

What's Wrong With Virtue Signaling?

James Fanciullo and Jesse Hill

Lingnan University

Forthcoming in *Synthese*

Abstract: A novel account of virtue signaling and what makes it bad has recently been offered by Justin Tosi and Brandon Warmke. Despite plausibly vindicating the folk's conception of virtue signaling as a bad thing, their account has recently been attacked by both Neil Levy and Evan Westra. According to Levy and Westra, virtue signaling actually supports the aims and progress of public moral discourse. In this paper, we rebut these recent defenses of virtue signaling. We suggest that virtue signaling only supports the aims of public moral discourse to the extent it is an instance of a more general phenomenon that we call *norm* signaling. We then argue that, if anything, virtue signaling will undermine the quality of public moral discourse by undermining the evidence we typically rely on from the testimony and norm signaling of others. Thus, we conclude, not only is virtue signaling not needed, but its epistemological effects warrant its bad reputation.

Since the advent of social media, allegations of virtue signaling have become ubiquitous, and the practice is generally met with repugnance. Calling someone a 'virtue signaler' is an insult that is supposed to be representative of a vice or serious moral failing. Recently, Tosi and Warmke (2016, 2020) have given a novel analysis of the concept of virtue signaling or as they call it moral

grandstanding (hereafter, we will use the phrase grandstanding¹ when referencing Tosi and Warmke’s work but will otherwise stick with the more common phrase virtue signaling).² Tosi and Warmke mostly agree with our ordinary usage of the term and argue that grandstanding is usually bad and should be avoided. They hold that grandstanding is a threat to free expression and productive moral discourse, disrespectful, and representative of an aretaic failing. Grandstanding is not something that the virtuous person engages in and when rampant carries a heavy social cost. However, Levy (2021) and Westra (2021) have pushed back on these claims. They hold that virtue signaling is not seriously morally wrong or representative of a character flaw but a central and necessary part of moral discourse in that it shows our commitment to certain norms and provides genuine, higher-order evidence to agents. Accordingly, they argue that while the practice can sometimes lead to bad results it does not deserve our condemnation. As Levy puts it, virtue signaling’s “virtues typically outweigh its vices” (2021, p. 9545).

Our aim is to settle this debate by focusing on whether the practice of virtue signaling is a reliable way to advance moral discourse, that is, whether virtue signaling, itself, serves a central and necessary function in moral discourse and whether the practice provides genuine, higher-order evidence to agents. To accomplish this goal, in section 1, we briefly recapitulate Tosi and Warmke’s account of grandstanding. In section 2, we first summarize Levy’s three-part defense of virtue signaling, and go on to argue that Levy apparently conflates virtue signaling with a more general phenomenon that we’ll call *norm* signaling. The upshot is that the positive characteristics that Levy attributes to virtue signaling are actually constitutive of norm signaling. In section 3, we

¹ One could, of course, grandstand in non-moral ways (for example, when, after hitting a homerun, a major league baseball player flips his bat and slowly rounds the bases). Our focus will be on moral grandstanding, particularly within the realm of public moral discourse.

² It should be noted, however, that Tosi and Warmke have reasons for preferring the term ‘grandstanding’ (Grubbs et. al 2019), and for them ‘virtue signaling’ and ‘grandstanding’ may not always be interchangeable. See Levy (2021, p. 9546, footnote 1) and Westra (2021, p. 157, footnote 3) for more on this issue.

argue that Westra similarly fails to show that virtue signaling reliably advances moral discourse. Next, in section 4, we argue that virtue signaling does not always provide genuine, higher-order evidence to agents—and, when rampant, actually undermines the evidence provided by testimony. We conclude, in section 5, by reiterating how virtue signaling is neither epistemically responsible nor does it advance moral discourse.

1. Tosi and Warmke's Account of Grandstanding

Tosi and Warmke give a paradigmatic account of grandstanding (or, for our purposes, the equivalent, that is, virtue signaling) that attempts to illuminate the core features of the phenomenon. According to Tosi and Warmke, grandstanding is characterized by two central features: a *recognition desire* and a *grandstanding expression*.

Tosi and Warmke define *recognition desire* as follows:

the grandstander desires that others think of her as being morally respectable with regard to some matter of moral concern ... she wants others to make a positive moral assessment of her or the group with which she identifies. (2016, p. 200)

In other words, the grandstander desires that others will be impressed with her putative moral qualities, for example, her moral beliefs, commitment to justice, purity, superior moral acumen, etc. (2020, p. 171). Importantly, the *scope* of whom the grandstander desires to impress can vary. For example, the grandstander might wish to be recognized as morally impressive by members of a certain group or groups (as in many cases of political grandstanding) or by the population at large. Furthermore, the grandstander might desire that others see her as meeting a minimum moral standard (for example, as someone who cares about women's rights) or as morally superior to others (for example, as someone who cares more than anyone about stopping puppy mills).

Grandstanding also typically involves a contribution on behalf of the grandstander, for example, a written expression, speech, or action.³ Tosi and Warmke call this feature the *grandstanding expression*:

when one grandstands, one contributes a grandstanding expression in order to satisfy the recognition desire. In other words, one's grandstanding expressions are attempts to get others to believe that one is morally respectable. (2016, p. 202)

A grandstanding expression may involve moral claims that are *true* or *false*. For example, to achieve higher status in his in-group of white power skinheads, a Nazi might make a false moral claim about Jewish people being vile, whereas—with the hope of being recognized as a good person within her social group—a social justice warrior might make a true moral claim about certain injustices that are faced by Aboriginal peoples. Grandstanding expressions can also be *effective* or *ineffective* both in terms of whether the expression increases the status of the grandstander and whether the expression makes a difference regarding its putative moral cause.

It is also important to note that Tosi and Warmke view the recognition desire as a necessary feature of paradigmatic instances of grandstanding: “a person grandstands only if the Recognition Desire is a significant motivator for what she says” (2020, p. 173). This means that grandstanding is not just the expression of a commitment to a certain social norm but also involves the desire that one is recognized as a morally respectable person.⁴ As Tosi and Warmke put it, “the grandstander’s motivation is largely egoistic; she is using public moral discourse to secure certain kinds of

³ Tosi and Warmke focus on grandstanding expressions that are either written or verbal (2016, p. 202). This makes sense given that they were originally concerned with grandstanding’s negative effect on public moral discourse. However, there can also be grandstanding acts such as when a celebrity makes an appearance at a children’s hospital because of her desire to be recognized as a caring person. Our focus will largely be on grandstanding language, which seems particularly problematic due to its low cost. Talk is cheap, whereas action requires time or money and, as a result, is costly to fake.

