal that many testimonies are driven by a desire to appear a certain way rather than sincere reports about emotions or judgments. These are instances in which deliberators provide false testimony to secure their reputation or self-image. If the Condorcet Jury Theorem relies on deliberators producing independent judgments, the above evidence of conformism represents a serious problem for the view.

So far, we have merely shown that a large minority of stereotypical virtue signallers lie. However, conformism may be generated in more than one way. In addition to conformism in the form of lying that is generated by the desire to preserve one’s reputation, there is also conformism in the form of an honest judgment that one holds simply because one sees the majority or some authority holding it.

A very large literature has followed in the wake of Solomon Asch’s work on conformity.¹⁵ We learned that people are very susceptible to the pressure to conform. When asked a simple perceptual question about the length of lines, conformity produced impressive results. When the experiment subjects were not influenced by a group’s view, they answered incorrectly only **1%** of the time. However, when faced with several confederates agreeing on the wrong answer, test subjects erred **37%** of the time. In other words, our ability to judge and think independently is quite vulnerable to conformity.

A meta-study of the literature found that anonymity did not diminish the drive to conform: whether or not a test subject’s views would be reported to the group had no significant effect on conformity.¹⁶ This is worth meditating. If the test subjects were lying, we ought to find greater conformity when the results are reported and lesser conformity when the results are not shared. Instead, conformity is essentially invariant. Merely knowing what other people think, rather than being subject to their adverse judgments, drives conformity. This is good evidence that these instances of conformism are not falsifications. Subjects in the relevant experiments simply take the testimony of others about the length of lines as evidence and think their own, as it turns out correct perception, must be mistaken. On our reading of the experiments, such subjects trust the testimony of the group more than they trust the testimony of their own eyes. And if the relevant subjects form their judgments on the basis of the testimony of others, then their own testimony is not independent and not to be double counted as evidence about the length of the lines.

Thus, conformity is not only a rational strategy to safeguard or improve one’s reputation. It is also generated by simply hearing the testimony of others. We hear the majority or the experts testify, and we believe them. Given everything we know about our evolutionary inheritance and our tribal nature, these findings should not be that surprising. We are groupish and conformist creatures. And in many ways that is a good thing. It is good that people trust testimony. And it is good that people trust experts. But what that means is that such judgments based on such trust are not independent. And so the CJT does not apply. If my

¹⁴ Ibid., p. 942

¹⁵ Solomon E. Asch, “Studies of independence and conformity: I. a minority of one against a unanimous majority”, *Psychological Monographs: General and Applied*, 1956, Vol. 70 No. 9, pp.1-70

¹⁶ R. Bond and P.B. Smith, “Culture and conformity: A meta-analysis of studies using Asch’s (1952b, 1956) line judgment task”, *Psychological Bulletin*, 1996, Vol. 119 No. 1, p.124

doctor tells me I'm sick, I'll believe it just on the basis of the doctor's expertise. But then we should not double count my judgment about whether I am sick as additional evidence beyond the doctor's testimony.

This sort of conformity does not occur merely in the context of perceptual judgments. There is a growing literature on moral conformity.¹⁷ Summarising past studies, Chituc and Sinnott-Armstrong conclude that conformism does affect moral judgments. Noteworthy, they cite a 2012 study by Aramovich, Lytle, and Stitka in which participants are asked the same question three times: privately, in a group, and privately again. 80% of the participants modified their views during the second phase where they interacted with confederates that disagreed with them. Tellingly, the participants did not return to their original view when given the chance in the final private interview. Thus, it is very unlikely that their responses in the group setting were merely strategic - they were given the chance to voice their views without disapproval but expressed the same view as in the group setting.¹⁸ Whether it was the result of an unconscious process or the conscious decision to trust the testimony of the confederates or something else, the 80% who modified their testimonies did not display independence.

Chituc and Sinnott-Armstrong attempt to explain this conformism. While they do not deny that some conformity is driven by strategic reasoning - to defend or promote one's standing - they claim that it should primarily be understood as an epistemic move. As we hypothesize regarding perceptual judgments, Chituc and Sinnott-Armstrong explain these results in terms of subjects' dispositions to believe that majorities are onto something.¹⁹ Conformity in this case is generated because we respond to numbers as evidence rather than respond to them as a threat in the face of which we must hide.

However, like us, Chituc and Sinnott-Armstrong worry about the epistemic credentials of this conformism about moral topics. In as many words, they cite the Condorcet Jury Theorem's requirement for independence among deliberators.²⁰ The problem, of course, is that conformism is a threat to independence: you cannot both be an independent deliberator *and* modify your view for no other reason than because you learn that large groups or experts hold that view.

