Synthese
https://doi.org/10.1007/s11229-020-02653-9

Virtue signalling is virtuous

Neil Levy[1,2]

Received: 22 September 2019 / Accepted: 3 April 2020

Abstract
The accusation of virtue signalling is typically understood as a serious charge. Those accused usually respond (if not by an admission of fault) by attempting to show that they are doing no such thing. In this paper, I argue that we ought to embrace the charge, rather than angrily reject it. I argue that this response can draw support from cognitive science, on the one hand, and from social epistemology on the other. I claim that we may appropriately concede that what we are doing is (inter alia) virtue signalling, because virtue signalling is morally appropriate. It neither expresses vices, nor is hypocritical, nor does it degrade the quality of public moral discourse. Signalling our commitment to norms is a central and justifiable function of moral discourse, and the same signals provide (higher-order) evidence that is appropriately taken into account in forming moral beliefs.

Keywords Virtue signalling · Cooperation · Cognitive science · Social epistemology

1 Introduction

Virtue signalling is not something any of us likes to be accused of. In this paper, I will argue that we ought to worry much less about the accusation. While some of the vices attributed to virtue signalling, and those who engage in it, are genuine, the fact that it gives rise to such problems is not a reason to abandon the practice. All practices, even the most valuable, have their risks and pathologies. Virtue signalling has its virtues, and these virtues typically outweigh its vices.

I shall develop the argument in dialogue with the only sustained consideration of virtue signalling in the philosophical literature: Tosi and Warmke's recent article on what they call *moral grandstanding* (Tosi and Warmke 2016). 'Moral grandstanding' seems to be identical to, or at any rate to overlap very considerably with, virtue sig-

✉ Neil Levy
neil.levy@philosophy.ox.ac.uk

1 Department of Philosophy, Macquarie University, Macquarie Park, NSW 2109, Australia

2 Uehiro Centre for Practical Ethics, University of Oxford, Oxford OX1 1PT, UK

Published online: 16 April 2020

nalling; since the latter term is more familiar (a column in the *Guardian* describes it as "out of control" (Shariatmadari 2016), I think it is better to use it than Tosi and Warmke's name for the phenomenon.[1] In what follows, I will first set out Tosi and Warmke's concerns about virtue signalling, before defending it against their accusations. As we'll see, their primary concern is that virtue signalling subverts a primary function of public moral discourse by substituting mechanisms of social comparison for exchange of reasons, and thereby making changes in opinion ir- or arational. I'll argue it does no such thing: rather, it provides genuine (higher-order) evidence to agents. Establishing this claim goes a long way by itself to vindicating virtue signalling (it is sufficient to vindicate if the benefits thereby reaped are greater than any costs that arise from it). But however things stand on this question, I'll argue, Tosi and Warmke are in any case wrong to think that signalling—independently of its argumentative function—is a perversion of the primary or justifying function of public moral discourse. Rather, signalling is *itself* a primary, and valuable, function of such discourse. Virtue signalling supports the deliberative function of moral discourse *and* it is valuable for its signalling role.

2 The vices of virtue signalling

What is virtue signalling? Tosi and Warmke define moral grandstanding as "making a contribution to moral discourse that aims to convince others that one is 'morally respectable'" (Tosi and Warmke 2016, p. 199). According to them, public moral discourse is justified by its role in improving the audience's moral beliefs and perhaps the world. But the virtue signaller is unduly concerned with *herself* rather than the issues she purports to discuss. While she may also aim to convince others and produce a better world, at least one of her primary motivations is *recognition*. She signals her supposed moral insight and her superior values, thereby turning moral discourse into a "vanity project".

What, precisely, is wrong with turning moral discourse into a vanity project? One worry might turn on hypocrisy. The virtue signaller claims to be concerned with——indeed, *outraged* by—injustice, but is instead (or to some important extent also) motivated by the desire to let others know how morally advanced she is. Often, this will be a motivation she herself is committed to condemning, insofar as she calls on us to

[1] In their recent empirical paper examining the personality traits associated with moral grandstanding, Tosi and Warmke (and co-authors) assert that moral grandstanding is preferable as a construct to virtue signalling on the grounds that the former is better defined. They also seem to redefine moral grandstanding in a way that makes it relatively more distant from virtue signalling. According to them, moral grandstanding is motivated "to a significant degree by a desire to enhance one's status or rank" (Grubbs et al. 2019). This seems a more demanding motivation that the motivation they had earlier taken to be at issue: signalling moral respectability. I can wish to be seen as respectable without aiming to enhance my prestige, let alone to establish dominance: I may simply seek to remind people that I am good *enough*. It is unsurprising, given this more demanding understanding of moral grandstanding and their concomitant operationalisation of it as a disposition to engage in conflictual interaction, that they find it associated with prestige and dominance-related personality traits. But given that virtue signalling need not seek prestige or dominance and need not lead to conflict, this paper is not illuminating for our purposes here. Note, moreover, that what matters in the end is not whether virtue signalling is or is not distinct from moral grandstanding, but whether the objections Tosi and Warmke level at the latter are (also) the best objections to the former.

place some alleged injustice at the centre of our concerns. She claims to be concerned about others and about justice, but reserves a great deal of concern for herself and our perception of her. Tosi and Warmke claim that this self-aggrandizing and narcissistic motivation deserves aretaic condemnation (Tosi and Warmke 2016, p. 215). Signalling virtue manifests vice.

