A Modest Defence of Somewhat Selective Outrage

Abstract: Many people think there is something objectionable about 'selective outrage'. After investigating how to best characterise what selective outrage is and what these objections target, this paper argues that many cases of supposedly selective outrage actually have important positive effects. Because we often have limited resources with which to enforce norms, it can be collectively prudent to prioritise enforcing norms that are well-established or collectively recognisable over those that are not. This will sometimes require responding to individual wrongs that seem less immoral, outrageous or in need of attention than others. We argue that when we encounter agents who are outraged about a violation of a genuinely valuable norm but not another relevantly similar violation, we should generally refrain from objecting unless we have good independent evidence the agent's outrage stems from objectionable motives.

Keywords: Outrage; blame; hypocrisy; anger; norms.

1. Introduction

Selective outrage, it is commonly thought, is bad. Certainly, few would be proud to admit that they apply their outrage selectively. The label is frequently used as a criticism to suggest that someone's outrage is somehow illegitimate or insincere, and to thereby shift attention from what the agent is outraged about to the agent themselves. To get an idea of the phenomenon in question, consider some examples that have been pointed to as examples of selective outrage.[1] Each of these follows a schema, which will be useful to detail since we have multiple agents, reactions and norm violations under discussion. Accusations of selective outrage typically concern one agent or group (S1) being outraged at one norm violation (N1) significantly moreso than another violation (N2). To this, a second agent or group (S2) objects to S1 being differentially outraged at N1 compared to N2, given N1 and N2 appear relevantly similar, and given S1 appears to lack reasonable justification for their difference in outrage. For example:

Chemical Weapons: When the Syrian government used chemical weapons to kill people in 2013 (N1), many citizens and politicians in the US (S1) were outraged. But the Syrian government had previously used non-chemical weapons to kill a much larger number of people (N2), and this generated nowhere near the same outrage, something that was objected to by many other citizens (S2) (Levs, 2013).

Terrorists: When France was subject to terrorist attacks in 2015 (N1), many people in Western nations (S1) expressed outrage and grief, and put pictures of the French flag on their social media profiles. However, much less outrage, grief, or social media flag postings occurred when Turkey was bombed by terrorists some months earlier (N2) (Saul, 2015).

Sexual Harassment: Stacey (S1) expresses outrage online upon finding out that her company's boss committed sexual harassment (N1). But since Stacey has never

[1] Or which could easily arise, in the case of *Despot*.

expressed outrage at the boss for firing employees indiscriminately (N2), or any other jerky things that bosses should not do and which arguably affect more employees, Bryan (S2) accuses Stacey of 'arbitrary deploring' (Caplan, 2017).

Despot: David asks Diane to join him in regularly donating to Amnesty International, to help fight for the rights of minority ethnic groups locked up in camps by authoritarian governments. When Diane refuses (N1), David is outraged (S1), accusing her of not caring about the rights of these ethnic groups. Diane (S2), having just read *The Most Good You Can Do* (Singer, 2015), accuses David of being selectively outraged about injustices, because he doesn't donate money to the Against Malaria Foundation (N2) like she does, which saves far more lives per dollar (McMahon, 2016).

Holiday: Hollie gets outraged at people (S1) who buy brand new SUV's (N1) rather than more efficient used cars, taking her response to be justified by their exorbitant, unnecessary emissions. However, she regularly has holidays (N2) that require international flights, which also have very high emissions. When Harold (S2) challenges her on this, she replies that she thinks these are justified because she's 'earned a break' (Gillepsie, 2011).

Though philosophers have given a lot of attention to hypocritical blaming which is similar to what occurs in *Holiday*, the kinds of behaviours in the other cases have not received much explicit investigation.[2] Our goal in this paper is to provide an investigation into the dynamics of accusations of selective outrage, and whether we should object to people who appear to exhibit it. To be clear, our primary focus is not on those who exhibit selective outrage, and whether their behaviour should be considered bad, or wrong, or objectionable, etc. (though this is relevant).

[2] Telech & Tierney (2019) argue that there is a 'comparative nonarbitrariness norm' on blaming, such that if we are to blame an agent for a norm-violation, we should also equally blame relevantly similar agents who violate the same norm. However, they don't consider different but relevantly similar norms, or norms between communities, and as we'll see, distinct considerations enter the picture once we take this broader scope.

Rather, our primary focus is on what we, the moral community, should do about it. Generally, we have reason to object to things that are bad, wrong, or objectionable (etc.), but distinct considerations can arise when it comes to thinking about practices and communities, as opposed to token actions. O'Brien and Whelan (2022), for instance, have argued that even if hypocrisy is bad, we should refrain from making accusations of hypocrisy in the political arena because we are often poorly placed to accurately assess whether a politician is being hypocritical, and such accusations disincentivise valuable practices of negotiation and compromise (cf. Piovarchy, forthcoming).

This paper proceeds as follows. First, we identify a range of ways outrage towards seemingly similar norm violations can be differential, but justified, to help home in on when outrage is or is not criticisably selective. We then outline a number of ways that selective outrage can be bad, and note that since objections to selective outrage can also be bad, this makes it unclear how we ought to respond to selective outrage. To anchor our theorising, we then use mixed-motive public goods games to model how norms can be upheld, and argue that selective outrage can sometimes have good effects because it represents a more efficient and co-ordinated enforcement of norms. We close by arguing that even if selective outrage is objectionable, these benefits support a presumption in favour of not objecting to selective outrage, unless independent evidence of objectionable motives can be found.

