This is a pre-print of an article forthcoming in the Journal of Ethics and Social Philosophy (JESP), uploaded December 28th, 2021.

Vice Signaling

Olúfẹ́mi O. Táíwò

Abstract. Tosi and Warmke discuss cases where the speaker intends for the audience to take their expressions as evidence of good moral character. However, another possibility exists that similarly exploits the social communicative architecture. A contribution to public moral discourse may also attempt to strut by demonstrating evidence of *bad* moral character, by purposely failing to meet the evaluative standards of its audience—or, paradigmatically for my purposes, a particular section of its actual or notional audience. I call this kind of communication *vice signaling*. On their face, virtue signaling and vice signaling may seem to be opposites, since the labels imply that they are signaling opposite things. But certain cases of vice signaling are in fact also cases of virtue signaling. These are the cases where someone flaunts the standards of an out-group in order to demonstrate solidarity, seriousness, or some other virtue to their in-group. In this paper I attempt to describe these cases and point out the moral risks and opportunities they present.

> *I'm going to say this and I mean — down to my subatomic particles — what I say. And I actually don't care what anyone might think about it:*
> *I don't give a* FUCK *about Justine Damond and what happened to her.*
> *I don't give a fuck because most white people didn't give a fuck when police murdered seven-year-old Aiyana Stanley-Jones as she lay on a couch, sleeping. What most white people — and some black people — did was blame Aiyana's family …*
> *Most white people rely on this idea that black people, in situations where white people are in pain, are only ever to be soothing and understanding; only ever to be Mammy or Uncle Remus; only ever to extend condolences; only ever to embody loyalty; only ever to offer the empathy and sympathy that most white people purposely and haughtily deny when the situation is reversed — almost as if most white people still see us as their property.*
> *When the situation is reversed, when we require empathy and sympathy, then suddenly we're all of the opposite things that these once-needy white people previously said we were. When the shoe is on the other foot, then they assess us as immoral, violent, criminal, subhuman, unworthy.*[1]

Forty-year-old yoga instructor Justine Damond had called police to her Minneapolis suburb to report a sexual assault. Officer Mohamed Noor arrived on the scene and, for unclear reasons, opened fire on Damond, killing her—a tragedy. Yet: Son of Baldwin does not give a fuck about Justine Damond. And neither, apparently, should you.

Son of Baldwin is a writer known for his skillfully crafted and widely circulated pieces about social justice issues in the US, and is known for hot takes on various aspects of white supremacy. His writing has been controversial at times: in particular, Professor Johnny Williams at Trinity College was the target

[1] The original Son of Baldwin post was deleted from Medium. Its text is available in Starr, "I Understand Why Some Black People Couldn't Care Less About Justine Damond."

of a coordinated right-wing media campaign and placed on administrative leave for a tweet that referenced Son of Baldwin's characteristically provocative piece, "Let them Fucking Die."[2]

By itself, Damond's death is tragic but unsurprising. We are not quite sure how many people the police kill—for years, the FBI's statistics on police homicides were calculated by voluntary disclosure of police chiefs, which seems to dramatically undercount—but it is probably more common than we realize.[3]

What was surprising, on the other hand, was the response to her death. Legal consequences for police shootings are not terribly common in the US: between 2005 and 2017, only eighty officers were even arrested on charges for shootings on the job, less than half of whom were convicted.[4] Just days after the Damond killing, the police chief resigned at the mayor's public request.

Other differences between this case and other high-profile cases help explain why there were consequences of this severity in this case, and also help explain why Son of Baldwin wrote what he wrote. In several high-profile cases involving Black victims of police violence, major media outlets have released photos or reported information predictably damaging to the perceived character of the victims. A particularly egregious example is the release by CBS media of the arrest record of Alton Sterling, who was shot in the back while fleeing a police officer, in an encounter recorded on video and widely circulated.[5]

But in Justine Damond's case, media targeted the Black police officer. Meanwhile, media venerated the white victim, showing video of Damond saving ducklings from a sewer and asserting that Damond is the "most innocent victim" of a police shooting that the attorney representing her family had ever come across.[6] That last one stings: among the high-profile cases of police violence are Aiyana Stanley Jones, a Black child killed while sleeping in her bed, and Tamir Rice, a Black child killed while playing in the park.

I assume that Son of Baldwin's core audience—the "in-group" for our purposes here—is predominantly Black and other people of color angry about racial injustice. Given the preceding, we have a lot worth being resentful about. But for our purposes, the important part of this assumption about the core audience is that it helps us understand what Son of Baldwin is up to in his polemic.

To signal one's bona fides as a member of the in-group, one can contradict, mock, or otherwise flaunt the moral standards of the out-group. This is what I take it that Son of Baldwin is doing when he edgily assures us that he does not care that Damond is dead, presumably either imagining the reproach of white liberals and conservatives with his core audience or ravenously waiting for actual reactions from this peripheral audience. It is also, from a different political vantage point and with very different moral and political implications, what the person who tells racist jokes in mixed company is doing, and what the person who refuses to use a person's stated gender pronouns is doing. This helps explain why such statements earn the label "vice signaling": these statements do what they do by virtue of the fact that some disfavored out-group is taken not to like it.

In April 2015, James Bartholomew wrote a column for *The Spectator* that used the term "virtue signaling," alleging that public indications of one's personal strengths of moral character were on the rise.[7] By October of that same year, Bartholomew declared that this term (that he invented, he hastens

---

[2] Flaherty, "Trinity Suspends Targeted Professor"; Son of Baldwin, "Let Them Fucking Die."
[3] Sullivan et al., "Four Years in a Row, Police Nationwide Fatally Shoot Nearly 1,000 People."
[4] Stinson, "Police Shootings Data," 29.
[5] Media Matters Staff, "CBS Report On Police Shooting of Alton Sterling Inappropriately Highlights Victim's Record."
[6] Goyette, "Justine Damond"; Perez, "Bride-to-Be Is 'Most Innocent' Police Shooting Victim."
[7] Bartholomew, "The Awful Rise of 'Virtue Signalling.'"

to remind us) had "taken over the world," citing its use by authors with large Twitter followings and articles in well-read publications like Breitbart, *The Daily Telegraph*, and *The Independent*.[8]

The following year, Justin Tosi and Brandon Warmke wrote an article preferring the term "moral grandstanding" to virtue signaling.[9] Their initial article, and an associated blog post about it, inspired long-form responses from Eric Schliesser, Liam Kofi Bright, and Justin Weinberg.[10] Tosi and Warmke have continued to investigate the phenomenon empirically, joined by psychologists, and have found preliminary evidence in favor of their explanation of the phenomenon.[11] This piece aims to supplement their account of moral grandstanding by offering a related concept of *vice signaling*, which typically is a special case of virtue signaling or moral grandstanding rather than a different kind of contribution to public discourse altogether. Analyzing how vice signaling works, then, will help us along in understanding both moral grandstanding and public moral discourse more generally.