They also claim that virtue signalling is pernicious due to its effects. They identify five problems that virtue signalling characteristically gives rise to:

1. *Piling on* The serial reiteration of a condemnation already made by earlier commentators is apt to occur as each person grasps the opportunity to signal they (too) belong on the right side.
2. *Ramping up* Rather than being recognized as (merely) on the right side, some or all of the virtue signallers may attempt to outdo earlier signallers by condemning more harshly, aiming thereby to be recognized as more morally serious and perceptive than others.
3. *Trumping up* Another way to signal one is more morally serious and perceptive than others is to detect a moral problem that others cannot. This may lead to virtue signallers claiming to see a moral problem where there is none.
4. *Excessive outrage* Signallers may attempt to demonstrate their moral seriousness by displaying a degree of anger well out of proportion to any actual offense.
5. *Claims of self-evidence*
6. Finally, moral perceptiveness may be signalled by an implicit analogising to sensory perceptiveness. One can just *see* that, and how, wrong an action or assertion is, thereby implying that those who lack this capacity are morally deficient in comparison.

Tosi and Warmke claim that these effects or accompaniments of virtue signalling have negative effects on the practice of public moral discourse. Virtue signalling, they claim, may induce moral cynicism, because those who engage it are not sincere in claiming to call attention to injustices. It thus causes a "devaluation of the social currency of moral talk" (Tosi and Warmke 2016, p. 210). Ramping up, trumping up and excessive outrage also devalue moral discourse: moral condemnation becomes cheap and moral language loses its force. Moreover, what should have been respectful debate that aims to uncover and explain the morally problematic features of states of affairs tends to have effects on people's behavior and beliefs that are not mediated by reasons at all. As people ramp up and pile on, group polarization (Sunstein 2002) may occur, with members of the group tending to shift toward more extreme viewpoints.

Virtue signalling almost certainly has some negative effects; those that Tosi and Warmke identify among them. It's not hard to think of instances of trumping up, involving (for instance) individuals who claim to detect racism or ableism in the most innocuous actions or cultural products, in order to draw attention to themselves or to proclaim their moral vanguardism. Piling on seems genuinely to occur, and may have the effect of alienating people whose offense is trivial and who might otherwise have easily acknowledged it and benefited from sensitive discussion. Claims of self-evidence may be used to shut down genuine debate, and so forth. But every practice has its pathologies. To assess whether a practice should be condemned as a whole, we also need to identify any benefits it may have, measure the relative weight of its costs

and benefits, and assess whether those benefits can be procured by some other, less costly, route. I will not embark on this full assessment of the merits of virtue signalling in this paper. Rather, I will advance some preliminary reasons to think that such an assessment will likely vindicate virtue signalling.

3 The argumentative function of public moral discourse

While there is no doubt that moral discourse has a number of functions, the function Tosi and Warmke highlight—changing minds about the moral properties of the world and thereby sometimes changing the world itself—is without doubt an important one. If Tosi and Warmke are right that virtue signalling interferes with the deliberative function of public moral discourse, this would be a very considerable mark against it. In this section, however, I will suggest that virtue signalling *supports* the deliberative function of public moral discourse.

Rational deliberation is deliberation that is appropriately responsive to evidence, as Tosi and Warmke emphasise. They object to virtue signalling because they claim it doesn't offer evidence. But they overlook *higher-order* evidence. Whereas first-order evidence bears directly on the truth of a proposition, higher-order evidence is evidence about the reliability of the processes that generate belief. For instance, the fact *that the consultant physician hasn't slept for 40 h* is higher-order evidence that her judgment that the patient is suffering from a rare disease is less reliable than her better slept colleague's dissenting judgment. Higher-order evidence is genuine evidence; evidence that we rationally take into account in deliberation. Virtue signalling provides higher-order evidence and thereby may play an important, and rational, role in deliberation.

Virtue signalling provides higher-order evidence by conveying *confidence* and the *numbers* of people who share a judgment. As a number of philosophers have emphasised, much—perhaps most—of what we know, we know in virtue of testimony (Coady 1992; Lackey and Sosa 2006). We are and should be responsive to a number of different features of testimony (its prior plausibility; how well placed the person is to know what they say; evidence that they might have ulterior motives, and so on). Among the features we ought to attend to is the confidence with which assertions are made. 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. Unsurprisingly, the empirical evidence shows that ordinary people use a *confidence heuristic* in assessing testimony (Price and Stone 2004; Pulford et al. 2018).

The number of people who offer testimony also provides higher-order evidence for (or against) the assertion made. The easiest way to see how numbers make this kind of difference is through a consideration of the epistemic significance of disagreement (on which there is a large and well-developed literature). Under a broad variety of conditions, disagreement with *one* epistemic peer constitutes strong pressure to conciliate: that is, to reduce our confidence in our beliefs. Consider *Restaurant Check* (based on Christensen 2007):

> 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.

Since both have (roughly equally) good track records with regard to mental arithmetic, and there is no reason to think that one is more likely to be mistaken on this occasion than the other, both should reduce their confidence in their judgment.[2]

With this in mind, let's move from the two person cases—where *A* disagrees with *B*—upon which the literature on the epistemic significance of disagreement has concentrated, to cases in which at least one of the parties is joined by other agents. An increase in numbers makes an epistemic difference in two ways. First, the larger the number of people who are known to disagree with me, the harder it is to dismiss them from peerhood by citing my better track record, intelligence, lack of bias, and so on, or by reference to what Lackey calls "personal information," such as my knowledge from the inside that I am sincere and attentive (Lackey 2010, p. 318). While it may be true that some of those who disagree with me can be dismissed on these kinds of grounds, the more numerous they are, the harder it is to dismiss them from peerhood (Levy 2019).

Second, sheer numbers make a direct difference to the significance of disagreement. Consider this variant of *Restaurant check*: I add up the bill and come to a different figure than Anika and Bindi, while they agree with one another. In a case like this, the fact that two independent individuals have come up with the same figure, whereas I'm in a minority (of one), entails that I am under more rational pressure to reduce my confidence than they are. 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).[3]

[2] The literature on peer disagreement often uses a highly idealized account of peerhood, such that we have few epistemic peers. As Lackey (2010) notes, this idealization threatens to cut the debate off from the real world cases which motivate investigation of the issue in the first place. In any case, the idealization is unnecessary. We routinely encounter cases in which disagreement puts pressure on us to conciliate. In actual cases like *Restaurant Check*, the agents should (and typically would) lower their confidence, without the need to investigate whether they are epistemic peers in the idealized sense supposed in the literature. If they are friends, they may know one another to be close enough to peerhood for the pressure to be significant. See Matheson (2015) for discussion of this point.