2. When is Outrage (Not) Selective?

Our characterisation of selective outrage contains a few things worth unpacking. We intend for outrage to be understood quite broadly. Moral outrage is often understood by psychologists as something like "anger provoked by the perception that a moral standard—usually a standard of fairness or justice—has been violated" (Batson et al., 2007, p. 1272). There is currently a lot of work being done trying to work out how outrage is related to nearby notions such as anger, hatred, criticism, blame, and indignation; as well as whether it can ever be dispassionate, whether

it focuses on victims or the wrongness of the violation itself, and whether it involves disgust (Molho et al., 2017; Royzman et al., 2014), among other things.[3]

For our purposes, these debates can largely be set aside. Accusations of selective outrage don't always target literal outrage; often they target harsh criticism, blame, anger, or reproach. Accusations of selective outrage seems unlikely to be defused by replying 'but I wasn't outraged, I was merely angry'. Moreover, because outrage motivates us to behave in certain ways, objections to selective outrage don't always target the literal emotional reaction. Sometimes they target the amount of attention being devoted to an issue, its prevalence in discussions or media airtime, the resources being allocated, and the sanctions, protests, or various other ways that S1 can express a negative reaction towards N1. Thulin and Bicchieri (2016), for instance, demonstrate that outrage not only motivates punitive responses, it also motivates compensatory responses to victims of norm violations over and above empathetic responses.[4] Nevertheless, people who are selective in these kinds of affective and behavioural responses to norm violations seem guilty of the same fault that people selectively exhibiting literal outrage are accused of.

Because accusations of selective outrage are concerned with a broad array of responses, and because our argument below does not turn on any particular theory of such emotions or responses, we intend for out treatment of 'selective outrage' to be ecumenical. Our primary interest here is the perceived selectivity, rather than the outrage. Our sense of S2's 'objecting' is

---

[3] Haidt (2003) proposes that outrage is 'disinterested', in the sense of not involving identification or empathy with the victim, but Hechler and Kessler (2018) find evidence that conflicts with this. A common finding is that anger and outrage are quite sensitive to the norm-violator's intent (Ginther, Hartsough and Marois, 2022). While the philosophical literatures on blame and anger are quite well-established, the literature on outrage independent of these other notions is relatively nascent. Nguyen and Williams (2020), for instance, have written about the nature and badness of 'moral outrage porn', although they do not provide a sustained examination of the nature of outrage itself. Hirji (2022) argues that 'outrage anger' is a distinct attitude which an attitude directed at circumstances in which a violation is not fully intelligible to the dominant moral community, whose central function is to close off the victim's ability to feel empathy for their abuser. However, she explicitly distinguishes this from the more common phenomenon of moral outrage, and doesn't aspire to capture psychologists' or folk usage of the concept (p. 9).

[4] Thanks to an anonymous reviewer for this source.

also intended to be broad, and can itself manifest as outrage. For readability though, we will use 'outrage' for S1's response to the initial norm violation, and 'objecting' to refer to S2's response to S1.

There are a number of ways one can argue that what appears to be selective outrage is differential, but justified and thus not 'selective' in any criticisable sense. For instance, it is often reasonable for people to be more outraged at their own governments violating the rights of fellow citizens than they are outraged at foreign governments violating other citizens' rights in comparable ways. One might have a personal relationship with their fellow citizens which grounds duties of partiality. That a fellow citizen's rights have been violated might also mean your rights are at risk of being violated, and outrage might be a means of rallying others to help prevent this. You are also likely to have a much better understanding of the facts of the case when it concerns your own government, and a greater capacity to enact change. Government representatives who want to stay elected typically need to be much more responsive to what their nation's citizens think of them than what other nation's citizens think. These are all justifications for not experiencing the same level of outrage in response to N1 and N2. Pointing to these justifications is a way of saying that, given the facts of the case (partiality, self-defence, greater capacity, lower risk of error), one *actually* has more reason to be outraged about N1 than N2.

Another sense of 'justification' is also relevant here, regarding things that are reasons by the agent's lights. There are instances where, although the facts of the case may not warrant differential levels of outrage, the agent's beliefs in some way count against attributing to them charges of selective outrage. A clear example of this concerns cases where the agent couldn't have been expected to know better. Suppose Suzy is outraged that Jeff was fired, but indifferent to Jess being fired, and there are no other differences between Jeff and Jess. Suppose, however, that Suzy has good but misleading evidence that Jess was stealing money from the company, and this explains why she is not outraged at Jess's treatment. Here, we would say that Suzy is not

exhibiting selective outrage, even if Jess and Jeff's treatment by their boss is, in fact, equally outrageous.[5]

Relatedly, our accusations of selective outrage are often sensitive to an agent's moral beliefs. Consider Singer's (1972) argument that failing to save a drowning child is relevantly similar to failing to donate to charity. Even if Singer is correct about this, many people who would be outraged at the former but not the latter do not exhibit selective outrage. Why? Because they reasonably take there to be important moral differences between the two cases (as many philosophers do), even if they cannot formulate a bullet-proof argument for why upon questioning. Being mistaken does not thereby make one unprincipled.