⁴ Note that this desire may be *witting* (that is, conscious to the grandstander) or *unwitting* (that is, an unconscious desire) (Tosi and Warmke 2020, p. 173).

recognition for herself ... To grandstand is to turn one's contribution to public discourse into a vanity project" (2016, pp. 198-199).

This raises interesting questions about the extent to which a person's action must be motivated by her desire for recognition for her action to constitute an instance of grandstanding.

Tosi and Warmke's view, which we will adopt, is as follows:

Our claim here is that the recognition desire plays a motivating role in paradigmatic cases of grandstanding. Although the recognition desire may not be the only desire motivating the grandstanding expression, or even the strongest motivating desire, the recognition desire does make a significant motivating contribution. That is, the grandstander says what she does in large part because she desires that others think of her as morally respectable. ... Just how significant? We think that the desire must be strong enough that if the grandstander were to discover that no one actually came to think of her as morally respectable in the relevant way, she would be disappointed. (2016, p. 202)

This leads to a threshold account of grandstanding wherein if a putative grandstander would not be disappointed if her (perhaps non-existent) recognition desire was not satisfied via the agent's putative grandstanding expression, then the agent's act is not an instance of grandstanding.

Another important distinction concerns whether the grandstander is *sincere* when she makes her grandstanding expression. We think that there are at least three possibilities regarding sincerity and grandstanding.⁵ First, is the "*whole-hearted*" grandstander who is significantly motivated by her desire for recognition but nevertheless does believe in the cause she is advocating. Such a person will be more likely to act on her beliefs. Second, is the "*half-hearted*" grandstander who is indifferent or does not have strong beliefs regarding the cause she is advocating. Her grandstanding expression amounts to mere lip service. Lastly, is the "*cold-hearted*" grandstander who does not believe in the cause she is advocating for. There is some overlap between the cold-

⁵ Tosi and Warmke make a similar but two-parted distinction in one of their responses to a PEA soup blog post, see (Coady et al. 2017).

hearted grandstander⁶ and other dubious moral monikers, for example, the liar, the bullshitter, the hypocrite, and the sophist.

Note, then, that the same coarsely defined event (for example, changing one's Facebook profile picture in support of Black Lives Matter) could be an instance of whole-hearted, half-hearted, or cold-hearted grandstanding—or not involve grandstanding at all! The whole-hearted grandstander is motivated by her desire to be viewed positively by a certain group of people, but her expression (that is, when she changes her profile picture as a show of support for African Americans) is sincere. The half-hearted grandstander is largely motivated by her recognition desire. She changes her profile picture to seek approval and will likely do nothing else to support the movement if such actions are costly. She is caught up in—what is for her—a fad. The cold-hearted grandstander is mainly motivated by her recognition desire. She changes her profile picture to maintain a good reputation and mask her actual beliefs. Of course, one could also change one's profile picture in support of Black Lives Matter without grandstanding. The difference between such a person (who is just signaling a view which she holds) and the grandstander is motivational.

We can now return to Tosi and Warmke's claim that one of the ways in which grandstanding is morally problematic is that it interferes with the main aim of public moral discourse, that is, "to improve people's moral beliefs, or to spur moral improvement in the world" (2016, p. 210). They argue (2016, pp. 203-208) that grandstanding does this, in part, by manifesting itself in five problematic ways:

- 1) *Piling on* occurs when people chime in to reiterate what others have already said. People who pile on contribute nothing new to a discussion and are trying to signal that they, themselves, are on the "right side".
- 2) *Ramping up* occurs when people make "increasingly strong claims about the matter under discussion" (p. 205). They do so not just to signal that they are on

⁶ This holds—perhaps to a lesser extent—for the half-hearted grandstander as well.

- the right side but to show that they are the ones in their in-group who are the most respectable. People who are guilty of such behavior turn moral discussion into an “arms race”.
- 3) *Trumping up* occurs when people “insist on the existence of a moral problem where there is none” (p. 206). By doing so, they attempt to show others that they are particularly respectable and have a keener moral sense. However, they are too eager to find fault as the “problem” that they are identifying is not morally objectionable.
 - 4) *Excessive outrage* occurs when people use exaggerated emotional displays to advance the cause they are advocating. Such emotional displays are, again, used to signal how much more the grandstander cares and how keen her moral senses are. Of course, emotional displays in response to immoral conduct are sometimes apt; however, the grandstander’s outrage is disproportionate to the seriousness of the offense.
 - 5) *Claims of self-evidence* occur when someone dismisses others’ views and will not engage in debate because their own moral views are self-evident and not in need of defense. As Tosi and Warmke put it, “What is not obvious to others is painfully obvious to the grandstander ... any suggestion of moral complexity or expression of doubt, uncertainty, or disagreement is often declaimed by the grandstander as revealing a deficiency in either sensitivity to moral concerns or commitment to morality itself” (p. 208).

Tosi and Warmke (2016, pp. 210-212) argue that when virtue signaling is rampant these five features negatively impact public moral discourse in three ways. First, grandstanding leads to increased *cynicism* about morality. When people abuse moral talk to promote themselves, it can cause other people to become disillusioned with moral claims in general. It can lead to people becoming skeptical of others’ motives and over whether altruistic or moral acts are possible. Second, grandstanding can lead to *outrage exhaustion*. Heightened emotions—when working properly—can serve as a kind of spotlight for genuine moral problems. But when such outrage is frequently faked or exaggerated, people can become desensitized to such cues, which can then lose their effectiveness. Lastly, grandstanding can lead to *group polarization*, that is, “the phenomenon by which members of a deliberating group tend to move toward more extreme viewpoints” (pp. 211-212). The problem isn’t that extreme views are, in a vacuum, less likely to be true than more moderate views. Instead, the worry is that when extreme views are formed via ramping up and

trumping up effects they will be motivated—not out of the pursuit of truth—but via peoples’ desire to be recognized as high status within particular groups. As a result, this leads to an increased likelihood that group members will hold false views. Furthermore, it can lead to a scenario where rival groups end up holding incommensurable views. Of course, token instances of grandstanding can sometimes avoid these worries and cause morally positive outcomes. However, Tosi and Warmke hold that this is typically not the case.