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

Cases like *Restaurant Check* also allow us to see that *agreement* is epistemically significant. On most nights, Anika and Bindi come to the same figure when they add up the bill. Their agreement should, and does, make each more confident in their mental math than they would have been had they added up the bill on their own. The harder the problem, and the greater the likelihood of error, the stronger the higher-order evidence constituted by agreement. Suppose that the bill itemized 117 different items, none of which were priced in whole dollars, and the gratuity calculated was 13%. If Anika and Bindi come to precisely the same (plausible) figure after doing the arithmetic independently in their heads, they should be very much confident in their sums than otherwise.

Now let's return to virtue signalling and to whether it tends to interfere with the deliberative function of public moral discourse. According to Tosi and Warmke, virtue signalling is epistemically objectionable. While it is capable of changing minds, the *mechanism* whereby belief change occurs through signalling is ir- or arational, and therefore unlikely to produce well-justified beliefs.[4] Rational deliberation occurs via the presentation of argument and evidence, and appropriate response to such evidence. Virtue signalling produces belief change through social comparison, they argue, and "social comparison is not truth-sensitive":

> By that we mean that what causes people to alter their views or stated positions is predominantly a desire to hold a prized place within the in-group. The relevant incentive, then, is not to cease modifying one's beliefs or stated positions once one arrives at the truth, but to stop once an even more extreme position would no longer impress one's in-group. Our objection, then, is not to radical or "extreme" views as such, but rather to the process by which group members arrive at them. That process does not reliably track truth, but rather something else (Coady et al. 2017).

Above I noted their claim that virtue signalling leads to group polarization: a tendency of groups to shift toward more extreme views. According to Tosi and Warmke, group polarization is (typically) irrational, and arises from social comparison. But it is far from obvious either that group polarization is irrational, or that it arises from social comparison (in any objectionable form).

Let's pause, first, to assess whether group polarization deserves condemnation (however it arises). There seems no a priori reason to think that the truth is more likely to lie in the middle of a group of deliberators, prior to their sharing their opinions with one another (and thereby providing first and higher-order evidence in favor of their views), than at the extremes. The only example Tosi and Warmke provide seems to make this point as well as any. Their example features a group of deliberators who, in the wake of a school shooting, are initially tentative in their support for stronger gun control but come to be more fervent through polarization. To me, that seems like group polarization serving the aims of truth. Obviously, in saying this I commit myself to a particular—controversial—normative claim. But there's no reason to think this

Footnote 3 continued
independence of individual informants from one another varies from case to case, but we can be confident that each filters testimony for plausibility to some degree.

[4] They argue for this claim at greatest length in a symposium on their paper at the blog *Pea Soup* (Coady et al. 2017).

normative claim should be rejected because it was initially held only by a minority of the deliberators. Everything depends on the composition of the group and the distribution of opinion within it. Extreme opinions about race and gender were more accurate than more moderate opinions in the antebellum United States, for instance.[5]

Of course, it might be the case that group polarization leads to more accurate beliefs in this or other cases only by chance. It might nevertheless be epistemically objectionable,[6] because it arises from social comparison and not argument and evidence. In fact, there is reason to think that rather than arising from an irrational process of social comparison, polarization arises from rational agents optimally taking into account both the confidence with which testimony is offered and the numbers of agents who share an opinion (Bordley 1983).

In cases of the kind Tosi and Warmke mention, the opinions of agents who are tentative in their support of gun control *should* be given less weight than others who are more confident. The expression of moral outrage is a particularly powerful cue to confidence. Rather than being an irrational influence on judgment, it works (at least in part) by providing higher-order evidence. If piling on—the serial repetition of claims—occurs, group members are provided with a guide to the numbers, and thereby further higher-order evidence. Their credences should change accordingly. If I tentatively think that p, my confidence that p ought to rise when I find out that many others also think that p, and rise still further if I discover that some are very confident that p (other things being equal, of course). Far from virtue signalling bypassing reasoning, it provides inputs into reasoning processes and leads to better justified beliefs.

That's not to say that group polarization may not cause individuals to move further from the truth. We may be subject to an *information cascade*. Such cascades sometimes occur when agents update their beliefs sequentially. In such cases, it may happen that agents disregard private evidence or prior probabilities because the evidence provided by the behavior of earlier agents outweighs their own. Consider this case (based on Anderson and Holt 1997): You are given the opportunity to draw a ball from one of two urns. Urn A contains white balls in a ratio of 3:1 to red; urn B the reverse. Your task is to identify which urn is which, by drawing a ball from one. Obviously, if you draw first, you should bet that the urn you draw from contains predominantly balls of the color you've drawn. Suppose, however, you draw after other agents. You don't

[5] In his classic guide to argument and fallacies, Thoules argues that any opinion at all can be represented as the mean between extremes, and identifies the tendency to think the truth lies in such moderation as an informal fallacy (Thouless 2011).