Some caution is needed with this line of thought though. We don't want to say that agents do not exhibit selective outrage so long as they take there to be *some* relevant difference between the two cases. Too often, people confabulate reasons for there being relevant differences warranting differential outrage, which don't stand up under scrutiny. To see what we mean, consider *Terrorists*. A plausible-sounding defence one might offer in favour of being more outraged about terrorism in France than Turkey is that the local media in Western nations are much more likely to run stories about the former. As a result, citizens are much more likely to be aware of these wrongs. Additionally, people have a limited amount of time, effort and attention to inform themselves about what is happening in the world. They cannot research everything, so they need to outsource their investigations to journalists, and if the journalists unjustifiably favour certain stories over others, that is their fault, not the citizens. The problem with justifications like these is that the causal story often runs in the other direction: media outlets give reduced airtime to acts of terrorism in certain countries because they know citizens are *already* disposed to care less about them. Moeller (1999) reports that, from interviewing journalists in the 90's, sayings such as

[5] It may be tempting to characterise this as an excuse. But, as we'll see below, because of the way that selective outrage implicates the agent's motives, misleading information doesn't make it the case that one was being selective but was excused, it means they were not being selective at all.

"one dead fireman in Brooklyn is worth five English bobbies, who are worth 50 Arabs, who are worth 500 Africans" (p. 22) were not uncommon. This seems like evidence that citizens in Western nations are, in fact, disposed towards selective outrage about wrongs that depend on the ethnicity of the victims, and that attempts to put the blame entirely on the media are *post hoc*.[6]

In discussing real-world cases, where all else is not equal, we also need to be careful not to take the fact that there is *some* relevant difference present to show that the difference in *actual* outrage is thereby principled. For example, even if chemical weapons cause more pain than non-chemical weapons, and this merits different levels of outrage all-else-being-equal, *Chemical Weapons* still seems to be a *prima facie* case of selective outrage because (a) this difference is not enough to justify the differences in outrage that were actually experienced (given murders by governments remain very outrage-worthy) and (b) even if being a non-chemical weapon did somehow justify a much smaller amount of outrage, that difference should be more than offset by the much larger number of deaths caused by non-chemical weapons.

However, the possibility that people have beliefs which count against considering all differential outrage to be selective is why we have included a 'reasonable' clause in our characterization: we think that it helps draw the line in approximately the right place. Someone doesn't exhibit selective outage just because there are, in fact, no moral differences between the two cases, but nor do they avoid exhibiting selective outrage just because they believe there are no moral differences. Their beliefs need to have some reasonable basis. We will not attempt to provide any further account of what kinds of circumstances make differential outrage reasonable and which don't (as debates about tolerance of 'unreasonable' comprehensive moral doctrines, and 'reasonable' disagreement, have struggled with for decades), but this will not pose any issue for our argument. It will suffice for us to focus on paradigmatic instances of selective outrage to make our case.

[6] Though the media cannot put all the blame on viewers either.

3. What's Bad About Selective Outrage?

Having looked at what counts as selective outrage, and some circumstances that plausibly defeat such accusations, let's now try to get clearer about why selective outrage seems objectionable. We propose that instances of selective outrage involve behaviour that can be bad for a number of (often overlapping) reasons. Sometimes, people in the position of S2 use the fact that the norm violations seem similar in order to argue that one of S1's reactions are unfitting or disproportionate. S2 is effectively saying that *given* how outrageous N1 is, N2 is equally outrageous, and thus S1 should be outraged. In *Terrorists*, for example, those complaining about selective outrage were not attempting to get citizens to temper their outrage towards the terrorists who attacked France. They were trying to get citizens to increase their level of outrage at the attacks in Turkey. But the opposite move is also sometimes made, where given how *not* outrageous N2 is, S2 tries to get S1 to think N1 is also not a big deal. This happens, for instance, in politics when one party argues that a new, supposedly outrageous policy is very similar to some other, already accepted policy, and thus the opposition's outrage is confected.

While concerns of fittingness and proportion account for some cases, they cannot be an exhaustive explanation of why selective outrage seems bad. We can see this by noting that if unfittingness were the primary concern, S2 should be equally concerned about other agents whose responses are unfitting, but not differential. On this framing, S1's lack of outrage towards N2 should be no more objectionable than towards someone else who has failed to experience outrage at either N1 or N2. (In fact, it should be *less* objectionable, since S1 is often correctly representing the outrageousness one of one norm violation). Additionally, this doesn't seem to fit the phenomenology of our accusations when we are S2: we don't just object on the basis that N2 is outrage-worthy. Frequently, part of our objection is about S1's level of outrage towards N1 *in tandem* with the differing level of outrage towards N2, which is why S1 is accused of selective outrage, not insufficient or disproportionate outrage.

Why might the conjunction of these two emotional reactions be objectionable? One possible answer concerns fairness. Outrage can confer certain benefits on victims. It draws people's attention to the wrongdoing they have suffered, it signals that this action was unacceptable, it communicates respect for the victim (Tierney, 2021). When people exhibit selective outrage then, they are effectively conferring a benefit on some victims but not others. It seems reasonable for the victims who miss out on these benefits to have a legitimate complaint here, and for others to object on their behalf. This seems especially the case in *Terrorism*, for instance. And even if the outrage towards French victims didn't lead them to gain more benefits, arguably the outrage implies something about the perceived status of the victims.