Tosi and Warmke discuss cases where the speaker intends for the audience to take their expressions as evidence of good moral character. However, another possibility exists that similarly exploits the social communicative architecture. A contribution to public moral discourse may also attempt to strut by purposely failing to meet the evaluative standards of its audience—or, paradigmatically for my purposes, a particular section of its actual or notional audience. Typically, this strutting takes the form of flaunting or violating out-group standards, behaving viciously or injuriously by the lights of an out-group. I call this kind of communication *vice signaling*.

In both an article in *Psychology Today* and in their recently published book on the topic, Tosi and Warmke argue against use of the terms "virtue signaling" and "vice signaling."[12] They maintain that "signaling" language is misleading since many signaling behaviors are unintentional, and moral grandstanding involves deliberate attempts to draw attention to one's self and affect how one is thought about by others.[13] They also anticipate the connection I aim to make here, to "vice signaling," but argue that debates about "virtue signaling" versus "vice signaling" would lead to "pointless arguments" about whether an action is best considered virtue signaling or vice signaling depending on "whether they are expressing good or bad values."[14] They do not say why the arguments would be pointless, but advise the reader to notice that either would fall into moral grandstanding as they define it: the combination of wanting to impress others with one's moral qualities ("recognition desire") and the attempt to satisfy this desire by way of "saying something in public moral discourse" ("grandstanding expression").[15]

My discussion here avoids these particular pitfalls. Since I take vice signaling to be, typically, a "special case" of virtue signaling, I agree that there is little to be gained from arguing which cases are which, or whether and to what extent the acts are good or bad. Accordingly, I will treat the terms "virtue signaling" and "moral grandstanding" interchangeably throughout this piece. The contrast between virtue signaling/moral grandstanding and vice signaling is instead used constructively, to build a more full picture of the stakes and dynamics of communication in public moral discourse, rather than to haggle about how to characterize individual cases. Moreover, since much of the discussion to come

---

[8] Bartholomew, "I Invented 'Virtue Signalling.'"

[9] Tosi and Warmke, "Moral Grandstanding." I use the term virtue signaling to draw out the intended parallel with vice signaling, which is key to the central aim of this paper. Tosi and Warmke express skepticism but stop short of denying that moral grandstanding and virtue signaling refer to the same phenomenon. I will generally use the terms interchangeably unless referring to their work specifically.

[10] Weinberg, "A Surprising Instance of Performative Philosophy"; Krishnamurthy, "Featured Philosopher."

[11] Grubbs et al., "Moral Grandstanding in Public Discourse."

[12] Thanks to an anonymous reviewer for pushing me to respond directly to this point.

[13] Grubbs et al., "Moral Grandstanding and Virtue Signaling."

[14] Tosi and Warmke, *Grandstanding*, 37–40.

[15] Tosi and Warmke, *Grandstanding*, 15.

appeals to social effects and dynamics that are likely outside of the conscious view of vice signalers, the fact that "signaling" encompasses both witting and unwitting forms of communication figures into this discussion as a feature, not as a bug.[16]

But my discussion also makes out the difference between virtue and vice signaling in a different way than Tosi and Warmke anticipate. Whether or not the values one expresses are "good or bad" full stop is not the difference between virtue signaling and vice signaling. People vice signal by behaving in a way that they expect out-group members to find injurious or vicious, and expect to thereby perform virtue and curry favor in the in-group.

This way of explaining vice signaling leaves open the question of whether or not the behavior is vicious or virtuous full stop in favor of an explanation where the act seems vicious to the out-group, and this very fact helps constitute it as virtuous for the in-group. The actual moral evaluation of the act itself—whether it is virtuous or vicious from the standpoint of morality, or a more cosmopolitan and less partisan perspective—plays no clear role in this aspect of social life. This is, arguably, is what is going on in the Son of Baldwin case: the moral fact about whether it makes any sense to curse a woman after her death is rendered secondary at best to the more salient fact that doing so will infuriate some out-group (presumably, white liberals who are insufficiently permissive of Black rage).

Whether we characterize such communicative acts as simple virtue signaling or also as vice signaling will depend on which sections of the evaluative community we take to be salient. In this paper I attempt to describe these cases, and point out the moral risks and opportunities they present.

1. Describing Vice Signaling

On their face, virtue and vice signaling may seem to be opposites, since the labels imply that they are signaling opposite things. But the Son of Baldwin case helps bring out the important point further suggested by the umbrella term "moral grandstanding": not only is vice signaling not the opposite of virtue signaling, but an important set of cases of vice signaling are in fact also cases of virtue signaling. These are the cases where someone flaunts the standards of an out-group in order to demonstrate solidarity, seriousness, or some other virtue to their in-group. This could help flesh out the connections investigated by Marcus Arvan between group polarization and moral discourse, which is often used to virtue and vice signal.[17]

Tosi and Warmke initially defined moral grandstanding as what one does when "one makes a contribution to public moral discourse that … attempts to get others to make certain desired judgments about oneself, namely, that one is worthy of respect or admiration because one has some particular moral quality."[18] Here, "public moral discourse" is "communication that is intended to bring some moral matter to public consciousness," in contrast to private moral discourse that is not intended for a wider audience.[19]

Vice signaling works by exploiting public information, much like more well-studied phenomena like assertions or questions. But, unlike assertion, vice signaling does not characteristically target the subject under discussion (in the case of conversation). Rather, the point of vice signaling is to change the social architecture that provides the scaffolding for conversation. To see how vice signaling works, it will help to revisit fundamental aspects of communication.