[6] Polarization might be problematic for non-epistemic reasons: say, because it makes political dialogue and compromise more difficult (in their response to comments at *Pea Soup* (Coady et al. 2017), Tosi and Warmke highlight potential non-epistemic costs of polarization). In support of their view, we might cite the widespread belief that social media has led to group polarization on both left and right. However, there is evidence that polarization has increased most among those who use the internet least (Boxell et al. 2017), which suggests that virtue signalling is not its cause. Moreover, while there has indeed been an increase in *affective* polarization, there has been no corresponding increase in polarization of policy positions (Iyengar et al. 2012; Mason 2018). That is, people dislike each other more than previously, but they are no further apart on the issues. Whether this is a serious problem for political dialogue, the extent to which it arises from signalling, and whether the problem is significant enough to outweigh the epistemic and coordination benefits of signalling are very difficult questions; questions I set aside (for lack of expertise, as much as space) here.

know what color ball they've drawn, but you do know how they've betted. Even if all agents are rational and all know that the others are rational, in certain circumstances an information cascade may occur that causes rising confidence in a falsehood. Suppose the first and the second agent both draw a red ball, and bet accordingly. Their behavior is rational, but their evidence may be misleading: they may have unluckily drawn red balls from the predominantly white urn. The agent choosing third in sequence may now rationally bet that the urn is predominantly red, *no matter what color she draws*, because the evidence stemming from the betting behavior of the earlier agents suggests the urn is predominantly red. From this point on, the total evidence available to each agent—their private evidence plus the evidence provided by the betting behavior of the previous drawers—favors red. As the sequence continues, the agents become increasingly confident that the urn is predominantly red, despite the fact that the private evidence favors white.

The possibility of such cascades seems to be a principle motivation for holding that rational agents should make their decisions independently of one another. Coady (2006) has pointed out that an independence requirement on decision-making is too strong: when some agents are more expert than others, we would lose important information were we to judge independently. But there is other information we would lose were we all to judge independently: higher-order evidence. While it is true that we risk information cascades in some cases, due to a failure to aggregate private evidence, epistemic vulnerability is just a sad fact of life. There is no failsafe way to firewall misleading evidence. We should follow the evidence where it leads, even knowing that sometimes it will lead us to falsehood.[7]

4 The purpose of public moral discourse

According to Tosi and Warmke, the "core primary function that justifies the practice" of public moral discourse is "to identify publicly certain moral features of a state of affairs, and sometimes additionally to explain the evaluation of that state or recommend some fitting response" (Tosi and Warmke 2016, pp. 209–210). They argue that because virtue signalling does not provide evidence, it cannot play that role. In the previous section, I've argued that this is false: that virtue signalling provides higher-order evidence and therefore can contribute to the (rational) evaluation of states of affairs. In this section, I set aside that claim, in order to assess whether that is indeed the "core primary function that justifies the practice" of moral discourse. I will suggest that public moral discourse has other primary functions, and these functions are supported by virtue signalling.[8]

As Tosi and Warmke themselves emphasise, it is unlikely that public moral discourse has any single function. In light of this fact, the claim that the function they

[7] Of course, a proper assessment of the rationality of sequential and non-independent decision-making, say on social media, requires a detailed formal model. The remarks I make here are no more than a sketch of the kinds of considerations such a model would have to take into account.

[8] Shoemaker and Vargas (forthcoming) also place virtue signalling in the context of signalling theory. Their aim is to develop a costly signalling account of blame, not virtue signalling; the latter is mentioned only as an aside. For them, virtue signalling is dishonest signalling. Virtue signalling is, for them, "aiming directly at the benefits [of signalling] by manipulating the signal" (13). As we'll see, virtue signalling need not, and typically does not, involve any manipulation: virtue signalling is honest signalling.

point to is the one that alone justifies it is implausible. It is more likely that more than one of its functions count in its favour and hence help to justify it. Among those functions, and playing an important role in justifying it, is the role it plays in solving coordination problems. Indeed, this role may be its single most important evolutionary function: the function that more than any other explains why we are in the business of making moral judgments at all.

We are a highly social species, and heavily dependent for our success on our capacity to share information and coordinate behavior. As such, we are at risk of being exploited by free-riders: conspecifics who attempt to reap the benefits of cooperation without paying the costs. We have evolved defences against free-riding. In hunter-gather societies, which are believed by many anthropologists to mirror to a significant extent the conditions to which we are most closely adapted, gossip about free-riders is an effective means of social control (Dunbar 1998). When gossip fails, harsher responses to persistent free-riders, such as ostracism or the imposition of punishments, are employed (Frey and Rusch 2012).

Gossip's role in solving coordination problems accords well with Tosi and Warmke's claim that the primary function of public moral discourse is improvement in beliefs or the world. By drawing your attention to *A*'s bad behavior, I bring you to have better moral beliefs about A and their character (and perhaps secondarily about the kinds of acts or omissions that constitute bad behavior), and I set the stage for improving our society by allowing us to place pressure on *A* together. However, as societies become more complex, the functions of moral discourse diversify, and signalling comes to play an increasingly central role.

While calling out bad behavior, and occasional escalation to the harsher responses that such calling out enables, may have been sufficient to stabilise cooperative norms in the small bands of the paleolithic era, in large groups, or in an environment in which individuals may easily move between groups, these responses are no longer sufficient. While you and I may now refuse to engage in cooperative enterprises with *A*, our gossiping may not reach the ears of others. *A* may be able to free-ride, secure in the knowledge that he will interact with a sufficient number of naïve agents to ensure that his behavior will not be punished effectively. The more complex the society, the harder it becomes to rely on reputation tracking to stabilize cooperative norms (Sterelny 2013).

One way to respond to these problems is by *signalling* that we are trustworthy. Animals use signals for a variety of purposes. 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). Good signals are hard to fake signals: if a signal is cheap, then defectors will co-opt it and it will rapidly lose its value. Stotting is a hard to fake signal because it is costly. The gazelle who can afford to waste energy it might have saved for fleeing is probably not worth chasing. The peacock's tail is an even more reliable signal, because the more spectacular the tail the more resources have been devoted to it and the better the health of the bird. A good signal of trustworthiness, too, will be hard to fake.

In human beings, hard to fake signals take a variety of forms. Some are costly, like the peacock's tail. 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), and particularly devout individuals may take vows of celibacy, of poverty or even enter small cells for life as anchorites. Some signals are not costly, but nevertheless are credibility enhancing (Henrich 2009). Crossing a bridge may not be costly for the person who crosses (she may benefit from doing so) but it is a reliable signal that she believes the bridge is safe.