Again though, while this accounts for why some instances of selective outrage are bad, this isn't a complete explanation of our practices. In many other cases, such as *Chemical Weapons*, the focus of S2's objection is not primarily concerned with victims; it seems to be much more centered on how S1 is responding to perpetrators. Additionally, even if S1 were to compensate or aid the N2 victims to the same degree as that received by the victims of N1, many would still feel there were grounds for objecting to S1 being selective. Agents accused of selective outrage are exhibiting some kind of *fault* that reflects badly on *them*.

One kind of fault which is often tied to accusations of selective outrage is the fault of hypocrisy. Hypocrisy is often described as a kind of false pretence, where the agent pretends to be more virtuous than they are. Rossi (2020), for example, describes hypocrisy as vicious, value-expressing inconsistency. The relevance to our target cases is that, if S1 is expressing outrage at N1, but not N2, this can be evidence that they don't care about N1 or value it as much as they are making out.

This seems to be a particularly apt characterisation of the fault displayed in *Holiday*. Hollie exhibits outrage that is both selective and hypocritical (where this selectivity contributes to it being hypocritical). But a few things count against thinking hypocrisy can provide a complete explanation, particularly for the other thought experiments given above. First, hypocrisy is much

more strongly associated with making an exception of one's self. The 'false pretence' stems from behaving in a way that communicates *they* are disposed to reliably abide by the norm when they are not. This is why the literature on hypocrisy and standing to blame has focused on hypocrisy's tendency to generate responses like "Who are *you* to blame *me*?" or "Look who is talking!" from others (Fritz & Miller, 2018; Piovarchy, 2020, 2023; Todd, 2019). We can also see this isn't quite the focus of accusations of selective outrage by noting that we gain an *additional* objection upon noticing that agents we are accusing of selective outrage don't abide by the norms in question themselves. It's one thing to object to some terrorists without objecting to others, it's another to do so while being guilty of terrorism yourself. When Western citizens are outraged at terrorism in France but not Turkey, S2 is not worried that these citizens are in some way sympathetic to terrorism or at risk of being terrorists themselves. Accusations of hypocrisy typically target one's standing, right, or entitlement to blame or make certain kinds of pronouncements. When we object to the citizens' selective outrage, we are typically not implying that they lack the right to experience outrage at all and that it would be preferable for them to remain silent. Hypocrisy is criticisable, and can manifest in selective outrage, but not all selective outrage is reducible to hypocrisy.

A final fault that seems relevant to selective outrage is grandstanding. Tosi and Warmke (2020) argue that what is objectionable about grandstanding (which often takes the form of outrage) is that the agent is motivated by a 'recognition desire': a desire to be seen as morally respectful, typically by their in-group. This is criticisable because it effectively devalues the currency of moral talk. As moral talk becomes a vehicle for self-promotion, rather than an important tool for moral co-ordination, people become cynical and start ignoring moral talk, even from agents who are sincere. In *Sexual Harassment*, if Bryan were to notice that Stacey experiences outrage about her boss when there is an audience of peers present, but fails to experience outrage towards relevantly similar wrongs when her peers are absent, this seems like evidence she is not being motivated by concern for the norm itself. That one's outrage is

differential is evidence the agent has a criticisable motivation, and this is what underlies the accusation that said outrage is 'selective'.

We think that worries about grandstanding are close to the mark: selective outrage is commonly taken to be evidence that the agent's *motives* are questionable. But a caveat is that many cases of questionable motives do not specifically involve a desire to be seen as respectable, and so grandstanding will not be a complete explanation; one might be motivated by some other kind of self-interest. Suppose, for instance, that George expresses outrage at Al's behaviour, but not Ralph's relevantly similar behaviour. Suppose George does this, not so that voters think he is respectable, but so that voters become dissatisfied with Al and go vote for Ralph, splitting the vote which then helps George win. George's selective outrage may have been motivated by a desire, but it won't be a recognition desire, and so he won't count as grandstanding.

We propose that what accusations of selective outrage primarily target is the agent's motivations, cares or commitments. Objections to selective outrage seem to communicate something like the following: "You profess to be outraged at N1. But you were not outraged by N2, which is relevantly similar. If you had genuine concern for the norm, you would care about N1 and N2 equally. Your failure to express equal outrage is evidence that your outrage towards N1 is being criticisably motivated by something other than concern for the norm violation itself." A concern with motivations accounts for the multiple ways selective outrage can be bad that we have identified. If S2's concern is the fittingness of S1's outrage, there is an implication that S1 does not care sufficiently about N2, or cares too much about N1. If the concern is unfairness, the implication is that S1 doesn't care enough about the victims of N2, or is perhaps disposed to objectionably favour victims of some groups over others. If the concern is hypocrisy, the implication is that S1 is not as motivated to uphold the norms or values they pretend to be. If the concern is grandstanding, the implication is that they are motivated by a

recognition desire. And there are many other criticisable reasons why someone might be motivated to selectively express outrage, such as when they will benefit from doing so.[7]

4. The Case Against Objecting to Selective Outrage

Having identified a range of cases in which selective outrage seems to be objectionable, clarified what kinds of things count as selective outrage and the nature of our objections to it, and identified that the heart of the objection seems to regard criticisable motivations, let us now consider an argument against objecting to selective outrage. Case (2017) points out that anyone who finds selective outrage objectionable is likely to be hoisted by their own petard. Suppose selective outrage is objectionable. People who object to selective outrage thus seem well within their rights to do so. But by their own lights, they ought to object equally to all relevantly similar cases of selective outrage. They cannot be selective with which instances of selective outrage they are objecting to.