---

[16] Thanks to an anonymous reviewer for calling my attention to this point.

[17] Arvan, "The Dark Side of Morality."

[18] Tosi and Warmke, "Moral Grandstanding," 199.

[19] Tosi and Warmke, "Moral Grandstanding," 197.

When someone communicates, they presuppose things. It is hard to see how interesting conversation could get off the ground if we had to rebuild a shared understanding of the world (including language itself!) from the ground up anew every single time. One aspect of a communicator's presuppositions is that at least some information is treated as public: that is, as available to other communicators for use in reasoning and other acts.[20] Such information makes up the content of what Robert Stalnaker calls the *common ground*.[21] The common ground is the set of background information we treat as mutual knowledge for, at least, the duration of the conversation. This set is neither all of the things that I know about the world nor the set of things that you know, but the set of things that I know that you know that I know that you know, *ad infinitum*. This is also the social architecture targeted by acts of virtue signaling and vice signaling.

Having the common ground as a communicative resource makes the kind of information-rich discussion that makes conversation possible, and, where we are clever and lucky enough, interesting. The common ground, as I analyze it, is not simply a list of things publicly taken to be the case. It also provides the set of expectations against which people guess which uses of public information will be accepted or rejected, valorized or shamed. The common ground thus understood is not simply a resource but also an incentive structure, and thus in a structural sense a causal structure.[22] When one acts communicatively, one updates the common ground—that is, one changes what information serves as public practical premises for the parties in conversation.

The paradigm communicative acts are those whose essential purpose is communicative: utterances, speech acts, signs (in sign language), gestures. But other kinds of acts also communicate. Remaining seated when one is expected to get up may communicate disdain and protest (say, if someone is singing the national anthem); a slap may communicate insult; and changing one's behavioral response to a claim communicated by another may not only communicate the like belief in the accepter but also respect for the person making the recommendation. Even though these acts are not speech acts, these acts also can communicate in that they can affect what social information is public—that is, the content of the common ground—through inferences that one makes about the significance of these actions and relies upon others making. When we speak of an action's communicative effects, we could reformulate that question as a question about what changes it caused to the common ground.

To investigate and characterize the communicative effects of an action, it will matter what was already in the common ground. There has been much discussion about how the content of the common ground determines or affects uptake of what is said or communicated, especially when the bare intelligibility of the act depends on particular presuppositions, in the way that "the present king of France is bald" might rely on a presupposition that France presently has a king.[23]

But when we communicate we are not just trying to transfer information, or tell others about what the world is already like. We are often also trying to change that world, or prevent unwelcome changes to it. We may seek to inspire, motivate, or agitate for a variety of ends. We may be trying to align preferences or objectives with others, or remind people of these commitments if they have forgotten or

---

[20] Stalnaker describes the content of the common ground as "mutual knowledge." But in his more careful moments, Stalnaker admits that we often treat things on the model of mutual knowledge even when we do not mutually know them: for example, when we suppose things for the sake of argument, or, along the lines I prefer to investigate, when we use the reasoning of a higher status person or theory because I do not want to take the social risks of challenging the view. I am indebted to Dan Zeman for this point.

[21] Stalnaker, "Common Ground."

[22] I discuss these aspects of the common ground under the heading of "agenda setting effects" at greater length in Táíwò, "The Empire Has No Clothes." The sense of structural causation used here is discussed in Malinsky, "Intervening on Structure."

[23] See, for example, Potts, "Presupposition and Implicature"; Abbott, "Presuppositions and Common Ground"; Stanley, *How Propaganda Works*.

(in our estimation) are failing to live up to them. Some communicative goals may center around concepts or ideas, even those that may not be perceptible at the level of granularity needed to evaluate an utterance's truth value. For example, a sentence explaining the results of a particular experiment may also be an attempt to establish the correctness or usefulness of the larger theory the experiment was designed to help establish, and recognition of that larger goal may be an important part of understanding what is socially at stake in communicating that particular sentence.

One aspect of the world that communicative acts can affect is the standing of things in relevant social categories and hierarchies within, among, and between them, whether those things are explanations, goals, or people. To the extent that information about these categories and hierarchies is public, they are also objects of public coordination and thus embedded in the content of the common ground in some sense or other. For example, a person's location in a prestige hierarchy may affect how the common ground updates in response to their speech. A full-fledged medical doctor's claim that a patient has cancer may affect her willingness to undergo chemotherapy in the way that an equivalent claim made by the patient's accountant would not; should she get another opinion, she will likely do so from another doctor rather than an accountant. Also, she may make use of differences in prestige to settle which doctor's claim to treat as a practical premise in the event that the doctors' claims conflict.

The aforementioned helps us more precisely distinguish vice signaling as a specific subset of virtue-signaling cases. Generally, virtue-signaling communicative acts are those that attempt to affect the location of the speaker in the social locations embedded in the common ground in desired ways by way of performing well by the lights of some public set of evaluative standards, paradigmatically those endorsed by the group one views as an in-group. Vice-signaling communicative acts are those virtue-signaling acts that aim to increase the speaker's prestige or standing in a specific way: by performing badly by the lights of a public set of evaluative standards ascribed to a disfavored out-group by the in-group.[24] This fits squarely into Tosi and Warmke's characterization of the root social explanation, which is the effect the speaker aims to have on their standing and prestige in the company of their audience.

This also helps us resist the temptation to view vice signaling and virtue signaling as opposites. Since our public information may allow for a multiplicity of groups, the same speech act may virtue signal when evaluated with respect to one group's preferred evaluative standards and vice signal when evaluated with respect to another group's. In the central cases of virtue-signaling-as-vice-signaling cases, like the Son of Baldwin case given in the introduction, it is *precisely because* an act is thought to vice signal with respect to the out-group's standards that it functions as virtue signaling in the in-group.

Moreover, since intergroup conflict is at the heart of this characterization of vice signaling, the distinction between virtue signaling and vice signaling is of clear interest to philosophers concerned about political polarization and other aspects of the social dynamics and consequences of this behavior, as Tosi and Warmke clearly are.[25] The more antagonistic the relationship between the in-group and the out-group, the likelier that inflaming the out-group will be sufficient grounds for one's action being received positively by the in-group.