Let us recapitulate. We are arguing against premise (2), namely that deliberators are sufficiently independent for the Condorcet Jury Theorem to be relevant. Our skeptical challenge rests on empirical evidence of our conformism. First, we conform strategically. We falsify our reports in order to avoid reputational costs. Second, we conform because we trust the judgments of the majority and of experts. We are conformists and our moral-cum-political judgments do not escape this tendency.

Once we consider the threat posed to independence both by *both forms* of conformism due to reputation preservation, on the one hand, and trust in the majority and authorities, on the other, we have strong reasons to doubt that premise (2) is true.

Stereotypical Virtue Signallers Are a Minority

As an aside, we think it is worth noting something else. The appeal to CJT does not help the cause of stereotypical virtue signallers. If the number of testimonies or the confidence with which they are expressed enhances their credibility, then it appears that in practice, numbers and confidence are often pitted against each other.

Take the "Defund the police" campaign as an example. While there is disagreement over what exactly it means, it is a prominent movement and people openly display their support or judgments. Yet, this

¹⁷ For a summary, see Vladimir Chituc and Walter Sinnott-Armstrong, "Moral conformity and its philosophical lessons", *Philosophical Psychology*, 2020, Vol. 33 No.2, pp. 262-282

¹⁸ *Ibid.*, p.271

¹⁹ *Ibid.*, p.273

²⁰ *Ibid.*, pp.274-275

movement has consistently been opposed by most Americans. While in June of 2020, support seems to have hit a high-water mark of **31%** national support - against **58%** opposed - this still means that more opposed than supported the proposition.²¹ Moreover, a March 2021 Ipsos poll found that support for “Defund the police” had fallen to **18%**.²²

If numbers matter, then “Defund the police” is a mistaken belief or set of beliefs. At the very least, the evidence against the truth of their views has only increased because the number of those who oppose it has a) always been greater than those who support it and b) that those who support it has dropped dramatically. If the supporters of this movement are virtue signalling, it is hard to see how their minority status and dwindling numbers shows that they are right.

If supporting the “Defund the police” campaign is a paradigmatic case of virtue signalling, then appeal to numbers would recommend that we consider the virtue signallers to be mistaken as they are fewer and dwindling. Moreover, if we recall that liberals were more likely to consciously conform and falsify their preferences, then we have further grounds to believe that some of the support for that movement - which is stronger among liberals than conservatives - was strategic rather than good faith reporting of judgments.

On many subjects, those we think of as paradigmatic virtue signallers appear to be a (small) minority. If we merely rely on numbers, then it is the credibility of views stereotypical virtue signallers reject that is enhanced and not the testimony of virtue signallers.

Let us be clear: We are not claiming that stereotypical virtue signallers are mistaken in their judgments. We are claiming instead that stereotypical virtue signallers are simply a loud minority. What happens to be the consensus view reported on social media is often not the actual consensus. And so if the numbers are taken to be especially relevant, the testimony of the quiet majority should be included. And the total number of testimonies, both loud and quiet together, often stacks against the positions of stereotypical virtue signallers.

Westra On the Value of Virtue Signalling

One important recent defense of virtue signalling is due to Evan Westra²³. Westra argues that virtue signalling is valuable because it is a tool to communicate social norms. As he (forthcoming, p.) puts it:

So understood, the potential utility of virtue signaling as a tool for positive norm change becomes clear. If a group of influential virtue signalers can be convinced that publicly committing to some new normative standard will increase their moral reputations, then they stand to play a valuable role in spreading that new norm throughout the broader population. This amounts to a three-step process: first, a group of sincere advocates for change seed a new, positive normative standard into the public discourse; second, virtue signalers eager to appear ‘on the side of the angels’ broadcast this new standard to a broader audience through a mix of positive avowals and public shaming; third, a much larger population treats the behavior of these virtue signalers as evidence that they should change their social expectations, and become motivated to conform to the new norm.