2. Levy on Virtue Signaling

Recently, several authors have jettisoned talk of “grandstanding” in favor of the recently popularized (and for our purposes equivalent) phrase “virtue signaling.” This includes Neil Levy (2021), who has defended virtue signaling against Tosi and Warmke’s charges, and argued that virtue signaling is, in fact, “virtuous.” He argues for his position on three main grounds, which we’ll consider in turn.

First, whereas Tosi and Warmke argue that virtue signaling undermines the deliberative function of public moral discourse—or, the function of “changing minds about the moral properties of the world and thereby sometimes changing the world itself”—Levy suggests that such signaling actually supports this function (2021, p. 9548). According to Tosi and Warmke, virtue signaling undermines this function by changing minds about the moral properties of the world specifically by way of an objectionable mechanism, which is unlikely to lead to properly justified beliefs. In particular, whereas paradigm cases of rational deliberation about the truth of a moral proposition involve subjects appropriately responding to evidence and arguments bearing on the truth of that proposition, virtue signaling produces changes in belief by way of “social comparison,” or a desire to adopt whatever position would impress one’s in-group. In this way,

the worry goes, the virtue signaler can produce changes in belief without any changes in evidence, which seems epistemically objectionable.

As Levy sees it, though, virtue signaling in fact supports the deliberative function of public moral discourse. It does this, he thinks, by offering *higher-order evidence*, or evidence “about the reliability of the processes that generate belief” (2021, p. 9548). If, for instance, Ann and Ben disagree as to the answer of some complex mathematical problem, the fact *that Ben is severely sleep deprived* is higher-order evidence that his judgment is less reliable than Ann’s. Similarly, Levy thinks, virtue-signaling provides higher-order evidence for the truth of certain judgments. And it does this by expressing *confidence* in, and contributing to the expressed *number* of, such judgments.

Here, Levy appeals to cases where subjects change their moral judgments in accordance with the testimony of others. When a subject is faced with testimony of an opinion expressed with greater confidence, or simply faced with more instances of testimony of that opinion, the subject will tend to give more credence to that opinion. Whereas Tosi and Warmke suggest that such updating is epistemically objectionable, since the evidence it relies on does not bear directly on the truth of the opinion itself, Levy notes that it is in fact perfectly rational, as both the number of people offering testimony in favor of an opinion, and the confidence with which they express it, provide valuable higher-order evidence for the truth of the opinion. If all three of one’s fellow interlocutors object to one’s opinion that abortion is wrong, for instance, and do so with great confidence, one *should* update one’s opinion accordingly. After all, one now has the higher-order evidence that the processes generating the dissenting opinion are reliable, making the probability of the dissenting opinion’s truth accordingly higher. Hence, it seems, by offering higher-order

evidence for the truth of moral propositions, virtue signaling in fact supports the deliberative function of public moral discourse.

Second, Levy argues that a further positive feature of virtue signaling is that it helps us delineate groups of reliable moral cooperators. As Levy notes, it is likely that public moral discourse has not just one particular function, but multiple. And another of these functions, he thinks, is plausibly that of helping us solve coordination problems. In particular, as social beings, our ability to share information and thereby coordinate behavior is crucial. As the groups with which we coordinate become larger, however, the risk of exploitation by free-riders becomes greater: some may go unnoticed in benefitting from the coordination while paying none of the costs. In a society like ours, where there are very many large social groups that can easily be moved between, it is no longer practicable for us to track the reputations of each member to try to prevent free-riding. As a result, Levy suggests, one way we might delineate groups of reliable moral cooperators is by *signaling* to others that we are trustworthy. Virtue signaling may thus support another key function of public moral discourse.

In support of this, Levy appeals to the function and efficacy of signaling in the practices of both non-human animals and religious followers. When signals are costly or credibility enhancing, he notes, they can be reliable sources of evidence of willingness to cooperate. Regarding non-human animals: “For instance, gazelles famously signal their fitness by stotting (jumping up and down on the spot) in front of predators (FitzGibbon and Fanshawe 1988). Peacocks even more famously signal their fitness with their spectacular tails (Zahavi and Zahavi 1999)” (Levy 2021, p. 9553). And regarding religious followers:

Many cognitive scientists argue that costly signalling is at the root of a variety of religious practises (Irons 2001; Sosis and Alcorta 2003; Sosis and Bressler 2003). Regular attendance at religious services is costly, insofar as it requires forgoing

more immediately rewarding activities. More directly, tithing is costly and religious rituals often involve some kind of privation. Fasting is a common signal of religious commitment (Lent, Ramadan and Yom Kippur all involve fasting, of course) (Levy 2021, pp. 9553-9554)

Similarly, Levy thinks, virtue signaling provides reliable evidence of willingness to cooperate. He writes:

The facial and bodily expressions of emotion are involuntary and therefore hard to fake: blushing and flushing are classic examples of typically involuntary, and therefore hard to fake, expressions of emotion. Virtue signalling is often accompanied by, perhaps even partially constituted by, strong emotions (“excessive outrage or other strong emotions”; Tosi and Warmke 2016, p. 206). At least when these signs are visible or otherwise perceptible, virtue signalling involves hard to fake signals. These signals are also potentially costly, inasmuch as in committing oneself to a moral position opens one up to condemnation if one fails to act consistently with it. (2021, p. 9555)

Accordingly, virtue signaling seems to play the further positive role of helping us delineate groups of reliable moral cooperators.

Third and finally, Levy defends virtue signaling against the charge that it involves hypocrisy on the part of those who engage in it. The charge here is that there is a mismatch between what a virtue signaler claims to be concerned with, namely the moral issue in question, and what they are actually concerned with, namely their own reputation. Here too, Levy thinks, a parallel can be drawn with religious followers. In particular, if he is right that signaling is at the root of many religious practices, or partly constitute their function, and yet followers who engage in these practices are in no relevant sense hypocrites, then the same will seem true of virtue signalers. After all, just as the fact that signaling may be a function of religious practices does not imply that those engaging in them do so specifically in order to signal, so the fact that virtue signaling may be a function of public moral discourse does not imply that those virtue signaling do so specifically in order to virtue signal. Thus, there need not be any hypocrisy involved in what the virtue signaler claims, on the one hand, and their motivations for making the claim, on the other.