We live in a world in which we cannot easily rely on others' moral record, as conveyed by gossip, to identify those we can trust. Our societies are too large for reputation-tracking to be reliable: gossip may not reach us, and agents move relatively freely from community to community. Formal systems of regulation may help, but their effective development and enforcement depends on a sufficient level of trust to avoid systematic corruption. Costly and credibility enhancing signalling help fill the gap between reputation tracking and formal regulation. For example, because religious observance involves hard to fake signals of trustworthiness, co-religionists may seek one another out as business partners. The role of Quakers in the early years of British industry is, for instance, well-known (Prior et al. 2006). Moreover, trust is not limited to co-religionists. Religious and non-religious people express more trust in religious people, regardless of their religion, than in atheists (Gervais et al. 2011, 2017).

Credibility enhancing displays and costly signals of religious commitment are moral signals (at least for those individuals who belong to the High Gods religions (Norenzayan 2013), with their moralized gods, which have a near monopoly on the faithful today). They are signals of willingness to abide by certain, publicly proclaimed, norms. They are ways of signalling our virtue. Displays of religiosity continue to play this signalling function today, especially in highly religious societies like the United States. But as societies secularise, such signals no longer have the same power. Small wonder we have turned to more secular virtue signalling.

At least some of the features Tosi and Warmke diagnose as typical of virtue signalling are features we ought to expect from signals that have the function of establishing one's bona fides as a trustworthy partner. Just as the faithful all join in public worship, with all singing, tithing or witnessing, so we all pile on in moral condemnation or—less often—praise (we pile on, moreover, in part to establish the boundaries of our group: our fellow believers, with whom we preferentially cooperate). Strong emotions are also predictable, given that emotions are hard to fake (Frank 1988); hence we see fervent religious devotion, on the one hand, and outraged moral condemnation, on the other. Claims of self-evidence may function to delineate the in-group, thereby serving the ends of signalling. Ramping up also has its religious analogues: think of Filipino self-flagellation or voluntary crucifixion at Easter, Shia self-flagellation during Muharram observances, or of the degradation of self that many Catholic saints engaged in. In these ways (and a myriad others, most much less dra-

matic: think of Christmas lights for example), believers compete to show how devout they are.[9]

Good signals are hard to fake, because they are costly, self-validating or involuntary. The peacock's tail is costly, while crossing a bridge to signal one's belief that it is safe is self-validating. 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. Of course, one may fear that in the contemporary environment (especially on social media, in which talk seems to be cheap and the hard to fake signs of emotion are not perceptible) virtue signalling is no longer hard to fake. This is an issue to which we will return.

Given that a central function of moral discourse is signalling commitment to norms, the claim that virtue signalling represents a perversion of the justifying function of such discourse is on very shaky ground. Virtue signalling is not merely a central function of public moral discourse; it is one that it plausibly *ought* to play. Delineating a group of reliable co-operators and signalling a willingness to abide by a publicly proclaimed moral code are surely aims worth pursuing.

5 Motivations for virtue signalling

In the previous section, I argued that signalling is not a perversion of the function of morality, but itself such a function. This claim does not, however, address a principal objection Tosi and Warmke level at virtue signalling. They suggest that it *devalues* public moral discourse, because it leads to cynicism in its audience. If moral discourse is signalling, "under the pretense of addressing injustice" (Tosi and Warmke 2016, p. 210), audiences who recognize this fact will become cynical. Virtue signallers will be seen to be hypocritical: they claim to be concerned with injustice, but are actually concerned with themselves. My claim that signalling is a function of moral discourse, not a perversion of its function, does nothing to allay this concern: it is the conflict between the content of the claim (*that's wrong!*) and its function (*I'm moral*) that gives rise to the worry, rather than any thought about the real function it is supposed to play.

Virtue signalling might also be hypocritical in another way. Not only might the virtue signaller really be concerned with signalling their moral respectability, they might also (or instead) be signalling *dishonestly*. The virtue signaller may fail to be virtuous. Is either accusation warranted? No doubt they sometimes are. However, the

[9] I leave aside 'trumping up'. No doubt it can be used as a signal of moral perceptiveness, but actual instances of alleged trumping up are likely to be controversial: the best explanation for why some people see some accusations as trumped sometimes just is their lack moral perceptiveness (or legitimate moral disagreement). Because almost all instances of alleged trumping up are controversial, I think we should refrain from concluding that it is common enough to count as a characteristic feature of virtue signalling.

comparison with signalling in the realm of religion should enable us to see that there's no reason to think either is generally true.

Consider, first, the claim that there is a mismatch between the content of the signal and its function that warrants an accusation of hypocrisy. The accusation is justified, it seems, only if the agent's motivation in making a moral claim is inconsistent with the content of their claim. There is no reason to think that that's the case with the virtue signaller. To see this, consider the parallel claim with regard to the signalling of religious commitment. As we've seen, many cognitive scientists of religion argue that a principal function of many religious rituals, practices and dress is such signalling. They do not suggest, however, that those who engage in such signalling do so *in order to* signal commitment. Evolutionary theory routinely distinguishes between *proximate* and *ultimate* explanations of behavior and other aspects of the phenotype. This distinction is crucial to responding to accusations of hypocrisy supposed to arise from the mismatch between content and function.

It was on analogous grounds that Michael Ghiselin famously argued that evolutionary theory entails that morality is shot through with hypocrisy (Ghiselin 1974). He suggested that because evolution can only reward selfish behavior—because behavior that benefitted other agents would be selected against—the pieties we mouth must be hypocritical. Of course, we engage in behavior that *seems* altruistic, aiding relatives and non-relatives alike, he conceded, but we are motivated to do so only because such behavior increases (inclusive) fitness. "Scratch an 'altruist', and watch a 'hypocrite' bleed" he wrote (Ghiselin 1974, p. 247).