But of course, people are selective about which cases of selective outrage they object to! For whatever reason, some instances seem more worth attending to and highlighting than others. Sure, there can be grounds for focusing on some instances more than others (e.g. partiality, greater capacity and familiarity, relevance to one's self). But we've already seen these kinds of justifications are available to the people exhibiting selective outrage too, and one needs to show

[7] One option at this point is to reserve the term 'selective outrage' for instances where the agent *in fact* has a criticisable motivation, and use the term 'differential outrage' for instances where we observe S1 display different reactions to N1 and N2 but are not commenting on their motives. While we think this has merit regarding philosophical accuracy (particularly when it comes to discussions where we are not yet quite sure if the agent has criticisable motives), people tend to use the term to imply S1 has criticisable motives *on the basis of* there being differential outrage without clear justification, or without checking if S1 has criticisable motives. Below we argue against these tendencies. Rather than encouraging people to push back against those who describe S1 as selective (given the difficulty of ascertaining motives, and established patterns of usage) we think it will be more productive to encourage people to refrain from objecting *even if* the outrage appears selective due to being differential. That is, responses like 'We don't know if it's selective because we're not sure of his motives' seem less productive to us than responses like 'Even if his outrage seems selective, there might be good reasons to let it slide'.

that these justifications warrant the difference in outrage that was actually experienced, not just some degree of difference. Given how common selective outrage is, it is extremely unlikely that people who object to it are doling out their objections consistently, or proportionately to the egregiousness of S1's selectivity. So even if selective outrage is, in some sense, objectionable, by their own lights people who are outraged about selective outrage exhibit a kind of moral inconsistency.

In response, it is worth noting that the above argument doesn't show that selective outrage *is not objectionable*. It shows that those who object to selective outrage are flawed humans, or that prohibitions against selective outrage are demanding, but this is not quite the same as showing that it is false. By comparison, many people think the fact that many utilitarians fail to maximize utility doesn't thereby show that utilitarianism is false, at least not without some supplementary argument regarding why moral principles cannot be overly demanding (Sobel, 2007).

But as we noted at the beginning, we are primarily interested in whether we should object to selective outrage as a matter of public discourse, or as part of our moral practices. Establishing that selective outrage is objectionable without establishing that we ought to, in fact, object to it would be rather a hollow victory. And Case's argument arguably counts against objecting in practice by showing that people who object to selective outrage cannot avoid being *hypocritical*. Their failure to abide by their own pronouncements—to talk the talk without walking the walk—arguably undermines their right to object to people who display selective outrage, or at least their right to have their objections taken seriously by others.

While this may seem to put the selective outrage opponent on the back foot, they can make a similar move. Although hypocrites are thought to lack the standing to blame, such that hypocritical blame is bad, or objectionable, it is commonly accepted that this isn't the same as showing that agents should never, in practice, express hypocritical objections. Indeed, some philosophers argue that worries about hypocrisy are ill-founded, or that important benefits are

lost if we get too hung up on who has the right to blame instead of acknowledging when we've done wrong (Bell, 2013; Dover, 2019).

This seems to leave us with something of a stalemate. In principle, selective outrage seems objectionable, but objecting to selective outrage seems hypocritical. It is hard to know what impact these facts should have on our discourse and actual practices, all-things-considered. Trying to resolve this by looking at the long-term effects of policies either allowing or prohibiting expressions of selective outrage seems likely to devolve into speculation and a trading of intuitions.

What would be helpful is if we had some way to anchor our theorising about the benefits and costs of outrage being applied selectively. This is what we shall provide. Both outrage and accusations of selective outrage seem to have a *point*, and it's worth considering what these points might be and how valuable they each are. We think that once we look at the broader purpose of our responses to norm violations, we will see that selective outrage can deliver important benefits that objections to it risk undermining.

5. How Selective Outrage Can Be Beneficial

To understand our argument, it helps to first understand what are known as *mixed-motive public goods games*. In these 'games', agents must choose to put some resources either into a 'pot' (co-operate) which will then be redistributed to everyone, or keep their resources ('defect'). They face the following incentives. The money that goes into the pot experiences a multiplier before being redistributed. If everyone co-operates, everyone does much better, particularly because they can *reinvest* the extra benefits, making everyone do even better yet again. But, if one agent defects while the others co-operate, that agent does fantastically as they get both their starting resources and some of the redistributed pot, while those who contributed end up worse off than if they'd never contributed at all. However, if everyone follows this strategy, then no-one contributes, and no-one gets any benefit.