2.1 Evaluating Levy on Virtue Signaling

Admittedly, Levy's defenses of virtue signaling have a good deal of initial plausibility. However, we think that he overlooks a distinction that, once appreciated, shows that what he ultimately defends does not seem to be "virtue signaling" at all.

Specifically, we believe there is an important yet (at least in Levy's case) unappreciated distinction to be made between *virtue signaling* and, more generally, *norm signaling*. Virtue signaling entails norm signaling, but the reverse is not true. Virtue signaling, in the relevant sense, is essentially just what Tosi and Warmke describe as moral grandstanding. It centrally involves the expression or communication of one's being committed to a certain norm, *alongside* a (partly) motivating recognition desire. Importantly, this analysis is meant (and seems) to properly capture what we mean in everyday conversation by "virtue signaling"—that is, it captures our basic folk conception of the concept. Norm signaling, on the other hand, centrally involves just one part of virtue signaling: namely, the expression or communication of one's being committed to a certain norm.⁷ We think that what makes Levy's defense of "virtue signaling" seem attractive is best explained in terms of this distinction. In particular, recall that in defending "virtue signaling," Levy appeals to (i) its providing higher-order evidence for the truth of certain moral judgments, specifically by expressing confidence in, and contributing to the expressed number of, such

⁷ Here it is worth noting that, as in the case of virtue signaling, norm signaling may come in a variety of forms. One may norm signal merely by avowing belief in a given norm. Alternatively, one may norm signal by performing a costly act associated with upholding the norm (as we will see in the religious cases below). In each case, what norm signaling ultimately signals is that one is confident in or accepts a given norm. The evidence others take from this, though—as, again, in the case of virtue signaling—may depend on the form of norm signaling in question, as we'll see below.

More generally, what is important here is that for any form of virtue signaling, there is a form of norm signaling with all the same features except for the recognition desire. And so, more specifically, for any case of virtue signaling that Levy appeals to (whether it involves actual acceptance of a norm, a given amount of confidence in an expressed belief, a willingness to pay a cost, etc.), we can imagine a case of norm signaling with all the same relevant features but without the recognition desire. We thank an anonymous reviewer for pressing us to clarify here.

judgments; (ii) its helping us delineate groups of reliable moral cooperators; and (iii) its not involving any hypocrisy on the part of its subject. In each case, we claim that what seems to support his defense is an appeal to norm signaling, as opposed to virtue signaling (properly understood). Let us explain.

Regarding (i), it seems Levy is right both that virtue signaling provides higher-order evidence for the truth of moral judgments, and that it does this by expressing confidence in, and contributing to the expressed number of, such judgments. Notice, though, that it seems to do this only in virtue of its being an instance of signaling of a more general kind—namely, norm signaling. That is, norm signaling—or the expression of one’s being committed to a certain moral norm—is by itself what provides higher-order evidence for the truth of moral judgments. When someone norm signals—say, about the permissibility of abortion—we rightly take this as evidence for the truth of the relevant proposition. When someone norm signals with greater confidence, we rightly take this as stronger evidence for the truth of the proposition. And when more people signal their acceptance of the same norm, we rightly take this as stronger evidence for the truth of the proposition. All of this is the case, regardless of whether the norm signaler also has a recognition desire that partly motivates their signaling—or, regardless of whether the norm signaler is a *virtue* signaler. Thus, it seems that virtue signaling contributes to the deliberative function of moral discourse in the way Levy describes just to the extent that *norm* signaling contributes to this function, and virtue signaling entails norm signaling. And, thus, whatever contribution virtue signaling makes to this function can be made by signaling without a relevant recognition desire. As far as the deliberative function of public moral discourse is concerned, virtue signaling, to the extent it differs from norm signaling, simply isn’t needed.

Regarding (ii), a similar point seems true. Levy plausibly claims that another function of public moral discourse is that of helping us delineate groups of reliable moral cooperators. And one effective way of doing this, he thinks, is by virtue signaling. Here, Levy appeals to signaling in the practices of non-human animals and religious followers. For one, though, notice that the contribution Levy claims virtue signaling makes to this function of moral discourse can again be attributed, more simply, to norm signaling. Specifically, the idea is that where virtue signaling is a costly and credibility enhancing type of signal, it can be a reliable source of evidence of a willingness to cooperate with others. Again, though, whatever costs or credibility enhancements that are involved in virtue signaling are involved, as well, in norm signaling more generally. Genuine norm signaling will equally involve an expression of emotion, and so involve attendant involuntary facial and bodily expressions. It will be equally costly in opening one up to potential condemnation, as well. It seems, then, that virtue signaling will be a reliable source of evidence of a willingness to cooperate with others just to the extent that norm signaling is a reliable source of evidence of a willingness to cooperate with others, and virtue signaling entails norm signaling. Again, as far as this function of public moral discourse is concerned, virtue signaling, to the extent it differs from norm signaling, simply isn't needed.

And for another, with this distinction between norm signaling and virtue signaling in hand, we can see that Levy's analogies with non-human animals and religious followers are actually inapt. True, animals like peacocks signal to express fitness, and religious followers signal to express their religious commitment. In each case, however, while the signal may have the function of expressing a willingness to cooperate, this function can be performed without any additional

desire specifically for recognition on the part of those who express the signal.⁸ Those signaling commitment to their religion, for example, may do so simply by regularly attending religious services, tithing, and fasting, where these things signal no more than their commitment to religious norms. Similarly, peacocks signaling their fitness may do so simply by displaying their tails, where this signals no more than their potential for being a good mate. There need not be any additional desire *specifically for recognition for* being committed (or, at least, for expressing commitment) to these norms. The religious follower may regularly attend religious services and thereby signal their commitment to religious norms, without also explicitly desiring recognition from others for what they signal their commitment to—or, they may do so without *virtue* signaling. Indeed, it seems we should expect that for many religious followers, this added egoistic desire plays no role in their religious practices. And, in the case of animals like peacocks, we should certainly expect that no such desire plays a role in their motivations, as beings like these simply seem incapable of desiring such complex things. It again seems, then, that as far as promoting cooperation is concerned, we can explain everything Levy describes merely in terms of norm signaling.