Suppose Westra’s model is correct. Then it supports our skepticism of (2). In his model, you have some sincere advocates of a view. Then you have a larger number of people conforming and broadcasting that view in order to appear virtuous. Then you have the much larger population accepting that view on the basis of the

²¹ A discussion of public support at the time can be found on the statistics focused website Five Thirty Eight, <https://fivethirtyeight.com/features/americans-like-the-ideas-behind-defunding-the-police-more-than-the-slogan-itself/>

²² All of the polling details can be consulted at https://www.ipsos.com/sites/default/files/ct/news/documents/2021-03/usat-ipsos_racial_injustice_topline_030421.pdf

²³ Westra, E. (2021) Virtue Signalling and Moral Progress *Philosophy and Public Affairs* 49 156-78

testimony. Here the collection of judgments is not independent. The large majority form their judgment on the basis of the testimony of the virtue signallers. The virtue signallers form their judgment on the basis of status seeking. *At most*, the minority sincere advocates are independent. If Westra's model is right, then virtue signallers and those who trust their testimony lack independence. And therefore our skepticism about premise (2) of the Argument from CJT is still warranted. The dependence of the virtue signaller's judgments means the numbers do not enhance the value of their testimony.

Levy on Independence and Discrimination

Levy offers some reasons for thinking that the judgments of virtue signallers are independent. The thought is that there is evidence that in general humans are typically discriminating in their judgments. As Levy puts it:

Numbers do not make a difference in either kind of way if additional agents are non-discriminating reflectors of a single individual (Goldman 2001). A non-discriminating reflector holds whatever opinion their guru holds, regardless of its plausibility, and therefore their agreement adds no independent epistemic weight to the initial opinion. In the actual world, agents are never or almost never non-discriminating reflectors of a guru, for at least two reasons. First, even if some agents are non-discriminating with regard to token opinions, the fact that they regard someone as a guru is good evidence that they take them to be reliable in general. Second, there is extensive evidence that agents are rarely genuinely non-discriminating. Even young children filter claims for plausibility, and will reject testimony from a familiar person, even a parent, in favour of more plausible testimony from an unfamiliar informant (see Harris 2012). The degree of independence of individual informants from one another varies from case to case, but we can be confident that each filters testimony for plausibility to some degree.

We agree but we doubt that this vindicates virtue signalling. We worry that conformism poses a challenge to independence even if we are discriminating. Above, we provided evidence from multiple studies to show that we conform. We conform strategically: we are seeking to preserve or promote our reputations. We also conform non-strategically when we respond to numbers and adjust our views in good faith by deferring to the greater numbers and to expertise.

Moreover, to the point that even children are discriminating: There is good evidence that children are conformists.²⁴ But given that virtue signalling is really about adults, we can turn to those whom we might expect to be the most independently minded - the most educated. Yet, even the educated or politically sophisticated display behaviour that is uncritically partisan and conformist. The politically sophisticated are not models of independent thought: they are more likely to uncritically accept or dismiss information depending on whether it supports or opposes their view.²⁵ In fact, political sophisticates are *more likely* to vote out of loyalty or partisanship than on the basis of content, i.e. the candidates beliefs, performance in office, etc.²⁶ We agree with Levy that humans are not non-discriminating reflectors. And we recognize that that is worth noting. But that is compatible with our doubt that virtue signallers are not sufficiently independent for

²⁴ E. B. Kim et al., "Does children's moral compass waver under social pressure? Using the conformity paradigm to test preschooler's moral and social-conventional judgments", *Journal of Experimental Child Psychology* (2016), Vol. 150, pp.241-251

²⁵ Charles s. Tabert and Milton Lodge, "Motivated Skepticism in the Evaluation of Political Beliefs", *American Journal of Political Science*, 2006, Vol. 50, pp.755-769

²⁶ John Zaller, "Floating Voters in the U.S. Presidential Elections, 1948-2000", in *Studies in Public Opinion: Attitudes, Nonattitudes, Measurement Error, and Change*, edited by Willem E. Saris and Paul M. Sniderman, Princeton University Press, pp. 166-212

the CJT to work and for us to be skeptical about the support the numbers give to the propositions virtue signallers endorse.

Westra on Epistemic Vigilance

Westra suggests another way in which the numbers might add credibility to testimony. As he puts it:

To protect ourselves from deception and misinformation, human beings have developed psychological mechanisms for epistemic vigilance, which enable us to monitor the coherence of a person's testimony and reasoning, and to keep track of their overall epistemic reliability. These mechanisms kick into action not when we reason in a solitary fashion—in that regard, people are relatively lazy and tend to conserve cognitive resources—but when critically evaluating the arguments of others. This asymmetry in the reasoning process amounts to a form of interactive quality control that enables groups of people to reject bad arguments and converge upon sound ones. In the process, it forces individuals to generate better arguments for their conclusions than they would have on their own. Thus, the claim that virtue signaling is unconstrained from the truth ignores the fact that it occurs within a discursive ecosystem filled with epistemically vigilant agents, where poor arguments and unfounded claims are subject [to] scrutiny.