These games model many situations agents find themselves in when deciding whether to act morally, and have been used to explain how moral norms developed (Curry, Mullins, & Whitehouse, 2019). Many moral norms provide important collective benefits when practiced at scale, over the long run. Individuals can do better for themselves by defecting, but if too many people follow these incentives, everyone ends up worse off. If I don't pay my taxes I can save for a new car, but if everyone thinks like this there won't be usable roads to drive on. If I break my contract to build your house I can have a fun weekend spending your money, but if everyone does this we'll be unable to count on anyone to provide us with services. Given these situations, many agents tend to act as conditional co-operators: they will co-operate if they believe that a sufficient number of others will too.

However, not everybody requires the same level of cooperation from others before being willing to cooperate themselves. This means that small numbers of initial defections can lead to cascades of more (Fehr & Gachter, 2000). For example, if I defect upon seeing three people defecting, this will then trigger anyone who would defect at seeing four people defect to do so, which will then trigger anyone who would defect upon seeing more than four people defect, and so on. It is thus very valuable to have some mechanism that can prevent small numbers of initial defections from occurring. This is where sanctions come in. Sanctions increase the cost of defecting, thereby deterring it. But to have this deterring effect, the threat of sanctions needs to be credible. If there is only a small chance of defections being sanctions, many would-be norm violators will take their chances.

One thing that makes it easier to get away with defections is an increase in the number of people defecting with you. There's a limited number of agents who are willing and able to enforce norms by sanctioning, and any given community has a limited amount of resources with which to enforce particular norms (e.g. time, effort, money, the opportunity cost of enforcing

other norms).[8] Since only so many people can be sanctioned at any given time, the more defectors there are, the less likely it is that any given defector will be sanctioned, and at a certain point the norm violators will overwhelm the norm enforcers, leading even more people to become norm violators.

How does this model help us think about selective outrage? Would-be defectors are far more likely to encounter higher costs when they defect in communities of agents prone to outrage than in communities of agents not prone to outrage (Gaus, 2011). Outrage (and associated reactions) draw the attention of everybody to your conduct (Brady, Gantman, & Van Bavel, 2020), increasing the chance of sanctions (Ginther, Hartsough, & Marois, 2022). It motivates people to stop assisting you, and to set your interests back. Additionally, the outraged person makes clear that they expect others to be on their side, and without some apology or atonement, your reputation is likely to suffer as outraged people are motivated to gossip about your transgressions to others. Outrage and associated reactions also signal something like 'this behaviour is unacceptable, and should not be tolerated' (Shoemaker and Vargas, 2021), giving assurance to others that the norm is important and will be complied with. Additionally, as norm-internalising beings, part of the badness of being a target of outrage is in knowing that people think poorly of you. It is unpleasant to be on the receiving end of outrage (Gavrilets & Richerson, 2017; Henrich & Ensminger, 2014). We are sensitive to our social status, and outrage is a reliable way of diminishing a target's esteem in their community.

With this background on the table, we are now ready to see how selective outrage can be beneficial. Because outrage contributes to norm enforcement, and because the ease with which norms can be enforced depends greatly on a number of factors (such as what the current levels of norm-compliance are), the efficacy of outrage at enforcing norms is also going to be

[8] Note that there are many costs in working out how to co-ordinate (e.g. form, hire, train, and run a functioning police force), which is not always obvious in these games where individuals can just decide to sanction.

significantly mediated by those factors. The difference that makes it harder for communities with more norm violators to enforce said norm than communities with fewer norm violators is the same difference that occurs *within* a community currently trying to uphold *different* norms with different levels of compliance. Suppose norm N1 is well-known and often complied with, while N2 has fallen out of favour and no longer has widespread endorsement. Suppose also that agents cannot punish defectors of N1 and N2 at the same time; perhaps it takes time and resources to monitor for any given defection. Within this community, individual members will often produce more benefit trying to enforce N1 rather than N2. Given limited resources, if they each tried to follow the maxim of punishing defections of N1 and N2 equally, or punishing proportionate to the number of defections of each, there is a great risk that they will not greatly improve rates of N2 compliance, while allowing a cascade of N1 defections to develop. In this kind of scenario, having agents who are disposed to be more outraged about N1 than N2 is going to lead to important benefits. Having agents who are *selectively outraged* about N1 will be beneficial.[9]

Once established, norms are only maintained if enforcers keep up enforcing, and make the threat of sanctions credible. From time to time, some agents will try to test the commitment of enforcers by violating the norm. If the enforcers don't demonstrate that defectors will be sanctioned (or give some credible signal that they will), others will then attempt to violate the norm too. It thus often makes sense for enforcers to sometimes spend resources upholding well-understood and accepted norms that is somewhat disproportionate to the prevalence or overall apparent badness of token violations.