But once we distinguish between the proximate and ultimate explanation for altruistic behavior, the accusation of hypocrisy loses its sting. It may be true that we are motivated to engage in altruistic acts because such actions are, on average, in our genetic interests, but this ultimate explanation does not entail, or even make plausible, the claim that we are motivated to engage in altruistic acts *in order to* increase the proportion of our genes represented in the next generation. The proximate mechanism is likely to be a genuine concern for others' well-being. Compare sex. The evolutionary explanation of sexual desire is obvious, but the content of desire isn't *to replicate our genes*. Notoriously, we are regularly motivated to engage in sexual acts when there is no chance of procreation, which is good evidence that the proximate mechanism is dissociated from the ultimate explanation.

Similarly, the person who wears religious garb, tithes or attends services may do so (in part at least) in order to signal commitment to a set of norms, but there is no reason to believe that this ultimate explanation figures in their proximate motivations. Indeed, there are good reasons to doubt that this signalling function is one that agents are usually aware of, since genuine commitment is likely to give rise to a more reliable signal than a merely instrumental commitment (Frank 1988). We therefore have good reason to think that those who profess religious belief or engage in religious behavior do so, in part, because they are sincere and take the behavior to be worthwhile.[10]

[10] Of course, people attend religious services for many reasons. For instance, there may be pressure from the community or family to attend, and a congregation may provide a social network or a source of emotional or even financial support (the relatively low level of spending on social welfare by the US has often been invoked to explain why it has resisted secularizing trends to a far greater extent than other wealthy and industrialized countries; see Gill and Lundsgaarde 2004). Agents are sometimes aware of these motivations. Nevertheless,

For exactly parallel reasons, we should expect that even if a major part of the explanation of why people engage in (particular instances of) public moral discourse is that such discourse can signal virtue (i.e. commitment to a set of norms or to in-group cooperation), in general people do not engage in public moral discourse *in order to* send these signals. No doubt some do, but the claim that this is routine or even common seems to be an expression of the same cynicism about morality expressed by Ghiselin, and equally unmotivated.

In fact, there is experimental evidence that people express moral outrage in order to signal virtue, but that the outrage they express is nevertheless real. In a recent study, Jordan and Rand asked participants to report their degree of anger toward an anonymous person who refused to share with another participant money they had been given in the course of the experiment. In some conditions, participants also had the opportunity to share money themselves (Jordan and Rand 2020). They found that the degree of condemnation of non-sharers diminished for those participants who had the opportunity to share themselves. The fact that those people who had this opportunity felt less outrage is evidence that expression of moral outrage has as one of its function the signalling of virtue (just as Tosi and Warmke suggest). Because the opportunity to share money provides a better opportunity to signal virtue, those who had this opportunity had a greatly diminished need to avail themselves of the outrage signal. But the entire experiment was anonymous: no one (not even the experimenters themselves) could identify the source of the signal. Jordan and Rand suggest that this shows that the outrage was genuinely felt, despite its signalling function. While the degree of outrage we feel is sensitive to the signalling function that the expression of outrage can play, we genuinely feel outrage to that degree. The ultimate explanation turns on signalling, but the proximate mechanism is the generation of outrage.[11]

Let's turn, now, to the second accusation: that those who signal virtue do not genuinely possess it. As we've seen, this accusation is dissociable from the first: whereas the first turns on the *motivations* for virtue signalling, the second turns on its *truthfulness*. This second accusation is surely sometimes warranted. The parallel phenomenon is common in biological evolution: when systems of signals evolve, they provide opportunities to organisms to mimic them. Consider *aposematism*: the use of signals to indicate that an organism is dangerous to others. For example, some animals and plants that are toxic to predators signal their toxicity through the use of colors that potential predators or grazers recognize as signals of toxicity (Harvey and Paxton 1981). This signal provides an opportunity for organisms that are not toxic: if they can mimic it, they can lower their risk of predation (Mappes and Alatalo 1997). If virtue signalling has benefits, it would not be surprising if this kind of mimicry developed, and some virtue signallers may be deceptive.

Footnote 10 continued
there is good reason to expect an intrinsic concern for religion and/or religious belief motivates a good many of those who participate in organised religion, even if it is also true that they would not be religious believers if it did not play some of these functions.

[11] Jordan and Rand suggest that we unconsciously calculate the signalling value of such expressions. It is important to be careful in assessing claims like this, however. Whether the best accounts of the mechanisms that make us sensitive to signalling value involves the attribution to us of an unconscious concern with signalling depends on tricky issues concerning the attribution of content to mechanisms. Right now, we don't know enough about the relevant mechanisms to know whether such an attribution is appropriate.

Considerations from evolutionary biology and game theory suggest that virtue signalling is unlikely to be dishonest, in the main, in the offline environment. First, the costs of false positives in this domain are high. Virtue signalling is supposed to be a solution to a coordination problem where the stakes are high: cooperating with a defector risks exploitation. For virtue signalling to be an effective solution to this coordination problem, false positives must be relatively uncommon. In the environment in which it evolved, therefore, we should expect such signals to be honest. Second, if the proportion of mimics surpasses a certain threshold (a threshold that will vary from case to case, depending on the costs and benefits of the signal), the signal will not be reliable enough to play its function and it will tend to fall into disuse. The fact that virtue signalling remains widespread (if it is a fact), is therefore some evidence that mimicry remains at a low frequency.

But online, for instance on social media (which might be thought to be the natural home of virtue signalling), the conditions are very different from those under which our disposition to virtue signal developed (by cultural evolution). We are disposed to signal virtue because such signals were adaptive in the large-scale societies that succeeded the small bands in which reputation tracking was sufficient to ensure cooperation (Norenzayan 2013), but these societies have been succeeded in turn by those characteristic of the post-industrial world in which we now live. Social media has substantially lowered the costs of virtue signalling, opening the way to mimicry, it might be thought. These developments are very recent; perhaps our dispositions simply haven't caught up with the changed incentives provided by our new environment.