We recognize that this sort of epistemic vigilance is operative in some contexts. But our developmental story is not about achieving objective and timeless truth. It is about out-competing other groups. We are tribal to the core.²⁷ Our ability and willingness to critically evaluate an argument depends largely on our pre-existing beliefs - we are all better at finding fault in the views of others than our own. We display favourable biases towards the in-group and disfavorable biases towards the outgroup, whether one is so-called liberal or conservative. Evolution has not produced neutral truth seekers, but loyal members of competitive coalitions or teams who are particularly apt at finding flaws in rivals and far less predisposed to become internal gadflies. The point is not that we are incapable of self-criticism or independent thought. Rather, a sober reading of the evidence shows that we are far less self-critical or independent than we would like to believe.

This is one reason we think it is so valuable to have people coming from different angles discuss sensitive topics like virtue signalling. We think that we, along with Tossi and Warmke, are especially well positioned to point out flaws in Levy and Westra style arguments. And we think that Levy and Westra are especially well positioned to point out flaws in our arguments. And we think that this idea fits well with Levy's broader points in his other work about the social nature of epistemology which we accept. And it fits with Levy's broader view even if, in this one respect, we disagree with him.

The Problem of Affective Polarization

In the remainder of our paper, we argue that even if virtue signalling were epistemically useful it would remain too costly to use. We begin by laying out Levy's response to the issue of polarization. Levy points out that polarization is not necessarily bad. He writes (forthcoming, p. 6-7) this:

[I]t is far from obvious either that group polarization is irrational.... Let's pause, first, to assess whether group polarization deserves condemnation (however it arises). There seems no a priori reason to think that the truth is more likely to lie in the middle of a group of deliberators, prior to their sharing their opinions with one another (and thereby providing first and higher-order evidence

²⁷ For a relatively recent summary, see Cory J. Clark et al., "Tribalism is Human Nature", *Current Directions in Psychological Science*, 2019, pp.1-6

in favor of their views), than at the extremes.... Extreme opinions about race and gender were more accurate than more moderate opinions in the antebellum United States, for instance.

Westra also addresses the worry about polarization. He says (p. 175):

As far as virtue signaling is concerned, the threat of polarization turns out to be a little less disturbing than it first appeared. Polarization does not seem to seriously undermine epistemic vigilance, at least for more reflective individuals. This means that when a virtue signaler makes an implausible but politically convenient claim about the outgroup, whether or not they are believed will depend upon whether their audience is disposed toward critical thinking, not just their partisan allegiance. Virtue signalers spreading such falsehoods cannot count on their most cognitively sophisticated audience members to rationalize their dubious claims. Polarization does not make bad reasoners of us all—it is just one more context where bad reasoners might reveal themselves.

We think that what Levy and Westra say here is plausible. But we do not think it addresses what we are most worried about. The problem with polarization is not *epistemic*, but *political and social*. Those who worry about political polarization are not primarily worried that people have a misguided view about how many ‘middle-ground’ propositions are true: they are worried about the breakdown of cooperation, compromise, or trust within a population. The problem is about cooperation and trust within human societies. The costs identified by Torsi and Warmke – cynicism, outrage exhaustion, and polarization – are primarily undesirable because they weaken our ability to engage and cooperate. So in short, what we are worried about is not strictly polarization simpliciter. We are worried about *affective polarization* - the phenomenon whereby partisans on opposing sides come to view one another more negatively.

No human society is composed of an ideologically homogeneous population. Or to put it otherwise, human societies require people who disagree to cooperate and trust each other. They must also allow for disagreement and productive discussion of competing views. Yet, virtue signalling undermines all of this.

On the one hand, we saw that virtue signalling seems to prevail in contexts where conformism is rewarded. Yet, a serious discussion about difficult moral questions is not helped by an environment that punishes dissent or rewards hyperbole for the sake of homogeneity. We strongly encourage the reader to consider how a need to constantly signal that one was a good Marxist and the corollary fear of being denounced affected the quality of moral and scientific debate. We would suggest that growing worry, occasionally expressed by academics in the popular press, about the increasingly difficult task of debating certain issues is in part a product of virtue signalling and the context it creates.²⁸

On the other hand, phenomena like affective polarization generate resentment and anger between fellow citizens. Rather than treating fellow citizens as partners in search of good ideas and policies for the nation, one starts to view others as obstacles or enemies. Think of the less than stellar quality of the conversation of gun control/gun rights. As each side becomes more radical, they become less and less willing to engage with the other. The reason they become distrustful and contemptuous is that as each party becomes polarized, they all consider that the gap between them widens.