To be clear, we are not saying that selective outrage is only beneficial when it favours one norm that currently has higher compliance than the other norm; other circumstances exist. For

[9] Perc and Szolnoki (2012) for instance, demonstrate that agents who increase their propensity for punishment upon noticing that defection is *spreading*, but who reduce their disposition to punish if they don't encounter any violations for long enough, can outcompete agents with members who are all stably disposed to punish defections to a fixed degree. If an increased disposition to monitor for defections and punish incurs increased costs, the former 'adaptive' form of punishment is much more efficient at stabilising cooperation than the stable form.

instance, suppose our community organized themselves to co-ordinate and make a sustained effort to enforce N2, and temporarily gave more resources to their pool punishments systems, perhaps having decided that the costs of continued violations were too great to bear.[10] They could feasibly get the number of defections down to a level at which they would no longer need to keep coordinating and spending those extra resources, and could then fall back to relying on uncoordinated individuals to peer punish any defections they observe (allowing other individuals to return to enforcing N1), without risking a return to the prior rate of N2 defections. In such a case, it would make more sense to exhibit selective outrage favouring N2 rather than N1, especially if the community is likely to decrease defections of N2 much faster than N1 compliance will drop.[11]

Other factors can also make it beneficial to selectively enforce some possible norms over others that seem morally similar. One important factor is the ease of finding agreement regarding what *kinds* of actions count as violations. For example, suppose it is widely accepted that there is a moral principle that 'governments should not kill civilians of any nation without a legitimate reason', and also that it is sometimes legitimate and necessary for a government to kill some civilians. Unfortunately, it is very difficult to find any consensus over what counts as 'a legitimate reason', or to recognize from the outside which killings were legitimate given the government in question may possess information about its targets that everyone else lacks. Given a civilian has

[10] Punishment can be doled out directly by individuals in what is known as 'peer punishment', or individuals can contribute to some kind of 'pool' which then doles out 'pool punishment' (such as a police force and judiciary) (Traulsen, Röhl, & Milinski, 2012). Some forms of punishment are in-between, or involve both. For example, in giving a journalist a scoop so they publish a story that tarnishes your reputation, I have indirectly punished you at little cost to myself, but individual media organisations are not something that everyone (or even most) individuals contribute to.

[11] Note also that outrage could also be used to build consensus around what is presently a controversial norm. Perhaps while the community overall doesn't accept a norm, a sub-group does, and by expressing outrage they could eventually cause the wider community to adopt their norm, perhaps by communicating how this norm is relevantly similar to some other norm the community already endorses. Thanks to an anonymous reviewer for this point.

been killed by a government, there may be relatively low certainty or consensus about whether this violated the principle.

Contrast this with the principle 'governments should not kill civilians with chemical weapons'. If there is consensus that, despite needing to sometimes kill civilians, none of the justifications for doing so require the use of chemical weapons, it will be far easier to find consensus and establish a norm around not using chemical weapons that governments actually comply with. This will be even easier still if the distinction between 'chemical weapon' and 'non-chemical weapon' is relatively easy to agree on, with few edge cases. Note that the higher level of compliance here comes *not* exclusively from the parties believing that chemical weapons are especially worse than non-chemical weapons, but partly from the ability of parties with varied and conflicting interests to *agree* on what counts as a chemical weapon and its lack of perceived overlap with other actions the parties sometimes want to allow. 'Chemical weapons' acts as a kind of recognizable, consensus-friendly category with which a norm can be built as a *line in the sand*, around which agents then co-ordinate. Especially if the norm has been successfully adhered to for many decades, it will be hard for governments who violate the norm to argue they have an excuse or plausible deniability, and act as evidence that they are disposed to violate other, similar norms, which are currently widely complied with, in the future.

These kinds of examples, we propose, help characterize many situations featuring agents who experience selective outrage. It is very rare that all else is equal regarding two seemingly morally similar norms. In many cases, these differences help explain why some norm violations receive more outrage than others. Outrage is often selectively experienced towards some wrongs more than others not because the former are the most important, or because the agent is disposed to arbitrarily favour those kinds of wrongs, but because those wrongs are *perceived to be violations of a norm that is well-established in this community*. If a norm violation is unusual, focusing on it rather than another that is more routine can have a greater net effect on norm-following overall, as it helps prevent cascades of defections from occurring and allows for a more efficient use of

enforcement resources. In *Despot*, David is not being objectionably selective regarding which lives matter most by favouring donating to Amnesty International rather than the Against Malaria Foundation, even though the latter saves more lives. This is because 'don't lock up minority ethnic groups in concentration camps' is the kind of norm that much of the international community takes to be well-established and historically important for stemming further widespread human rights violations. Allowing violations of this norm to pass without outrage risks signalling to other nations considering doing the same that they could get away with it too, or to the nation in question that other human rights abuses won't be met with penalties.

To bring this all home: outrage is a mechanism by which moral agents can *signal* to everyone that this kind of thing is not to be done, *co-ordinate* with other norm enforcers by communicating that *this* is where they need to attend, and *deter* violators in order to uphold norm-following. By expressing outrage, one is effectively saying to others that *this* wrong warrants *immediate* attention, that it is *not acceptable* and that this kind of conduct will not go unchallenged. But how effective it is at in fact upholding the norm depends greatly on a range of factors which, in practice, we cannot abstract away. These include: the current level of norm compliance for each norm, the number of people willing and able to sanction, the ability of both defectors and sanctioners to co-ordinate with others in their group, the robustness of conditional co-operators' willingness to co-operate, the information agents have about other agents, the amount of resources (time, effort, attention etc.) that enforcers have to devote to enforcing and the opportunity costs of favouring any given approach over others, how easy it is for parties to find consensus on what types of actions violate which norms, what counts as justification and what counts as evidence of wrongdoing, and whether there are currently any efforts underway to build support for a new norm, among other things. These factors interact in important ways that can make selective favouring of the enforcement of some norms over others (including by being outraged at their violation) more effective at upholding norm compliance overall, providing important benefits.