With this picture of how vice signaling works in individual conversational interactions, I will point out two potentially positive functions of the practice and two potentially negative ones in section 2.

2. Evaluating Vice Signaling

[24] Of course, an individual may simply wish to signal hostility at an audience without wanting to thereby affect some in group, or even without there being an in group to thereby affect. I do not focus on these cases here.

[25] The new book devotes a full chapter to discussion of these: Tosi and Warmke, *Grandstanding*, ch. 4.

Thucydides provides a helpful early discussion of vice signaling and related problems in his discussion of conflict in Ancient Greece:

> The meanings of words had no longer the same relation to things, but were changed by them as they thought proper. Reckless daring was held to be loyal courage; prudent delay was the excuse of a coward; moderation was the disguise of unmanly weakness; to know everything was to do nothing. Frantic energy was the true quality of a man … the lover of violence was always trusted, and his opponent suspected …. He who plotted from the first to have nothing to do with plots was a breaker-up of parties and a poltroon who was afraid of the enemy. In a word, he who could outstrip another in a bad action was applauded; and so was he who encouraged to evil one who had no idea of it.
>
> The tie of party was stronger than the tie of blood, because a partisan was more ready to dare without asking why …. The seal of good faith was not divine law, but fellowship in crime. If an enemy when he was in the ascendant offered fair words, the opposite party received them not in a generous spirit, but by a jealous watchfulness of his actions. Revenge was dearer than self-preservation…The cause of all these evils was the love of power, originating in avarice and ambition, and the party-spirit which is engendered by them when men are fairly embarked in a contest …. An attitude of perfidious antagonism everywhere prevailed; for there was no word binding enough nor oath terrible enough to reconcile enemies.[26]

Many of the observations Thucydides makes about vice signaling correspond to phenomena pessimistically predicted by Tosi and Warmke about moral grandstanding (virtue signaling), of which vice signaling is typically a special case. Much of the passage claims that Hellenes attempted to one-up each other on savagery toward enemies. Similarly, Tosi and Warmke predict "ramping up," where the signaling value of strong moral claims results in a "moral arms race" in which each individual attempts to demonstrate their commitment to justice by making a claim more extreme than the last individual.[27]

Tosi and Warmke note that people want to avoid being seen as cautious or cowardly by members of the in-group. Thucydides, similarly, comments that "[r]eckless daring was held to be loyal courage; prudent delay was the excuse of a coward; moderation was the disguise of unmanly weakness."[28] Tosi and Warmke predict that "excessive outrage" will result from moral grandstanding, where some will exploit the mistaken tendency to judge those with the most outrage about an issue to be the most morally reliable and upstanding people, either with respect to that issue or generally. Thucydides: "Frantic energy was the true quality of a man."[29]

One important difference, however, between Thucydides' analysis and the one offered by Tosi and Warmke is the level of generality for their claims. Tosi and Warmke focus their attention primarily on the effects of virtue signaling on discourse, perhaps corresponding to a strong distinction between discourse and acts in general. But on the view of things advanced in section 1, communication is something that acts can do in general. Language or discourse concerns the sort of action where communication is usually the point, but does not nearly exhaust the domain of action where communicative effects are salient. This thought is at home in Neil Levy's recent rebuttal to Tosi and Warmke, in which Levy points out that "public moral discourse" serves many social functions, thus doing more than just providing a forum for rational deliberation on moral matters (the singular role assigned

---

[26] Thucydides, *The Peloponnesian War*, book III, as quoted in Robertson, *Patriotism and Empire*, 93–94.
[27] Tosi and Warmke, "Moral Grandstanding," 205.
[28] Robertson, *Patriotism and Empire*, 93–94. Supplemented with lines added from Thucydides, *The Peloponnesian War*.
[29] Thucydides, *The Peloponnesian War*, book III, as quoted in Robertson, *Patriotism and Empire*, 93–94.

to public moral discourse by Tosi and Warmke).[30] Thucydides' account provides a telling real-world example of Levy's objection, on the safe assumption that the "plots" and "crimes" he refers to were not merely verbal dressings-down or pronouncements in the town square.

That is: we can and should ask quite generally what the behavioral consequences of both virtue and vice signaling will be. We would then follow Thucydides in investigating social life beyond speech acts or discourse. If the previous section is onto something, then virtue and vice signaling adjust incentive structures not simply for essentially communicative acts but for all acts that communicate, at least where the communicative effects are salient for the overall payoff of the act or otherwise taken into account by actors. Denigrating speech acts communicate insult, but rolled eyes, slaps to the face, revenge plots, and ignored invitations do as well. Then, our phenomena of interest will include speech acts, but it will also include many other sorts of actions.

The discussion of the pros and cons of vice signaling in this section will presume this level of generality to the insights about moral grandstanding discussed so far. I take it that vice signaling has many of the same potential benefits and upshots that virtue signaling or grandstanding have generally, as Levy's article explains: vice signaling can express genuinely held moral commitments and contribute to public discussion.[31] But it is nevertheless worth mentioning two benefits that are especially salient for the vice-signaling subset of virtue-signaling actions.

3. Potential Benefits of Vice Signaling

3.1. Vice Signaling Can Serve as a Basis for Solidarity

The example of etiquette in Southern Rhodesia both provides an example of non-speech acts that communicate and signal in the relevant sense, as well as demonstrating some potential benefits of vice signaling as a practice.

Nathan Shamuyarira was a high-ranking member of Zimbabwe's African National Union—Patriotic Front (ZANU-PF, the party of Robert Mugabe, the country's first prime minister and longtime president). Before taking this role, he was a key member of its nationalist struggle against colonial domination while the country was still known as "Southern Rhodesia." In his historical and autobiographical book *Crisis in Rhodesia*, Shamuyarira recounts not only that nationalist leaders deliberately flaunted the prevailing norms of etiquette, wearing hats in the presence of white officials, but that their willingness to do so became a marker of political credibility.[32]

It is not hard to see the wisdom of this. To follow the prescription of (then) Southern Rhodesia that "natives" (Black Africans) were not to wear hats in the presence of white people was to govern one's self by the moral expressive norms of an apartheid regime. Thus, it was not simply the case that each Black person had intrinsic reason to ignore the norm, part and parcel of a racist and oppressive social structure as it was. It was also the case that each person had reason to broadcast their willingness to defect from such norms, and thereby build social awareness that people were willing to stand up to apartheid in at least this small sense. That sense could, and did, build into a larger and more influential form of resistance, culminating in the successful Zimbabwean War of Liberation.