It will not do to reply that one side is wrong in this debate. Let us assume that the defenders of gun rights are wrong. Still, affective polarization and distrust do not disappear once a law is passed, or a policy adopted. It is because there are and always will be other issues that require cooperation that it is dangerous to sow distrust and resentment. If you treat people as moral lepers or unworthy of citizenship on issue X, you

²⁸ A popular piece that received a fair amount of attention can be consulted at <https://www.vox.com/2015/6/3/8706323/college-professor-afraid>.

can hardly expect their good faith cooperation on issues Y and Z. In other words, at best virtue signalling and its intolerant climate can generate short-term gains at the expense of long-term costs.

Finally, one might think that polarization can serve ‘the good’ or progress. Discussing a plausible case of gun control/gun rights after a school shooting Levy, for instance, finds value in polarization. Those who favour gun control and develop more radical views are correct. He reminds us that we ought not reject this normative view merely because it is extreme.

We disagree with Levy about this. First, given polarization, those who were initially wrong now hold an even worse view. If polarization leads to both radical correct and radical false views, then polarization yields mixed results. Worse, it tells us nothing about the relative political power or motivation of those concerned. If those who hold the wrong views are better funded, or organised, or motivated then polarization has made them more dangerous. This is a high price to pay for some epistemic gains.

Second, we are not really concerned about polarization *per se*. Again, what worries us is affective polarization. The general phenomenon of virtue signalling and the atmosphere of cynicism and resentment seems to produce awfully bad consequences. If the likely consequence of more virtue signalling is serious blowback, then we ought to wonder if the gains of such polarization are worth its costs. For instance, some research suggests that people’s aversion to political correctness increased their willingness to support Donald Trump.²⁹ This does not mean that aversion to something like virtue signaling - assuming that what is often called political correctness overlaps with it - caused Donald Trump’s election. But it certainly seems to be one factor. In other words, there is reason to believe that less virtue signalling would have made American voters less receptive to Donald Trump. Given the closeness of the 2016 election, it is not wildly improbable that with less political correctness Donald Trump would never have become president.³⁰

Objections and Replies

First Objection: Suppose the percentage of progressive white Democrats who falsify their preferences is, as we have suggested, over 40%. The reliability and independence of the majority of progressive white Democrats as a group remains intact.

Reply: Here we need to return to the idea that conformism can be generated in more than one way. Those who falsify their views are lacking independence. But as we saw, there is a considerable body of evidence to show that our vulnerability to social pressure is not only strategic. Even if 60% of white Democrats honestly report their views, that would still not secure independence. For example, suppose that a remaining 10% of the honest progressive white Democrats hold their views because it seems like the cool thing to do based on what they observe on social media. Or they are deferring to what they think the majority or the people they take to be experts think. In that case that puts the number of progressive white Democrats that arrive at their judgment independently at most at .5. And that is not enough for the CJT to apply.

Put another way, suppose we are right that 40% of progressive white Democrats lack independence because they lie in order to preserve their reputations. We would be surprised to learn that all of the remaining 60% that do not lie are independent. It is not far-fetched to suppose that progressive white Democrats also generate reports of beliefs in other conformist ways. Others may do it to be cool, because they value the connection of having shared beliefs with others, because they simply trust the testimony of others in their tribe, etc. Whatever the reason, and there are many, it is enough these judgments are affected by social pressures. Any that is sufficient to undermine independence. Compare: Imagine we are at a restaurant. You

²⁹ Lucian Gideon Conway III et al., “Donald Trump as a Cultural Revolt Against Perceived Communication Restriction: Priming Political Correctness Norms Causes More Trump Support”, *Journal of Social and Political Psychology* 2017, Vol. 5, No. 1, pp. 244-259

³⁰ For well-known version of this argument, see Mark Lilla, “The End of Identity Liberalism”, *The New York Times*, November 18th 2016

calculate the tip and tell me it is \$42. I believe you. Then someone asks us what the tip is and we both say “\$42”. In that case my testimony is not independent of yours. It should not be double counted.

Second, a point of clarification: we think all of this is beside the point. Distinguish between considerations about reliability and considerations about independence. CJT doesn’t merely presume that greater than .5 of the deliberators in the group are independent. It presumes that *all of them* are independent. The bit about .5 is that the individual judgments of all the deliberators has to be true with a probability greater than .5. So even if we were to grant the objector’s point, that 60% are independent, white progressive Democrats still do not meet the minimal conditions about independence for the CJT to kick in.