Finally, regarding (iii), Levy's denial of any hypocrisy on the part of virtue signalers again relies on an analogy with religious followers that seems inapt. According to Levy, just as the fact that signaling may be a function of religious practices does not imply that those engaging in them do so specifically in order to signal, so the fact that virtue signaling may be a function of public moral discourse does not imply that those virtue signaling do so specifically in order to virtue signal. Notice, however, that the type of signaling that is plausibly a function of religious practices here is—as we've just seen—merely norm signaling, as opposed to virtue signaling. And,

⁸ In fact, Jesus tells his followers not to virtue signal: “Therefore, when you do a charitable deed, do not sound a trumpet before you as the hypocrites do in the synagogues and in the streets, that they may have glory from men” (Matthew 6:2).

analogously, the type of signaling that plausibly supports the cooperative function of public moral discourse is merely norm signaling, as opposed to virtue signaling. So, the core of Levy's thought is right: the fact that some type of signaling may support the cooperative function of public moral discourse does not imply that those signaling do so specifically in order to signal. However, the relevant type of signaling here is *norm* signaling, not virtue signaling. In fact, we already know that, contrary to what Levy suggests, those virtue signaling *do* in fact do so (at least in part) specifically in order to virtue signal: they do so at least partly in order to get recognition for what they express. That's just what virtue signaling is. But it's not, of course, what norm signaling is.

It seems, then, that the plausibility of Levy's defense of "virtue signaling" is best explained by the fact that what he ultimately defends is in fact *norm* signaling, though presented under the guise of "virtue signaling." Every good feature he ascribes to virtue signaling—its providing higher-order evidence for the truth of certain moral judgments, its helping us delineate groups of reliable moral cooperators, and its not involving any hypocrisy on the part of its subject—is actually a good feature of norm signaling. In fact, it seems these good features are ascribable to virtue signaling *ultimately and only because* they are good features of norm signaling, and virtue signaling entails norm signaling. Thus, it again seems that in each case, virtue signaling, to the extent it differs from norm signaling, simply isn't needed.

3. Westra on Virtue Signaling

Another recent defense of virtue signaling comes from Evan Westra (2021). While Westra is notably more careful in distinguishing between virtue signaling and the straightforward expression of commitment to social norms (that is, norm signaling), at least some of his points seem to overlap with Levy's. He rightly points out that virtue signaling provides evidence for the truth of certain moral judgments or social norms and that since virtue signalers will often be pressured to "practice

what they preach,” we should not expect the practice to create widespread hypocrisy of the sort that could undermine the functions of public moral discourse (2021, pp. 164-169). As we’ve seen in our evaluation of Levy, however, the same things may be true of more basic norm signaling: whatever benefits virtue signaling may provide, it does so solely in virtue of being a form of norm signaling. Virtue signaling, to the extent it differs from norm signaling, simply isn’t needed.

More important for our purposes, then, is a rather different defense of virtue signaling that Westra provides. According to Westra, there are some cases where virtue signaling can spur social progress in a way that mere norm signaling could not (or, at least, would not, or would not as quickly). If that’s right, then there seem to be cases that undermine our main point about virtue signaling thus far, or our point that for any benefit virtue signaling seems to provide, it can do so without involving any recognition desire: the same benefit would be produced if it were replaced by mere norm signaling.

The cases in question are inspired by the point that, generally, “most ordinary folk are of decidedly uneven moral character, motivated by a mix of egoistic and moral reasons and rarely aiming for more than moral mediocrity” (Westra 2021, p. 166). Focusing on the case of flight-shaming, Westra notes that whereas a jet-setting lifestyle may once have been seen as glamorous or enviable, it is now more commonly seen as morally dubious. And, importantly, whereas the first few flight-shamers who sparked this social change may have been motivated strictly by their sincere moral concerns, Westra notes: “It is likely that a number of people who subsequently engaged in flight-shaming were motivated at least in part by reputational considerations, which exert subtle, unconscious influence over such acts of moralistic punishment. But for the purposes of changing social norms, this is a good thing” (2021, p. 166). It is a good thing, of course, because it accelerates when—or, indeed, may even make a difference as to whether—the positive change

in social norm occurs. After all, if the virtue signaling flight-shamers had not participated in the moral discourse, it is unclear when our conception of a jet-setting lifestyle might have changed, or whether it would have at all. Since “most ordinary folk” are not motivated strictly by their sincere moral concerns and are instead more likely to be motivated partly by reputational considerations, it seems the progress here may not have been realized (as quickly) without their efforts—or, without their virtue signaling. Accordingly, as Westra puts it: “By appealing to human beings’ deeply engrained disposition to enhance their moral reputations, virtue signaling exploits a much more effective channel for transmitting social information throughout the broader population than appeals to pure virtue ever could” (2021, p. 167). In that case, we seem to have reason for thinking virtue signaling can indeed produce a benefit that norm signaling alone could not. Perhaps virtue signaling is uniquely effective in spurring moral progress.

3.1 Evaluating Westra on Virtue Signaling

It may be true that, as a mere empirical fact, virtue signaling has or would ensure that some moral progress was realized more quickly than it would otherwise have been. Perhaps if no virtue signalers had flight-shamed, for instance, the movement simply wouldn’t have occurred, or at least occurred yet. This is certainly possible. But does this give us reason to think there is a uniquely positive effect on moral progress, or on public moral discourse, that virtue signaling provides? That is, does virtue signaling’s role here make it such that it ultimately benefits public moral discourse in a way that mere norm signaling does not? We think the answer is No. This is because, while virtue signaling may play this positive instrumental role in bringing about moral progress in cases like flight-shaming, it may just as easily, and in just the same way, play a negative instrumental role in bringing about moral *regression* in other cases. Of course, virtue signalers themselves will presumably never be expressing their commitment to a norm they take to

contribute to moral regression—it is not that regression is their aim. Instead, the problem is that what counts as moral progress to one depends on what counts as moral progress to one’s in-group. And, when one’s in-group’s conception of what counts as moral progress conflicts with what *actually* counts as moral progress, it seems virtue signaling may play the very same role it does in cases of (actual) moral progress yet be equally harmful.