Shoemaker and Vargas call this signaling role "moral torch fishing" in the case of blame, arguing that signaling one's adherence to moral norms and willingness to enforce adherence in others is an important moral function that helps social systems cement stable cooperation over time.[33] Similarly,

[30] Levy, "Virtue Signalling Is Virtuous."
[31] Levy "Virtue Signalling Is Virtuous."
[32] Shamuyarira, *Crisis in Rhodesia*.
[33] Shoemaker and Vargas, "Moral Torch Fishing."

Neil Levy points out that the strong feelings involved in acts of virtue signaling—and thus, as this paper has argued, of many cases of vice signaling—are *constitutive* of possession of the moral virtues they exemplify.[34] In the case of Southern Rhodesia, this kind of anti-apartheid signaling proved efficacious (or at the very least, a survivable mistake), as it played a part in a successful revolt against colonial rule. A pro-solidarity effect of moral grandstanding is consistent with Tosi and Warmke's follow-up empirical investigations, which suggested a positive relationship between moral grandstanding and the tendency to grow closer to people of similar moral and political beliefs.[35]

3.2. Vice Signaling Can Restructure Social Relationships

Vice signaling can help publicize and cement opposition to the status quo, and thereby help restructure society by means of subsequent organized political action. This was the story in the previous example of the Zimbabwean War of Liberation. But vice-signaling communicative acts can directly challenge social relationships and thus relations of power and domination.

Social structure consists of both formal and informal elements. Formal elements, like laws and institutions, are easy to recognize and to specify pathways for changing. But informal elements like norms of civility and etiquette are also influential aspects of social structure. Philosopher Chenyang Li goes as far as to suggest that these aspects of social structure are partially constitutive of individual behavior, as the "cultural grammar" that decides whether some individual's behavioral "sentences" are well formed—that is, whether they succeed or fail by the lights of the going interpretive and evaluative norms.[36] Deliberate flaunting of the going norms can call them into question and provoke a wide reconsideration of those norms.

Historian Robin D. G. Kelley and sociologist James C. Scott describe the cultural importance of this kind of broadcasting to various marginalized groups of people, including working class African Americans, and South Asian peasant populations.[37] They credit it with preserving collective self-respect, cultural opposition to injustice, and persistent material challenge to oppressive power relations.[38] Li's "cultural grammar" view helps make sense of the last claim. If norms of civility and social conduct are an aspect of social structure, and vice signalers flaunt this aspect of social structure in a way that can provoke reconsideration of the attendant norms, it follows that vice signalers can provoke a reconstitution of social structure itself.

4. Potential Drawbacks of Vice Signaling

Though vice signaling has similar benefits to virtue signaling and other grandstanding acts, its differences and unique dangers show up when considering two interrelated drawbacks.

4.1. Vice Signaling Changes the Subject

---

[34] Levy, "Virtue Signalling Is Virtuous."

[35] See study 5. Grubbs et al., "Moral Grandstanding in Public Discourse," 16.

[36] Li, "Li as Cultural Grammar."

[37] Kelley and Scott often emphasize the cases of vice signaling that are inscrutable to the socially dominant groups ("hidden transcript"), but this is not a necessary aspect of vice signaling. Moreover, as social media changes the incentive structures of public communication, I would guess that the hiddenness of the opposition of marginalized groups will decline in political significance. Kelley, "'We Are Not What We Seem'"; Scott, Domination and the Arts of Resistance."

[38] Kelley, "'We Are Not What We Seem,'" 78.

Andrea Long Chu provides a telling example of how vice signaling changes the subject. In "On Liking Women" she comments on political lesbianism, a movement that advocated for a connection between same-gender relationships between women and the fight against the patriarchy. She writes: "I take to be the true lesson of political lesbianism as a failed project: that nothing good comes of forcing desire to conform to political principle …. Perhaps my consciousness needs raising. I muster a shrug. When the airline loses your luggage, you are not making a principled political statement about the tyranny of private property; you just want your goddamn luggage back."[39] Her point, as I understand it, is that the demands of this wave of the radical feminist movement for signaling one's commitment to women's liberation in one's personal relationships problematically dominated other reasons and motivations that would otherwise guide members' choices in romantic and sexual partnerships.

I agree with Tosi and Warmke that it is perhaps additionally morally problematic for individuals to use public moral discourse toward their own individual ends. But the effects on the group dynamics as a whole are my primary concern. Vice signaling can fundamentally change what is being pursued by the group, above and beyond its effects on individual conversations.

One way that vice signaling can change the subject operates through the relationship it can establish between the in-group and out-group. Generally, vice signalers are in constant contact with their group's own moral commitments. These, after all, will decide whether their performance in public space or contribution to public moral discourse succeeds or fails at instantiating virtue as the group defines it. Vice signaling, on the other hand, puts the in-group in a relationship of epistemic dependence to the out-group. For the vice signaler to successfully vice signal, it is the *out-group's* thoughts, moral compass, and evaluative norms that serve as the primarily relevant factors for vice signaling, not the in-group's.

One may object that I have overstated the case here, since I have left out discussion of what role the in-group's moral commitments play.[40] But, if the in-group's moral commitments are relevant at all to these acts—and it is not obvious that they are—they likely factor as a constraint on which violations of out-group morality will be tolerated. But this fact, even if true in the short term, is little consolation. Consider the following conjectures. First, that the higher the level of antagonism between in-group and out-group, the lower the extent to which in-group moral commitments will constrain vice-signaling acts, since inflaming the out-group is more valued when they are more hated. Second, that acts of vice signaling are likely to help create more antagonism between groups, as they involve deliberately inflaming the out-group and then celebrating this fact. Both of these, together, imply that the effective constraint of in-group morality on acts of vice signaling *weakens* as more vice-signaling acts occur. There are then two related dangers: that in-group moral commitments are not an initially effective constraint on vice-signaling acts and that, however effective they might be when vice signaling is rare, they will become increasingly irrelevant as vice signaling proliferates.