Second Objection: This suggests a different way to run the CJT. Let’s take as our group the 60% of progressive white Democrats that are honest. They are independent. So as long as they are each individually better than .5 at getting to the truth, we can run the CJT.

Reply: We do not find this objection to be persuasive. First, we think the higher prevalence of lying among progressive white Democrats is a canary in the coal mine. We suspect that this group will display other tendencies to conform and display a lack of dependence. But we can make our point in a more cautious way. We recognize that the empirical evidence we cite does not, by itself, settle whether the relevant subjects are sufficiently independent for the purposes of CJT. But we think that, at best, this just shows that we don’t know to what degree members of this group lack independence. Of course, we could be wrong about this. Maybe there is other empirical evidence defenders of the Argument from CJT might offer to support the claim that the relevant group is sufficiently independent. But the proponent of the Argument from CJT owes a defense of this claim. They may not simply assume it. If we are told that the relevant testimony is independent, and then we are given evidence that some of it is very dependent, it is on the proponent of the Argument From CJT to defend premise (2) by showing that the relevant subjects are sufficiently independent. We have no reason to think the testimony of virtue signallers is sufficiently independent for CJT to be relevant. We are happy to be proven wrong. What this suggests is it is not so easy to support the relevant premise. What needs to be done is further empirical work. If further studies are done that show that virtue signallers are independent in the relevant way, we will gladly change our minds. Until then, we think the Argument from CJT lacks sufficient support.

Second, if we really could identify the independent progressive white Democrats and run the CJT with respect to them, we would still need to factor in the people with other political views that form their judgments independently and have a probability greater than .5 of getting things right. We see no reason to think this would come out in favor of progressive views. As we said earlier in the paper, what matters is the majority judgment of the total number of independent deliberators and not the majority judgment of the subset of independent deliberators that are progressive white Democrats.

Third Objection: We are assuming that we are entitled to hold that virtue signallers lack independence. But we are not entitled to that assumption. It goes beyond the empirical data that we cite.

Reply: Again, we agree that this assumption goes beyond the empirical data that we cite. But we think this objection depends on a misunderstanding of what we are trying to do. Our project does not depend on this assumption. Here is the dialectic as we understand it: We are told that virtue signalling is significant evidence. In particular, it is testimony that is high in confidence and numbers. And that is a reason to think it is good evidence. We argue that that is not sufficient to treat it as good evidence. We then consider the possibility that it is good evidence because of the Argument from CJT. We reply that that that argument isn’t any good because there is not sufficient reason to think premise (2) is true. We think it isn’t supported well. We consider a possibility in which premise (2) is false. For one thing, a significant number of virtue signallers lie. For another, for all we know there are a bunch of other ways in which virtue signallers lack independence. In the experiments discussed by Chituc and Sinnott-Armstrong, for example, 80% of the subjects modified

their moral judgments based on the numbers. Indeed, we think it would be surprising if many virtue signallers conform for strategic reasons, but that all the others are radically independent in the way needed for CJT to be relevant. If they are conformist in such a way that they lie more than typical humans in such contexts, then it is natural to think that they are conformist in other ways as well.

Of course, it may turn out that there is empirical literature demonstrating that although virtue signallers lie in order to conform, they avoid conforming and are especially independent in other ways. And we would welcome such empirical evidence. But we have not seen it. In its absence, a defense of (2) cannot simply presume that although virtue signallers lie more than the general population, they are otherwise very independent. And in that case premise (2) lacks support. If we show that, the skeptical case stands.

Think of it this way: In Levy's defense of virtue signalling, he appeals to studies about stotting antelopes to make his case about humans. If Levy can extrapolate from studies about stotting antelopes to make his case, we think we can extrapolate from studies about virtue signallers and humans in general being conformists to argue that it is plausible that virtue signallers are too conformist for the CJT to kick in.

Compare: Imagine you are evaluating various formulations of the Common Consent Argument for the Existence of God. The core idea is as follows: most people believe that God exists. If most people believe that God exists, then you should believe that God exists. So, you should believe that God exists. You are told that the testimonial evidence for the existence of God is strong because most people testify to the existence of God, those people are independent, and they have a reliability greater than .5. Let's imagine that you then present empirical evidence that a significant minority of people who loudly testify to the existence of God lie about it in order to protect their reputations. And the number of liars professing belief in God for this reason is significantly greater than the number of liars professing atheism. Let's imagine that the totality of people who testify are actually in favor of atheism but they are quieter about it. Let's imagine that you provide empirical evidence that lots of people sincerely believe in God just because they trust the numbers and experts. Now imagine you hear the following reply to all this: "You haven't proven that the judgments of people who claim to believe in God lack independence. You've only shown that people who report belief in God lie more than people who don't. So the CJT based Common Consent Argument is sound." In this case, it seems like you win the debate. You've given enough grounds in this case to doubt the common consent argument. It is on the proponent of the argument to provide support for their claim.