To see this, consider two especially relevant moral issues in the United States at present: abortion access and gun control. If one’s in-group is a set of conservative religious followers, and the debate is over introducing legislation restricting women’s access to abortions, it is easy to see how virtue signaling may serve only to amplify what is in fact (at least in our view) moral regression. We can imagine, for instance, that debate in one’s in-group begins from the majority opinion that abortion should be outlawed except in cases of rape, incest, or to preserve the life and physical health of the mother, before the 22nd week of pregnancy. For the virtue signaler, however, this is not enough: partly in hopes of gaining recognition, they may insist that abortion should be outlawed except in the mentioned cases before the tenth week of pregnancy. After all, they may claim, a fetus after the tenth week of pregnancy is a person. Then, a fellow virtue signaler—again partly in hopes of gaining recognition—may insist on a further step: a fetus is a person, not after the tenth week, but upon conception. Thus, this person may insist, abortion should be outlawed without exception. The act of aborting an unborn person is never permissible, they may claim, even if this means we must sometimes “allow” the mother to be gravely harmed or die. Ultimately, some members of the in-group are not convinced by the virtue signalers’ arguments, but others are. Either way, the in-group’s conception of permissible abortion access has become more restrictive—thanks, we can suppose, to the virtue signalers’ efforts.

Similarly, consider gun control. Suppose one's in-group is a set of hunters and gun owners in a conservative state, and the debate is over introducing legislation, inspired by a recent school shooting, to restrict people's access to certain semi-automatic guns. Suppose, moreover, that debate in one's in-group begins from the majority opinion that semi-automatic guns should be readily available to anyone of age who passes a modest background check. After the school shooting, however, at least some members of the group seem more willing to cede further restrictions, like more extensive background or mental health checks, for accessing certain guns. Other members, meanwhile, stick to their guns (literally). Those potentially sympathetic to the restrictions point to the school shooting and bemoan the fact that it might have been prevented. Those unsympathetic to the restrictions appeal once more to their adage that "Guns don't kill people, people kill people." And, here again, the virtue signaler may arrive, serving only to amplify what is in fact (at least in our view) moral regression. Motivated partly by their hope of gaining recognition, they may insist that this restriction is just the first on a slippery slope that ends with all their guns being taken away. They may then claim that what is really to be bemoaned is the fact that lawmakers have "politicized" the school shooting as a way to restrict access to guns—that the school shooting is really what the opposition *wanted*. As a result, they may insist, any further restrictions on access to guns should be resisted. Then, a fellow virtue signaler—again motivated partly by their desire for recognition—may insist on a further step: that if anything, restrictions on access to guns should be entirely abolished. For support, after all, we need look no further than the Second Amendment. The fact that others act wrongly provides no reason to restrict *my* basic right. Ultimately, some members of the in-group are not convinced by the virtue signalers' arguments, but others are. Either way, the in-group's conception of permissible access to semi-automatic guns has become less restrictive—thanks again, we can suppose, to the virtue signalers' efforts.

Our point, then, is that while Westra may be right that some moral progress has or would require virtue signaling to be (as quickly or efficiently) realized, the same is true of some moral regression. If virtue signaling uniquely benefits public moral discourse by supporting moral progress in a way that mere norm signaling does not, then it also uniquely harms public moral discourse by supporting moral regression in a way that mere norm signaling does not. Of course, whether these unique benefits and harms will ultimately be positive or negative in total is an empirical matter. But we seem to have little reason to think, in line with the defender of virtue signaling, that the benefits would ultimately outweigh the harms. After all, if people's beliefs move in whatever direction appeals to their in-group, it would be quite lucky to find that this resulted in moral progress, or in more true beliefs.⁹ Even if the results were more modest, though, and these unique effects of virtue signaling roughly canceled out, our main point would remain. It seems, again, that there is no (net) benefit that virtue signaling provides that norm signaling does not.

4. Virtue Signaling and Higher-Order Evidence

So far, we have argued that virtue signaling isn't needed. Whatever benefits the practice might bring about are constitutive of norm signaling; virtue signaling *qua* virtue signaling does not have a net positive effect on moral progress. In this section, we return to Levy's claim that virtue signaling can have a positive effect on moral discourse in that it serves as higher-order evidence. This is false when it comes to half-hearted or cold-hearted virtue signaling. Furthermore, we argue that virtue signaling—even of the whole-hearted variety—will typically have a negative effect when it comes to discovering the truth. In this way, virtue signaling—unlike norm signaling—is epistemically problematic.

⁹ We thank two referees for helping us with these points about cumulative benefits and harms.

Higher-order evidence is, again, “evidence about the reliability of the processes that generate belief”, and Levy claims that “Virtue signalling provides higher-order evidence by conveying *confidence* and the *numbers* of people who share a judgment” (p. 9548, emphasis original). Thus, Levy claims that virtue signaling is neither irrational nor arational but, instead, contributes to the rational evaluation of moral claims by providing higher-order evidence. For example, according to Levy, when millions of people change their profile pictures in support of Black Lives Matter and make passionate tweets in support of the movement, this provides us with higher-order evidence that the cause’s tenets are correct.

Levy supports these claims via an appeal to two popular examples in the disagreement literature. First, consider *Directions*:

Suppose you’re looking for the railway station in an unfamiliar town and you stop someone to ask for directions. A confident response (“straight ahead two blocks and then first right; you can’t miss it”) will probably have you on your way, whereas a tentative response (“I think it might be down that street?”) will have you seeking another opinion. (2021, p. 9548)

This example shows that the confidence with which assertions are made can, itself, count as evidence in favor of a proposition.

Second, consider *Restaurant Check*:

Anika and Bindi are old friends who eat out together once a fortnight. They always split the bill. As they always do, each calculates her share on her own, dividing the check by 2 and adding 15% to the total for a gratuity. They are both pretty good at mental arithmetic, and they almost always agree on the total. When, in the past they have disagreed, checking has shown that Anika is right about half the time. Tonight is one of those rare occasions when they disagree: Anika announces that each owes \$43, while Bindi comes up with the figure of \$45. (2021, p. 9549, based on Christensen 2007)

Intuitively, Anika should lower her confidence that she owes \$43 because Bindi who is her peer disagrees and *vice versa*. Furthermore, suppose that Anika and Bindi bring in several other people who are unbiased and competent to help calculate the bill. Here, again, numbers matter. If many

of these consultants agree that the figure adds up to \$43, then we have good evidence in favor of this claim. In such a scenario, Anika ought to become more confident that her arithmetic is correct, while Bindi should conciliate further and perhaps no longer believe that the correct figure is \$45. As Levy puts it, “The numbers on each side count, simply because it is more likely that the less numerous side has made an error than the more numerous (other things being equal, of course)” (p. 9549).

