

Thank you for this feedback. In responding, I note that the suggestions are presented as optional (R3) and only a suggestion (R2)/food for thought (R1).

Reviewer 3

Comment: This was a well-argued, compelling paper that brought Kierkegaard's views to bear in an interesting, illuminating way on an important topic. I enjoyed reading the paper, and I think that it is fit for publication.

Response: Thank you for this judgement.

Comment: I do have several questions and comments, though, which I think the paper could benefit from addressing. (Though to be clear, I don't think these comments need to be addressed for the paper to be published.)

One question I was left with—especially in section 2—is whether the dispute between critics and defenders of virtue-signaling is, in large part, *terminological*. The third and fourth kinds of moral grandstanding that Tosi and Warmke identify—Trumping Up and Excessive Outrage—are defined such that the badness/viciousness/inaptness of grandstanding is *part of the very concept*. As you note on page 7, defenders of virtue-signaling (like Levy and Westra) resist such attempts to define virtue-signaling in such a way that it's *by definition* vicious. You respond that virtue signaling is a species of norm-signaling with “recognition desire” as its differentium. I think this would be worth foregrounding more clearly earlier in the paper.

Response: Good suggestion. I have done so at lines 117-8.

Comment: But I also think it would be worth explaining why Hill and Fanciullo think norm-signaling doesn't necessarily entail an attempt to gain recognition. On its face, it seems to me that *signaling* essentially involves communicating something to another party, where that communication aims to get *uptake* in which the other party *recognizes* what you've communicated. What distinguishes the specific kind of recognition virtue signalers desire from this generic recognition desire involved in any kind of signaling adherence to a norm?

Response: I think the key difference is that the “recognition desire” requires more than recognition in this minimal sense: rather, it seeks a kind of recognition for which I gain some *credit* that enhances in some way my *moral reputation*. To clarify, I have elaborated this at lines 271-2.

Comment: Similarly, I think it would be helpful to clarify whether you are defending the claim that virtue-signaling is always vicious—either as a matter of definition, or on substantive ethical grounds—or whether you think it's just *typically* vicious. To my mind, you make a strong case for the claim that virtue-signaling inherently has *tendencies* to corrupt individuals' character and have bad social effects, and you do an excellent job of diagnosing what these corrupting tendencies are. But I don't think it shows that these tendencies are *always* realized or manifested. That is, your argument leaves open that *some* genuinely virtuous agents engage in virtue-signaling while avoiding the pitfalls you identify.

Response: My claim is that it is *typically* vicious. I have aimed to make this clearer in both the abstract (line 19-20) and in the conclusion (line 1032). It is not part of my claim that the corrupting tendencies *always* manifest - though I would still object to virtue-signaling on the “vanity project” and “lack of character” grounds argued for in the paper, so I would not endorse the claim that “some virtuous agents engage in virtue-signaling”. I have added an extra footnote (24) to clarify.

Comment: This also bears on your critique of Westra on pp. 10-11. You say it “still seems a weak defence of the practice that it can lead to desirable outcomes provided the moral change is ‘positive’, when precisely the same process can spread negative moral changes.” But if Westra is just arguing for the claim that contrary to its critics, virtue-signaling is virtuous when performed in the service of promoting better social norms, I don't think that's a weak defense. Likewise, one way to defend the *epistemic* function of virtue-signaling is to say that it's epistemically valuable when it contributes to

a truth-conducive epistemic environment, but it's epistemically disvaluable when it's not. (Cf. Jennifer Lackey's defense of echo chambers.)

Such a defense of virtue-signaling would be especially natural if one accepts an externalist account of epistemic justification.

Response: See response above: since on my account virtue-signaling is not virtuously motivated, it is not itself virtuous even if it promotes admirable ends. The later comment (439-442) on Westra's consequentialist assumptions is also pertinent here.

Comment: I also think it may be worth pointing out some of the deeper points of disagreement between you/Kierkegaard and proponents of virtue-signaling. Kierkegaard thinks a) that striving to imitate others/conform to social norms is generally vicious, and b) that we should not care so much about social recognition, since both of these often hinder fulfilling our fundamental ethical task of becoming a single individual before God. But many philosophers (likely including some defenders of virtue-signaling) reject this view and think that conforming to social norms is, on the whole, a good, pro-social human disposition, and likewise think that human flourishing requires—or is at least significantly enhanced by—social recognition. (As Adam Smith famously said in the *Theory of Moral Sentiments*, we naturally desire *both* to be lovely and to be loved.) On this view, conforming with norms the “recognition-desire” of which Kierkegaard is critical can both be healthy and natural. While I don't think you need to spend more time arguing against these views, it may be helpful to more clearly diagnose what may be some of the underlying disagreements.

Response: Thank you for this suggestion. Since the paper aims to argue for the superiority of what I take to be Kierkegaard's view (by showing weaknesses in the alternative), I don't believe there would be room for this in what is already quite a long paper. But this could well be a useful suggestion for future work.

Comment: Just one comment regarding the structure: as the paper stands, the first half of the paper explains your own responses to Levy and Westra, and the second half explains how Kierkegaard provides the resources to critique virtue-signaling. I wonder whether the paper could be shortened and streamlined if you combine and integrate parts of each section.

Response: Thank you for this suggestion. In fact this structure is intentional: I aimed for the arguments against Levy and Westra to work on their own merits, before going on to show what additional resources Kierkegaard brings to the debate. That way, a reader unconvinced by Kierkegaard's assumptions may still be convinced by the arguments against Levy and Westra.