Fourth Objection: We say that 30.4% would deny opposition when asked directly. But they would reveal it clearly when tested indirectly. This is a case of falsely reporting views in an anonymous survey. That is very different from proclaiming lying in public for all to hear.

Reply: Agreed. What that study gives us is that people lie on anonymous surveys. Now, in light of this, here are two hypotheses:

H1: People lie on anonymous surveys, but they are honest when in public and their reputation is at stake.

H2: People lie on anonymous surveys and they also lie when in public and their reputation is at stake.

We place our bets on H2. We think that if people are willing to lie on an anonymous survey when their reputation is not at stake, then they will be even more willing to lie in non-anonymous contexts in which their reputation is at stake. Of course, we welcome empirical data to the contrary. And if given such data we will change our minds. But in the absence of such data, we think H2 is the better hypothesis.

Fifth Objection: We claim to have shown that the general tendency to lie to protect or promote one's reputation is well established. But we overreach. We start with evidence that some people sometimes lie about

certain issues to preserve their reputation. We then conclude that reputation promoting speech acts are generally unreliable. That inference is unsound. We have no clue what the ratio of true reputation preserving speech acts to false reputation preserving ones are., Thus we are not in a position to make the latter claim.

Reply: Recall, we think that the relevant issue in this context is not whether individual virtue signallers are reliable with a probability greater than .5. The issue, we think, is instead whether virtue signallers are *independent* in their judgments. We agree with the objector that the inference to unreliability is unsound. But our claim is that lying is sufficient to establish lack of independence. If I lie and say that I agree with you in order to shield my reputation from your criticism, that is sufficient to show that my reported judgment is not independent of yours. And so you shouldn't double count my agreement with you as extra evidence beyond your single judgment. If lots of stereotypical virtue signallers lie to promote their reputation, that is sufficient to show that their reported judgments are not independent. And that is sufficient to undermine the application of CJT to virtue signalling.

Sixth Objection: The term 'virtue signaling' is a slur used by the right. That is why it seems strongly connected with the left. This is why Tosi and Warmke use the term 'moral grandstanding' instead. Obviously, reputation-seeking moral speech is also a phenomenon on the right - it is just not often called 'virtue signaling'. So when we appeal to accusations of virtue signaling as armchair evidence that virtue signallers are most likely to be progressives. We are relying on a mere stereotype.

Reply: Regarding our use of 'virtue signalling' rather than 'moral grandstanding': Recall that we see our project as, in part, a reaction to and continuous with Neil Levy's important defense of virtue signalling. His paper is called 'Virtue Signalling is Virtuous'. He talks about virtue signalling instead of grandstanding. Since we think of our paper as in conversation with him, and since we think it would be valuable for our papers to be read side by side, we follow Levy by using 'virtue signalling' rather than 'moral grandstanding'. Also consider Evan Westra's paper 'Virtue Signalling and Moral Progress'. It is the other main defense of virtue signalling in the literature. And it uses 'virtue signalling' rather than 'grandstanding'. Since the main defenders of virtue signalling prefer to call it 'virtue signalling' we think it is permissible for us to do so as well. We want to use the term that defenders of virtue signalling themselves use.

Seventh Objection: Our argument in part depends on the claim that stereotypical virtue signallers lie for reputational benefit (thereby violating one of the conditions for Condorcet's jury theorem). But we use armchair inferences about who is most likely to be accused of virtue signaling as evidence for where virtue signaling is most prevalent.

Reply: Regarding whether our argument depends on the truth of the stereotype that white liberals are the main virtue signallers: We do not think it does. We note that some conservatives do lie to protect their reputations. They just lie in smaller numbers than white liberals. We also note that our point about there likely being a bunch of other ways in which liberals are conformists applies to conservatives as well. In contexts in which there are a bunch of conservative virtue signallers, we wouldn't take their convergence toward some perspective to be enough for the CJT to kick in. We doubt that the conservatives would in that case be sufficiently independent. For that there are very likely a number of other respects in which conservatives are conformist that undermines their independence.