The Son of Baldwin case provides a tidy illustration of this possibility. Vice signaling sidelines the in-group's conception of virtue, treating "fuck Justine Damond" as a virtuous expression of righteous Black anger, pearl-clutching white moderates be damned. But vice-signaling acts and the culture built around them thereby treat speech acts like "fuck Justine Damond" as an instance of a general virtuous kind of action—as an "expression of righteous anger"—obscuring the moral evaluation of the specific token act that it is, which is an insult to a homicide victim. Son of Baldwin does not even attempt to argue that Justine Damond herself did anything to merit being spoken about like this, or otherwise justify the specific thing being said. Rather, the expressive act justifies itself by reference to the hated racist political context and its out-group defenders, directing social attention away from the content of what

[39] Chu, "On Liking Women."

[40] I am indebted to an anonymous reviewer for the importance of this point.

was said and to the people that it involves, except insofar as they can be instrumentalized to express the speaker's and audience's well-deserved anger.

The possibility of the initial or gradual irrelevance of in-group moral commitments is especially hard to square with a version of social justice where the marginalized in-group wants freedom and self-determination. If this strategy is supposed to be how the in-group escapes the influence of the out-group, this result could hardly be worse. It requires in-group members to make constant reference to what the out-group thinks and believes, even though they aim to play contrarian. Groups that vice signal too often and for too long risk forgetting who they are culturally, ideologically, and politically as they subordinate themselves to antagonism for its own sake—and, in so doing, subordinate themselves to the very out-group they may have aimed to liberate themselves from.

A second way that vice signaling can change the subject is by directly affecting the basic character of social interactions around the topic groups are squaring off against each other over. On social media, our speech acts have quantified, measurable reactions from the audience: likes, replies, and retweets. C. Thi Nguyen argues that this can have structuring effects on our agency much like the rules and point systems of games, which structure our behavior by making the full range of practical possibilities quantitatively commensurable and thus making some decisions more "valuable" (often measured in points) than other decisions.[41] This produces "value clarity," an artificially simplified decision-making environment, which is pleasurable in and of itself and a key aspect of the fun of many kinds of games.

When social interaction around real-world issues is gamified in this way, social life is distorted. Nguyen and Bekka Williams use the term "moral outrage porn" to describe one way that discourse can shift people's antecedent relationship to their moral values. They define moral outrage porn as "representations of moral outrage engaged with primarily for the sake of the resulting gratification, freed from the usual costs and consequences of engaging with morally outrageous content."[42] The value clarity provided by Twitter as a platform, when combined with a culture permissive of internet vice signaling, might change how people interact with issues online and offline.

4.2. Vice Signaling Can Undermine In-Group Goals

The changes vice signaling makes to social interactions can have serious, long-term consequences on in-group's political interests.[43] Today, vice signaling changes the subject of discussion in public moral discourse. But this same drawback, considered on a different timescale, could have even deeper consequences: a month, year, or decade from now, vice signaling could change the practical orientation of a whole group of people or the course of a political project.

Take, for example, a progression of values and decisions we could make as organized opponents of mass incarceration. When we first start engaging online about the issue, we are clearly focused on destroying the current carceral system. We view social media instrumentally: we aim to intervene in online public moral discourse to win converts to our cause and proliferate better strategies among those who currently agree with our goals. Over time, our behavior changes, given the susceptibility of our organizing culture to the gamifying effects of social media platforms. Rather than tweeting and organizing about mass incarceration to figure out how to close jails and prisons, we begin tweeting to excite fellow abolitionists and inflame defenders of the carceral status quo and even make organizing

[41] Nguyen, *Games*, ch. 9.

[42] Nguyen and Williams, "Moral Outrage Porn."

[43] A small but growing body of empirical evidence suggests that there may be positive feedback between number of participants in signaling kinds of moral discourse at a given time and subsequent recruitment of people into similar kinds of moral discourse. Johnen, Jungblut, and Ziegele, "The Digital Outcry"; Pfeffer, Zorbach, and Carley, "Understanding Online Firestorms."

decisions for the same reasons. The simpler, social media–inflected version of our values replaces our original values and concerns: we measure how well we are doing by likes and retweets, not by the population of incarcerated people or the closures of jails and prisons.

This subtle shift in goals is what Nguyen calls "value capture": a gradual reorganization of one's goals and values, where things that were initially secondary or even tertiary goals climb the preference-ordering ranks and function as primary goals.[44] Our moral beliefs, the communities we were originally fighting for, and the events we are trying to bring about or prevent can all become instrumental servants to the symbolism of social interactions if signaling behavior goes unchecked. In the case just offered, the instrumental relationship of social media to concrete political goals is entirely reversed by the end of the process. The importance of the fates and lives of the people currently and at risk of being incarcerated falls by the wayside in favor of the group's new selfish and masturbatory ends: they figure in insofar as they enable us to declare victory online, to the extent that they are relevant at all.[45]

Pervasive vice signaling presents dangers, then, because of its long-term political effects: namely, that it might alter the incentive structures of patterns of discourse, political strategy, and behavior in general around the pursuit of ends that are less important or less coherent with our initial values than the ones we would pursue without them. Vice signaling risks a perverse trade between the communicative performance of taking sides in a political contest and the actions that could lead to winning the contest.

The previous point explains how vice signaling could harm political goals through its effect on our attention, and how antagonism can distract us from trying to make actual progress on changing the social world in the way our group wants. Another way vice signaling could undermine political goals is in the way it distorts deliberation about our group's political issues: that is, how we think about our political goals when we *are* paying attention to them.

If in-group members cannot express or act on ideas that smack of agreement or sympathy with the out-group, this might distort group deliberation that otherwise might have converged on some true or effective outlook. Similarly, an idea that would be rejected if evaluated on independent grounds might instead be embraced because it seems combative or militant, its effectiveness or principledness aside. These possibilities present strategic problems for social movements because the epistemic distortions affect the group's understanding of aspects of the world and the political context that are key to the group's success in political campaigns. This corresponds to Thucydides' observed response to vice signaling in Hellas: the "meaning of words no longer had the same relation to things, but were changed by them as thought proper."[46]

Sustained patterns of vice signaling can lead to the kind of conflict for conflict's sake that Thucydides describes, which is a likely result of the "ramping up" and "trumping up" that Tosi and Warmke consider in their discussion of moral grandstanding, that Arvan links to group polarization, and that relate to the short-sightedness diagnosed by Nguyen and Williams' discussion of moral outrage porn.[47] Tosi and Warmke's prediction about moral grandstanding applies just as well to vice signaling: it might generate an arms race to decide who is the most antagonistic to the mutually hated out-group (marginalizing the least antagonistic folks). It also functions as a way for to jockey for higher positions

---

[44] Nguyen, *Games*, ch. 9.