Levy then argues that our intuitions about confidence and number being epistemically relevant in *Directions* and *Restaurant Check* also apply in cases of virtue signaling. For example, Levy points out that when virtue signalers “pile on” what they are doing is letting us know how many people believe in the truth of some moral proposition, say, *p*. Thus, as in *Restaurant Check*, if we originally believed that *p*, then the fact that there are numerous virtue signalers who also claim that *p* entails that we should increase our confidence in *p*, whereas if we originally believed not-*p*, we now have a reason to conciliate or withhold belief. Furthermore, virtue signalers can signal their confidence that a moral proposition is true by use of “excessive outrage”, and such displays of emotion—rather than being arational—can, as in *Directions*, serve as valuable higher-order evidence. Thus, all things being equal, if there are many people who passionately advocate for the truth of a moral proposition, then this serves as higher-order evidence in favor of that belief.

We agree with Levy that confidence and numbers are epistemically relevant in *Directions* and *Restaurant Check*; however, we believe that these examples are disanalogous with instances of virtue signaling in three important ways.

First, in *Directions* and *Restaurant Check* everyone is dispassionately weighing the facts and trying to arrive at the truth. However, the virtue signaler—by definition—has ulterior motives. The virtue signaler desires recognition. She acts the way she does in large part because she desires

that other people view her in a positive light. The problem is that acting out of a desire for recognition is an unsound epistemological principle. As Tosi and Warmke put it, the virtue signaler acts based on the following principle: “adopt moral views that one believes will thereby lead those in one’s in-group to be morally impressed” (Coady et al. 2017). Such a principle will sometimes coincide with the truth or lead to good results, but this is accidental.¹⁰ Forming or expressing beliefs out of a desire to be recognized by others is not related to an evidence-based approach about how the world actually is, and, as a result, is epistemically irresponsible.

Second, in *Directions* and *Restaurant Check* everyone is forming their beliefs about the bill or location of the railway station *independent* of one another. When numerous experts come to the same conclusion about a complicated matter independently from one another, we should be confident that they are correct. Suppose Anika and Bindi bring in three consultants each of whom calculate the figure independent of each other (for example, in separate rooms) and by using different methods—one does the math in her head, one with paper and pencil, and another uses a calculator. If these consultants all agree that the total is \$43, then Anika should increase her confidence that she is correct. However, if the consultants’ work is not independent (for example, they all use the same calculator or in some way share results or influence each other), then we should be wary of “double counting” their testimony. This is because if the calculator is broken or they influence each other’s answers, then their agreement is no longer a sign that they are getting at the truth.¹¹ The problem with virtue signaling is that it takes place in a highly dependent environment that is often based around a particular group’s identity. Once a majority position in such a group becomes apparent, many virtue signalers will pile on to align themselves with popular

¹⁰ For example, if the person who would otherwise respond tentatively in *Directions* was significantly motivated to impress you by providing directions, she may answer differently despite her evidence being exactly the same.

¹¹ We would like to thank Thomas Mulligan for help with this point.

sentiment. To count such a person's testimony as always adding additional weight in favor of a position, is a mistake. Doing so amounts to a kind of double counting that is no longer based on the position's merit.

Lastly, genuine confidence is epistemically relevant, but we think that the virtue signaler often isn't genuinely confident. Genuine confidence matters insofar as it serves as a proxy for the likelihood that one's belief is correct. However, virtue signalers often show strong emotions not as a proxy for the conviction with which they hold their beliefs but to signal to others that they are morally superior or should be recognized as meeting a certain standard. Such faked or disproportional responses, which occur in instances of trumping up and excessive outrage, are no longer reliable signals. Additionally, in instances of piling on, ramping up, and self-evidential claims, any confidence that the virtue signaler displays will often be based on appealing to in-group dynamics and not on the evidence at hand. We think, then, that virtue signalers are likely to be overconfident in their convictions and, as a result, are not reliable moral guides. In fact, it seems to us that many people adopt the beliefs of certain groups specifically because they *lack confidence* about the issues and are looking for an easy heuristic to help them make decisions.

Imagine, then, a case like *Restaurant Check* but where virtue signaling is rampant. Such a case begins with Anika and Bindi's disagreement but suppose that this time the consultants that they bring in are virtue signalers. These consultants have ulterior motives that are not truth related. Perhaps they, themselves, will benefit if the bill is higher as they cash out tips with the waiters at the end of the night. Furthermore, they are not independent but work together in delivering their reports on the figure. Lastly, they are prone to excessive displays of outrage about the figure that are not representative of their actual beliefs or confidence. If you knew these facts about these consultants, you would no longer consider their testimony as higher-order evidence.

Levy would likely object that the above example is an imperfect analogy and that many real-world cases of virtue signaling lack these problematic features. Let us examine, then, the logical possibilities.

As argued in section 1, virtue signalers can be separated into three types based on whether they are being sincere, that is, as cold-hearted, half-hearted, or whole-hearted. The cold-hearted virtue signaler does not believe in the truth of her own grandstanding expression. Thus, if one knew that someone was guilty of cold-hearted virtue signaling, one would never take such a person's grandstanding expression as evidence in favor of whatever cause the virtue signaler was putatively advocating for.

The half-hearted virtue signaler is indifferent and merely paying lip service to whatever cause her grandstanding expression is about. As a result, she is likely to be strongly motivated by her desire for recognition, and her statement will be formed in a way that, as previously argued, is epistemically problematic.

Interestingly, Levy concedes these points about the cold-hearted and half-hearted virtue signaler:

If virtue signalling is to provide higher-order evidence, it *must* be honest. The outrage expressed must bear some reliable relation to the person's assessment of the moral wrong; piling on must occur in a way that actually reflects agents' judgments. It is only if virtue signalling is on the whole honest that the higher-order evidence it provides is reliable. (2021, p. 9559, emphasis original)

This means that only whole-hearted virtue signaling can count as higher-order evidence.