Although our argument doesn't depend on whether the stereotype is true, we think it would be fine if it did. We see no problem with this. In particular, one of the points we are making is this: We think that if we are evaluating whether virtue signalling is significant evidence, it wouldn't hurt to look at the people who are most often accused of virtue signalling. It seems like a natural thing to do. And it is interesting enough to learn that the stereotype virtue signalling doesn't count as significant evidence. There is the further question of whether empirical work will vindicate the stereotype. And we're happy to learn that the stereotype is wrong. But it is interesting enough to point out our doubts that when it comes to the stereotypical virtue signallers,

there isn't a good reason to treat what they say as significant evidence. And in any case, as we have said, we think our argument does apply to conservative virtue signallers.

Regarding whether the stereotype is true: We think the claim that it is true is an implicit assumption in the literature. We are bringing that assumption to the surface to be scrutinized. We see ourselves as inheriting this assumption from the relevant literature. If it is mistaken, we are happy for it to be corrected.

Eighth Objection: Our argument depends on the assumption that it is possible to identify virtue signallers. But virtue signaling is constituted by underlying desires and not observed behavior. So it is difficult to identify virtue signallers.

Reply: We do not think our argument depends on our ability to identify instances of virtue signalling. At most, what our argument depends on is the claim that a lot of political discourse is polluted by virtue signalling. We think this is an assumption we inherit from Levy and Westra and Tossi and Warmke. And if virtue signalling doesn't pollute our discourse, then we think it isn't that interesting or worth talking about in the first place. So what we assume is that there is a lot of virtue signalling. Even if it is difficult to tell whether a particular instance of an outraged person on Twitter is an instance of virtue signalling, that is significant enough to cast doubt on the significance of a large collection of outraged people on Twitter. And anyway, it isn't really about what counts as virtue signalling and what doesn't. The concern is that the testimony we see on social media about moral claims is not independent.

Ninth Objection: Our argument depends on the assumption that virtue signalling is the main cause of polarization. To quote a referee "An alternative view is that political polarization is both a byproduct of structural economic, political, or social development, such as deep economic inequality and the rise of right-wing media such as talk radio and Fox News, as well as a cause of a host of other bad outcomes."

Reply: We agree that we have provided no reason to think virtue signalling is the main cause of polarization. But we think that is compatible with our point. As we understand it, the dialectic is this: Some people say virtue signalling is bad. One of its bad features is that it causes polarization. And polarization is bad. Other people agree that it causes polarization. But they deny that polarization is bad. Levy and Westra agree with Tossi and Warmke that virtue signalling causes polarization. But they think polarization isn't a problem. We argue that polarization is bad. And we agree with all these authors that virtue signalling is one thing that causes or contributes to polarization. Doing so does not commit us to the claim that virtue signalling is the main cause of polarization. We can insist that virtue signalling is bad without at the same time insisting that it is worse than everything else. We are not fans of Fox News or conservative talk radio. We are not fans of some of the kinds of inequality we find in the U.S.. We can agree that all of these things contribute to polarization. We agree that it is unlikely that virtue signalling is the one main driver of polarization. Although maybe part of what is constitutive of the badness of Fox News and conservative talk radio is the virtue signalling that they participate in and promote.

Tenth Objection: We said that in a very similar world with less political correctness, it is plausible that Donald Trump would never have been president. But studies have shown that the best predictor for Trump support is racial resentment. So what we said is mistaken.

Reply: We have no desire to deny that racial resentment might be the single best predictor of support for Trump. But we deny that this conflicts with what we claim. On the one hand, we doubt that racial resentment is independent of polarization. The more polarized you make people, the more people who have racial resentment will increase in their racial resentment. Consider the phenomena of people who voted for both Trump and Obama. In the Midwest, Donald Trump won 209 counties that twice voted for Obama.³¹ Go back to Levy's discussion of polarization. As polarization increases, people become more radical. We think it is

³¹ For a very early description of these surprising switches see <https://www.washingtonpost.com/graphics/politics/2016-election/obama-trump-counties/>

plausible that people grew in their racial resentment due to polarization. Furthermore, we believe that such occurrences are not only explained by racial resentment, but also by resentment towards phenomena like virtue signalling or political correctness. So, first, we think that racial resentment is increased by polarization. And, second, we think racial resentment need not be the only factor that brought about Trump's victory even if it is the main one. It is plausible that racial resentment, by itself, would not have been enough to get Trump into office.

Acknowledgements

For comments and discussion we are very grateful to Nils Holtug, Justin Tossi, and Brandon Warmkie. Special thanks to Neil Levy and the referees for *Synthese*.