[45] Nguyen and Williams also point out the pleasure in consuming content that fits a person's moral perspective. I focus on the social aspects of moral outrage porn here for the sake of drawing out the political significance of changing the subject, but self-pleasure is yet another sense in which moral outrage porn and virtue signaling could "change the subject" ("Moral Outrage Porn," 23–26).

[46] Thucydides, [add citation].

[47] Arvan, "The Dark Side of Morality," 99; Nguyen and Williams, "Moral Outrage Porn"; Tosi and Warmke, *Grandstanding*, 51–57.

within the in-group hierarchy, threatening to supplant solidarity based on a group's positive goals with a perverse solidarity based on mutual hatred of an out-group or out-groups, bearing no necessary relationship to a positive set of moral and political commitments.

Thucydides chronicled ramping-up effects in his history: "He who plotted from the first to have nothing to do with plots was a breaker-up of parties and a poltroon who was afraid of the enemy. In a word, he who could outstrip another in a bad action was applauded; and so was he who encouraged to evil one who had no idea of it."[48] The danger is that maintaining solidarity in an atmosphere where vice signaling reigns will require yet more vice-signaling acts, generating a perverse feedback loop of pointlessly antagonistic actions that might erode the very social institutions that would be needed to address the grievances that kicked off the process in the first place.

All the effort put into resolving the in-group and between-group crises and battles could have been spent on positive projects: reviewing and working toward the in-group's positive commitments. The necessary behaviors for these positive projects (conversations, research tasks, organizing childcare and carpools) risk being distorted or crowded out entirely by the incentive structure that vice signaling often exploits, cements, and propagates.

Finally, it follows from the preceding that patterns of vice signaling also risk undermining the in-group morally. What makes some out-groups worth opposing is their coherence around fundamentally unjust group goals and practices. But the injustice of the dominant out-group does not by itself make the in-group worth joining: if prisons should not exist, then fighting to abolish prisons is a just struggle. But the struggle *against the people who support prisons* bears no such inherent relationship to justice, and is compatible with prisons' continued existence. If the in-group does not organize itself and cohere around just goals and practices—perhaps better yet, the pursuit of justice itself—then it risks cultivating a purely cosmetic relationship to justice.

5. Conclusion

In the preceding, I have primarily discussed the possible results of sustained patterns of vice signaling. Both my criticisms and hopes for vice signaling are primarily strategic or tactical. The goodness or badness of instances of vice signaling depends importantly on the moral status of the political project to which they contribute or fail to contribute. But even conceding this much, vice signaling seems to represent an especially intense form of the risks that have been associated with moral grandstanding.[49] In particular, the way that vice signaling incentivizes the irrelevance of one's own in-group moral commitments seems to pose a much more fundamental risk to public morality than other kinds of grandstanding—perhaps it is no coincidence that Thucydides' discussion of vice signaling is a description of social collapse and endemic conflict.

Interdisciplinary research can help identify the short-, medium-, and long-term risks of vice signaling. Tosi and Warmke are onto something by beginning to study grandstanders empirically, but an investigation of the psychology or goals of individual people who vice signal is of limited value. If the analysis offered in this paper is right, then the basic social dynamics that explain vice signaling are group level and inter-group. Future research should ask fewer questions about what grandstanders are after or whether or not they are hypocrites—these criticisms and preoccupations themselves risk participating in the erosion of the public moral discourse they purport to defend in a manner much like vice signaling itself does, to the extent that they change the subject to whether or not individuals have the standing or

[48] Thucydides, *The Peloponnesian War*, book III, as quoted in Robertson, *Patriotism and Empire*, 93–94.

[49] I am grateful to an anonymous reviewer for encouraging me to rethink this point.

conviction to properly express emotions like outrage and away from the circumstances being responded to.

Instead, future research should shed light on how patterns of communication between networks of people manifest in group-level psychological differences (e.g., a group's "affective tone") and patterns of social and political behavior, including political organizing and electoral participation.[50] Psychologists, sociologists, economists, and political scientists would all have much to contribute to a project of this kind.

There is, however, an ethical conviction motivating the arguments that I have pursued here. I believe that the battle for justice will only be won by defeating the current system of injustice if its replacement is just, and we will not figure out what that looks like just by opposing enough specific elements of the status quo, whether its political factions or its values. More importantly, we will not *be* what that replacement looks like merely by way of opposition, and we will not build what that replacement looks like through pure opposition.[51]

Georgetown University
Olufemi.Taiwo@georgetown.edu

References

Abbott, Barbara. "Presuppositions and Common Ground." *Linguistics and Philosophy* 31, no. 5 (2008): 523–38.

Arvan, Marcus. "The Dark Side of Morality: Group Polarization and Moral Epistemology." *Philosophical Forum* 50, no. 1 (2019): 87–115.

Bartholomew, James. "The Awful Rise of 'Virtue Signalling.'" *The Spectator*, July 7, 2018. https://www.spectator.co.uk/2015/04/hating-the-daily-mail-is-a-substitute-for-doing-good.

———. "I Invented 'Virtue Signalling.' Now It's Taking over the World." *The Spectator*, October 10, 2015. https://www.spectator.co.uk/2015/10/i-invented-virtue-signalling-now-its-taking-over-the-world.

Chu, Andrea Long. "On Liking Women." *N+1*, November 29, 2017. https://nplusonemag.com/issue-30/essays/on-liking-women.

Flaherty, Colleen "Trinity Suspends Targeted Professor." *Inside Higher Ed*, June 27, 2017. https://www.insidehighered.com/news/2017/06/27/trinity-college-connecticut-puts-johnny-eric-williams-leave-over-controversial.