What then should we make of the whole-hearted virtue signaler's testimony? First, as previously argued, virtue signaling adds nothing good over and above whatever benefits there are to norm signaling. Second, the whole-hearted virtue signaler is still motivated by her recognition desire. This desire is neither evidence-based nor conducive to truth. Because of this, we think that

in many cases the whole-hearted virtue signaler is, at the least, over-confident in her beliefs. Still, it must be admitted that virtue signaling is a diffuse phenomenon. The extent to which a virtue signaler's grandstanding expression is motivated by her recognition desire as opposed to other reasons lies on a continuum, and, in theory, one could be a whole-hearted virtue signaler in the sense that, although one's statement was solely based on sound reasoning, one was still slightly disappointed not to be recognized as virtuous for making said statement. As a result, one should not dismiss the whole-hearted virtue signaler's testimony, but the degree to which one should weigh such testimony depends on the quality of the evidence and arguments given in favor of the virtue signaler's view.

However, even with this concession to Levy that some instances of whole-hearted virtue signaling do count as higher-order evidence, this raises a new problem for the defender of virtue signaling. The problem is that in the real world it is incredibly difficult to know what a person's actual motivations are. As a result, when virtue signaling is rampant, the presence of cold-hearted, half-hearted, and whole-hearted but partisan virtue signalers weakens the strength of testimony, which, in turn, has a deleterious effect on the quality of public moral discourse and can lead to widespread cynicism. In particular, it does this by introducing into public moral discourse a motivation that aims to affect the discourse but that *does not aim at the truth*. This added motivation, when rampant, *should* lead us to discount or have reduced confidence in the testimony of others, and so the higher-order evidence their testimony provides. After all, in such cases, we will be unaware of people's motivations, and yet be well aware of the (then-widespread) risks of double counting and misleading confidence.

5. Conclusion

In this paper, we have challenged two recent defenses of virtue signaling. Whereas Levy ascribes a number of good features to virtue signaling—its providing higher-order evidence for the truth of certain moral judgments, its helping us delineate groups of reliable moral cooperators, and its not involving any hypocrisy on the part of its subject—it seems these good features are ascribable to virtue signaling ultimately and only because they are good features of norm signaling, and virtue signaling entails norm signaling. Similarly, whereas Westra suggests that virtue signaling uniquely benefits public moral discourse by supporting moral progress in a way that mere norm signaling does not, it seems virtue signaling also uniquely harms public moral discourse by supporting moral regression in a way that mere norm signaling does not. It therefore seems that in each case, to the extent it differs from norm signaling, virtue signaling simply isn't needed.

Moreover, we have suggested that, if anything, virtue signaling will undermine the higher-order evidence we typically can and should rely on from the testimony of others. Virtue signaling essentially involves a motivation that aims at affecting public moral discourse but that does not aim at the truth. When virtue signaling is rampant—when we are aware that this ulterior motive is common among our peers—we should give less weight to the higher-order evidence provided by the testimony of others than we otherwise would, on pain of double counting evidence and falling for unwarranted confidence. We conclude, therefore, that not only is virtue signaling not needed, but its epistemological effects warrant its bad reputation.

References

- Christensen, D. (2007). Epistemology of disagreement: The good news. *The Philosophical Review*, 116(2), 187–217. <https://doi.org/10.1215/00318108-2006-035>.
- Coady, C. A. J., J. Tosi, & B. Warmke. (2017). Philosophy & public affairs discussion at PEA soup: Justin Tosi and Brandon Warmke's 'Moral Grandstanding,' with a Critical Précis by C.A.J. (Tony) Coady. PEA Soup. Retrieved August 21, 2022, from <https://peasoup.deptcpanel.princeton.edu/2017/08/philosophy-public-affairs-discussion-pea-soup-justin-tosi-brandon-warmkes-moral-grandstanding-critical-precis-c-j-tony-coady/>.
- FitzGibbon, C. D., & Fanshawe, J. H. (1988). Stotting in Thomson's Gazelles: An honest signal of condition. *Behavioral Ecology and Sociobiology*, 23(2), 69–74. <https://doi.org/10.1007/BF00299889>.
- Grubbs, J. B., Warmke, B., Tosi, J., James, A. S., & Campbell, W. K. (2019). Moral grandstanding in public discourse: status-seeking motives as a potential explanatory mechanism in predicting conflict. *PLoS ONE*, 14(10), e0223749. <https://doi.org/10.1371/journal.pone.0223749>.
- Irons, W. (2001). Religion as a hard-to-fake sign of commitment. In R. Nesse (Ed.), *Evolution and the capacity for commitment* (pp. 292–309). New York: Russell Sage Foundation.
- Levy, N. (2021). Virtue Signalling Is Virtuous. *Synthese*, 198, (10), 9545–62. <https://doi.org/10.1007/s11229-020-02653-9>.
- Sosis, R., & Alcorta, C. (2003). Signaling, Solidarity, and the sacred: The evolution of religious behavior. *Evolutionary Anthropology: Issues, News, and Reviews*, 12(6), 264–274. <https://doi.org/10.1002/evan.10120>.
- Sosis, R., & Bressler, E. R. (2003). Cooperation and commune longevity: A test of the costly signaling theory of religion. *Cross-Cultural Research*, 37(2), 211–239. <https://doi.org/10.1177/1069397103037002003>.
- Tosi, J., & Warmke, B. (2016). *Moral grandstanding*. *Philosophy & Public Affairs*, 44(3), 197–217. <https://doi.org/10.1111/papa.12075>.
- Tosi, J., & Warmke, B. (2020) Moral Grandstanding as a Threat to Free Expression. *Social Philosophy & Policy*, 37(2), 170–89. <https://doi.org/10.1017/S0265052521000108>.
- Westra, E. (2021). Virtue Signaling and Moral Progress. *Philosophy & Public Affairs*, 49(2), 156–78. <https://doi.org/10.1111/papa.12187>.
- Zahavi, A., & Zahavi, A. (1999). *The handicap principle: A missing piece of darwin's puzzle*. Oxford: Oxford University Press.