But this worry cuts both ways. If it is true that we can expect a lag between changes in incentives and changes in our disposition to *respond* to signals, we should also expect a lag between changes in incentives and changes in our dispositions to *emit* signals. As we saw, the available evidence suggests that people feel genuine outrage in response to cues that are (or were) reliably associated with defection from moral norms. It will or would take time for the disposition to feel such outrage to respond to the new incentives, and therefore we can be reasonably confident that deceptive signalling of moral outrage remains relatively uncommon.

The speed with which any new opportunity for deceptive signalling will be exploited can be expected to depend, in important part, on how difficult it is to mimic the relevant signals. Social media makes virtue signals easier to fake because it is very much harder to observe the involuntary concomitants of genuine emotion, and because it is harder to monitor behavior across time and in different contexts online. But since the point of virtue signalling is establishing one's reputation, and that requires—at minimum—a stable name across time and ideally use of a real name (insofar as one seeks a good reputation for oneself, and not just an online avatar), deceptive signals remain at least somewhat difficult to fake. Stability of name raises the costs of online hypocrisy; use of real name raises the costs of hypocrisy across the board (especially to the extent to which one's network on social media includes people one interacts with "in real life"). Given that the opportunities for deceptive signalling have increased only recently, and

that the signals remain somewhat hard to fake, there seems little reason to believe that a very significant proportion of virtue signallers are deceptive, even on social media.[12]

But there's another, simpler, reason to think that virtue signalling is unlikely to be wholly dishonest. As we've seen, those who engage in such signalling seem to genuinely feel the emotional response they express. Such feelings are partially *constitutive* of possession of the relevant virtues. To that extent, we ought to be confident both that virtue signallers take themselves to be honest and that they have some rational basis for this judgment. That fact goes some a long way toward excusing them from a charge of hypocrisy.

6 Conclusion

The charge that someone is engaged in virtue signalling is widely felt to be a serious one. It is an accusation that stings. I hope we can now see that it should sting very much less. Virtue signalling is not an ir- or arational influence on belief formation. Rather, it provides (higher-order) evidence, which serves as an input into rational deliberation. Moreover, signalling is not a perversion of the central function of moral discourse. Independently of the role it plays in deliberation, signalling is a central function of public moral discourse, with an important role to play in enabling cooperation. Virtue signallers are not, in the main, hypocritical in their motivations and we have some grounds for thinking they are not dishonest in the signals they send.

It is important to note that the two functions of virtue signalling—its role in the provision of higher-order evidence, on the one hand, and its role in the solution to a coordination problem—are not wholly independent. 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.[13] We saw that there is good reason to think that virtue signalling is unlikely to be hypocritical in its motivations, and at least some reason to think that it is likely to be an honest signal, in the sense of expressing possession of the matching dispositions. For virtue signalling to play its role in the provision of reliable higher-order evidence, it is the first—the expression of a veridical judgment—and not the second that matters more (and it is the first for which we have better evidence). So we have good reason to think that its two functions are mutually supportive and not conflicting.

Though there are indeed cases in which virtue signalling will likely lead us to worse beliefs, vulnerability to such problems is the price we pay not (just) for allowing the

[12] Anecdotally, the accusation of virtue signalling is typically directed by the political right at the political left. Nothing in the account given here predicts that it will be politically partisan, however, and the available empirical evidence suggests that it is at least as prevalent on the right. *Expressive responding*—reporting factual beliefs that are congenial to one's political 'side' in order to express support and not because they are sincerely held—is very common, and amply demonstrated on the right. See Bullock and Lenz (2019) for review, and Schaffner and Luks (2018) for an extremely convincing demonstration of expressive responding by Trump supporters.

[13] I thank a review for *Synthese* for helping me to see the importance of this point.

expression of moral claims to play their signalling function, but also as the flip side of the epistemic *benefits* of such signals. Virtue signalling deserves condemnation neither on aretaic grounds nor on epistemic grounds. We can and go ahead and signal in good conscience.[14]

References

Anderson, L. R., & Holt, C. A. (1997). Information cascades in the laboratory. *American Economic Review, 87*, 847–862.

Bordley, R. F. (1983). A Bayesian model of group polarization. *Organizational Behavior and Human Performance, 32*(2), 262–274. https://doi.org/10.1016/0030-5073(83)90151-4.

Boxell, L., Gentzkow, M., & Shapiro, J. M. (2017). Greater internet use is not associated with faster growth in political polarization among US Demographic Groups. *Proceedings of the National Academy of Sciences of the United States of America, 114*(40), 10612–10617. https://doi.org/10.1073/pnas.1706588114.

Bullock, J. G., & Lenz, G. (2019). Partisan bias in surveys. *Annual Review of Political Science, 22*(1), 325–342. https://doi.org/10.1146/annurev-polisci-051117-050904.

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. (1992). *Testimony: A philosophical study*. Oxford: Clarendon Press.

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 1, 2017, from https://peasoup.us/2017/08/philosophy-public-affairs-discussion-pea-soup-justin-tosi-brandon-warmkes-moral-grandstanding-critical-precis-c-j-tony-coady/.

Coady, D. (2006). When experts disagree. *Episteme, 3*(1–2), 68–79. https://doi.org/10.3366/epi.2006.3.1-2.68.

Dunbar, R. (1998). *Grooming, gossip, and the evolution of language*. Cambridge: Harvard University Press.

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.

Frank, R. H. (1988). *Passions within reason: The strategic role of the emotions*. New York: Norton.

Frey, U. J., & Rusch, H. (2012). An evolutionary perspective on the long-term efficiency of costly punishment. *Biology & Philosophy, 27*(6), 811–831. https://doi.org/10.1007/s10539-012-9327-1.