George, Jennifer M. "Personality, Affect, and Behavior in Groups." *Journal of Applied Psychology* 75, no. 2 (1990): 107–16.

Goyette, Jared. "Justine Damond: Video Shows Australian Rescuing Ducklings Near Minneapolis Home." *The Guardian*, July 19, 2017. http://www.theguardian.com/us-news/2017/jul/19/justine-damond-video-shows-australian-rescuing-ducklings-near-minneapolis-home.

Grubbs, Joshua B, Brandon Warmke, Justin Tosi, A. Shanti James, and W. Keith Campbell. "Moral Grandstanding in Public Discourse: Status-Seeking Motives as a Potential Explanatory Mechanism in Predicting Conflict." *PLOS ONE* 14, no. 10 (October 2019).

[50] George, "Personality, Affect, and Behavior in Groups," 107.

[51] Thanks to Meena Krishnamurthy, Liam Kofi Bright, Joel Michael Reynolds, Abigail Higgins, and Shelbi Nahwilet Meissner for their support and comments during the writing of this article.

Johnen, Marius, Marc Jungblut, and Marc Ziegele. "The Digital Outcry: What Incites Participation Behavior in an Online Firestorm?" *New Media & Society* 20, no. 9 (November 29, 2017): 3140–60. https://doi.org/10.1177/1461444817741883.

Kelley, Robin D. G. "'We Are Not What We Seem': Rethinking Black Working-Class Opposition in the Jim Crow South." *Journal of American History* 80, no. 1 (June 1993) 75–112.

Krishnamurthy, Meena. "Featured Philosopher: Liam Kofi Bright." *Philosopher*, January 21, 2017. https://politicalphilosopher.net/2017/01/20/featured-philosopher-liam-kofi-bright.

Levy, Neil. "Virtue Signalling Is Virtuous." *Synthese* 198 (2021): 9545–62.

Li, Chenyang. "Li as Cultural Grammar: On the Relation between Li and Ren in Confucius' Analects." *Philosophy East and West* 57, no. 3 (2007): 311–29.

Malinsky, Daniel. "Intervening on Structure." *Synthese* 195 (2017): 2295–2312.

Media Matters Staff. "CBS Report on Police Shooting of Alton Sterling Inappropriately Highlights Victim's Record." *Media Matters for America*, July 7, 2016. https://www.mediamatters.org/video/2016/07/07/cbs-report-police-shooting-alton-sterling-inappropriately-highlights-victims-record/211411.

Nguyen, C. Thi. *Games: Agency as Art*. Oxford: Oxford University Press, 2018.

Nguyen, C. Thi, and Bekka Williams. "Moral Outrage Porn." *Journal of Ethics and Social Philosophy* 28, no. 2 (August 2020): 147–72.

Perez, Chris. "Bride-to-Be Is 'Most Innocent' Police Shooting Victim." *New York Post*, July 21, 2017. http://nypost.com/2017/07/21/bride-to-be-is-most-innocent-police-shooting-victim-lawyer.

Pfeffer, Juergen, T. Zorbach, and Kathleen M. Carley. "Understanding Online Firestorms: Negative Word-of-Mouth Dynamics in Social Media Networks." *Journal of Marketing Communications* 20, nos. 1–2 (March 4, 2014): 117–28. https://doi.org/10.1080/13527266.2013.797778.

Potts, Christopher. "Presupposition and Implicature." *The Handbook of Contemporary Semantic Theory*, 2nd ed. Oxford: Wiley-Blackwell, 2013.

Robertson, John Mackinnon. *Patriotism and Empire*. London: Grant Richards, 1899. https://archive.org/details/cu31924021032366.

Scott, James C. *Domination and the Arts of Resistance: Hidden Transcripts*. New Haven, CT: Yale University Press, 1990.

Shamuyarira, Nathan M. *Crisis in Rhodesia*. New York: Transatlantic Arts, 1966.

Shoemaker, David, and Manuel Vargas. "Moral Torch Fishing: A Signaling Theory of Blame." *Noûs* 55, no. 3 (2017): 581–602.

Son of Baldwin. "Let Them Fucking Die." Medium, July 23, 2017. https://medium.com/@SonofBaldwin/let-them-fucking-die-c316eee34212.

Stalnaker, Robert. "Common Ground." *Linguistics and Philosophy* 25, no. 5 (2002): 701–21.

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

Starr, Terrell Jermaine. "I Understand Why Some Black People Couldn't Care Less about Justine Damond." *The Root*. https://www.theroot.com/i-understand-why-some-black-people-couldn-t-care-less-a-1797189837.

Stinson, Philip M. "Police Shootings Data: What We Know and What We Don't Know." Urban Elected Prosecutors Summit, Atlanta, GA. April 20, 2017.

Sullivan, John, Liz Weber, Julie Tate, and Jennifer Jenkins. "Four Years in a Row, Police Nationwide Fatally Shoot Nearly 1,000 People." *Washington Post*, February 7, 2019. https://www.washingtonpost.com/investigations/four-years-in-a-row-police-nationwide-fatally-shoot-nearly-1000-people/2019/02/07/0cb3b098-020f-11e9-9122-82e98f91ee6f_story.html.

Táíwò, Olúfémi O. "The empire has no clothes." *Disputatio* 10, no. 51 (2018): 305-330.Thucydides. *The Peloponnesian War*. Translated by Benjamin Jowett. New York, E. P. Dutton. 1910. http://www.perseus.tufts.edu/hopper/text?doc=Perseus%3Atext%3A1999.01.0200.

Tosi, Justin, and Brandon Warmke. *Grandstanding: The Use and Abuse of Moral Talk*. New York: Oxford University Press, 2020.

———. "Moral Grandstanding." *Philosophy & Public Affairs* 44, no. 3 (2016): 197–217.

Warmke, Brandon, and Justin Tosi. "Moral Grandstanding and Virtue Signaling: The Same Thing?" *Psychology Today*, August 11, 2020. https://www.psychologytoday.com/blog/moral-talk/202008/moral-grandstanding-and-virtue-signaling-the-same-thing.

Weinberg, Justin. "A Surprising Instance of Performative Philosophy." *Daily Nous*, n.d. http://dailynous.com/surprising-instance-performative-philosophy.