Gervais, W. M., Shariff, A. F., & Norenzayan, A. (2011). Do you believe in atheists? Distrust is central to anti-atheist prejudice. *Journal of Personality and Social Psychology, 101*(6), 1189–1206. https://doi.org/10.1037/a0025882.

Gervais, W. M., Xygalatas, D., McKay, R. T., van Elk, M., Buchtel, E. E., Aveyard, M., et al. (2017). Global evidence of extreme intuitive moral prejudice against atheists. *Nature Human Behaviour, 1*(8), 0151. https://doi.org/10.1038/s41562-017-0151.

Ghiselin, M. T. (1974). *The economy of nature and the evolution of sex*. Berkeley: University of California Press.

[14] I am grateful to two reviewers for this journal and to audiences at Deakin University, Melbourne and Tilburg University for very helpful comments on this paper. Research leading to the publication of this paper was supported by the Australian Research Council (DP180102384).

Gill, A., & Lundsgaarde, E. (2004). State welfare spending and religiosity: A cross-national analysis. *Rationality and Society, 16*(4), 399–436. https://doi.org/10.1177/1043463104046694.
Goldman, A. I. (2001). Experts: Which ones should you trust? *Philosophy and Phenomenological Research, 63*(1), 85–110. https://doi.org/10.1111/j.1933-1592.2001.tb00093.x.
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.
Harris, P. (2012). *Trusting what you're told.* Cambridge: Harvard University Press. https://www.hup.harvard.edu/catalog.php?isbn=9780674503830.
Harvey, P. H., & Paxton, R. J. (1981). The evolution of aposematic coloration. *Oikos, 37*(3), 391–393. https://doi.org/10.2307/3544135.
Henrich, J. (2009). The evolution of costly displays, cooperation and religion: Credibility enhancing displays and their implications for cultural evolution. *Evolution and Human Behavior, 30*(4), 244–260. https://doi.org/10.1016/j.evolhumbehav.2009.03.005.
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.
Iyengar, S., Sood, G., & Lelkes, Y. (2012). Affect, not ideology: A social identity perspective on polarization. *Public Opinion Quarterly, 76*(3), 405–431. https://doi.org/10.1093/poq/nfs038.
Jordan, J. J., & Rand, D. G. (2020). Signaling when no one is watching: A reputation heuristics account of outrage and punishment in one-shot anonymous interactions. *Journal of Personality and Social Psychology, 118*(1), 57–88. https://doi.org/10.1037/pspi0000186.
Lackey, J. (2010). A justificationist view of disagreement's epistemic significance. In A. Haddock, A. Millar, & D. Pritchard (Eds.), *Social Epistemology* (pp. 298–325). Oxford: Oxford University Press. https://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199577477.001.0001/acprof-9780199577477-chapter-15.
Lackey, J., & Sosa, E. (2006). *The epistemology of testimony*. Oxford: Oxford University Press.
Levy, N. (2019). No platforming and higher-order evidence, or anti-anti-no-platforming. *Journal of the American Philosophical Association, 5*(4), 487–502.
Mappes, J., & Alatalo, R. V. (1997). Batesian mimicry and signal accuracy. *Evolution, 51*(6), 2050–2053. https://doi.org/10.2307/2411028.
Mason, L. (2018). *Uncivil agreement: How politics became our identity* (1st ed.). Chicago/London: University of Chicago Press.
Matheson, J. (2015). *The epistemic significance of disagreement*. New York: Palgrave Macmillan.
Norenzayan, A. (2013). *Big Gods: How religion transformed cooperation and conflict*. Princeton: Princeton University Press.
Price, P. C., & Stone, E. R. (2004). Intuitive evaluation of likelihood judgment producers: evidence for a confidence heuristic. *Journal of Behavioral Decision Making, 17*(1), 39–57.
Prior, A., Kirby, M., & Kirby, M. (2006). The society of friends and business culture. In D. Jeremy (Ed.), *Religion, business and wealth in modern Britain* (pp. 1700–1830). London: Routledge. https://doi.org/10.4324/9780203025352-14.
Pulford, B. D., Colman, A. M., Buabang, E. K., & Krockow, E. M. (2018). The persuasive power of knowledge: Testing the confidence heuristic. *Journal of Experimental Psychology. General, 147*(10), 1431–1444. https://doi.org/10.1037/xge0000471.
Schaffner, B. F., & Luks, S. (2018). Misinformation or expressive responding? What an inauguration crowd can tell us about the source of political misinformation in surveys. *Political Opinion Quarterly, 82*(1), 135–147.
Shariatmadari, D. (2016). 'Virtue-Signalling'—The putdown that has passed its sell-by date. *The Guardian*. Retrieved January 20, 2016, from https://www.theguardian.com/commentisfree/2016/jan/20/virtue-signalling-putdown-passed-sell-by-date.
Shoemaker, D., & Vargas, M. (2019). Moral torch fishing: A signaling theory of blame. *Noûs*. https://doi.org/10.1111/nous.12316.
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.

Sterelny, K. (2013). Life in interesting times: Cooperation and collective action in the holocene. In K. Sterelny, B. Calcott, R. Joyce, & B. Fraser (Eds.), *Cooperation and its evolution*. Cambridge: MIT Press.

Sunstein, C. R. (2002). The law of group polarization. *Journal of Political Philosophy, 10*(2), 175–195. https://doi.org/10.1111/1467-9760.00148.

Thouless, R. H. (2011). *Straight and crooked thinking* (5th ed.). London: Hodder & Stoughton.

Tosi, J., & Warmke, B. (2016). Moral grandstanding. *Philosophy & Public Affairs, 44*(3), 197–217. https://doi.org/10.1111/papa.12075.

Zahavi, A., & Zahavi, A. (1999). *The handicap principle: A missing piece of darwin's puzzle*. Oxford: Oxford University Press.