A benefit of our account is that it can explain some of the disagreement between people in the positions of S1 and S2. From the latter's perspective, S1 is often objecting to something that is relatively rare and which may not seem to need much attention, given violations are presently infrequent.[12] If what ultimately justifies outrage is the level of wrongness of individual acts, consistency does seem to demand that S1 be equally, if not more, outraged about N2. Given S1 does not experience this outrage, this makes S2 suspect S1's motivations. We think that, while S2 makes a reasonable inference, once we have an understanding of collective action problems of the kind that mixed-motive public goods games model, and the way norms are upheld, S2 should be much more hesitant to infer that S1 has criticisable motives. Though level of wrongness is relevant to determining what is outrage-worthy, and the moral similarity of wrongs is relevant to determining when someone is applying their outrage reactions consistently, *these are not the only factors relevant to assessing whether someone's outrage is criticisable*. Frequently, they are responding to norms that currently have different levels of current compliance, different risk of subsequent compliance if not enforced, different ease of identifying action-types, and the agents in question often possess different levels of evidence regarding each of these factors.

6. Motivations and Objections

We have just provided a model which can account for ways selective outrage can occur without any clearly criticisable motives. How frequently does this represent the actual instances of selective outrage we observe? It may not be possible to put any rough percentage on what proportion of selective outrage is accounted for by this kind of scenario, rather than e.g. a recognition desire. It is in this regard that our defence is 'modest': we are seeking to establish that people should be much more reluctant to interpret instances of selective outrage as undesirable

[12] Another relevant factor concerns roles; people often can contribute much more to particular types of norms in virtue of their implicit or explicit roles, such as by being members of a particular political body, and this can plausibly generate role-specific duties to care about that norm more than others.

and thus worth objecting to than they currently do, even though we do not aim to establish that selective outrage is always or in most cases beneficial.

A few things count in favour of thinking that it is not uncommon for this model to account for actual instance of selective outrage that we encounter. First, research from psychology demonstrates that people's moral judgments are sensitive to factors relevant to co-ordinating e.g. salience of the wrong (Amit & Greene, 2012), extent to which norm is endorsed by others (Marton-Alper, Sobeh, & Shamay-Tsoory, 2022), and perceptions of usualness (Effron, 2022). There is considerable evidence that one of the traits that makes humans unique is that we have a *norm psychology*: "a suite of genetically evolved cognitive mechanisms for rapidly perceiving local norms and internalizing them" (Chudek, Zhao, & Henrich, 2013, p. 443; see also Chudek & Henrich, 2011 and Gelfand, Harrington, & Jackson, 2017). The ease with which we internalize norms (then becoming motivated to uphold them) suggests that much of our outrage at norm violations stems from a genuine concern for the norm, rather than some other criticisable motive masquerading as concern. Additionally, our model does a good job of accounting for our intuitions about various cases. We have explained how there are differences in factors relevant to establishing stable norm-compliance which favour focusing on 'don't kill citizens with chemical weapons' rather than 'don't kill citizens without a good excuse', as well as 'don't put minorities in concentration camps' rather than 'save the most number of lives'. Attempts to create greater recognition of norms like 'don't sexually harass employees' plausibly account for why it receives more attention than 'don't be a jerky boss' in *Sexual Harassment*. We think that these are the cases featuring selective outrage that readers were least likely to consider criticisable. In contrast, it's not clear such differences are present when it comes to comparing buying a high-emissions car with having a high-emissions holiday, and so the differential outrage in *Holiday* does seem objectionable.

Let us now consider a challenge. Although our story may explain how individuals who experience selective outrage can be rational or produce good effects from the perspective of

collectives in public goods game-style settings, this is not quite the same as demonstrating that individuals are not acting objectionably. We can strengthen this line of argument by pointing out that most agents do not have an understanding of collective action problems and norm-enforcement. People who exhibit selective outrage do not think to themselves that N1 is well-understood and thus being more outraged at N1 than N2 is likely to be a more effective use of limited resources, nor are they aiming to prevent cascades of defections from occurring. The benefits that come from being selective with one's outrage isn't *their* motivating reason even if it helps explain why they experience selective outrage (Sandis, 2015). When we pay attention to what is or is not motivating *them*, they seem motivated by particular features of *this* norm violation but not others without being able to offer a principled justification. And perhaps this *lack* of justification for responding differently *just is* what warrants attributing an objectionable motive to them.

We think this line of thought is reasonable, but have two replies. The first is to return to the distinction between what is objectionable in principle, and what should be objected to, all-things-considered, in practice. Grant for the moment that selective outrage reflects an objectionable motivation or attitude. We would then ask S2, "What is the point of your objecting? What do you want to achieve?" Usually when we object to things, is because we want that kind of behaviour to stop, or change in some way. Many people who object to selective outrage do so because they want the outraged person to also be outraged about some other matter. But there are two ways that S1 can defuse the charge of selective outrage: take up being outraged at N2, or cease being outraged at N1. While S2 is presumably aiming for the former, there is a large risk that, upon being objected to, S1 avoids charges of selectiveness *not* by opting to get outraged about N2, but by ceasing to express their outrage at N1 to avoid criticism. As Dover (2019) puts it: "the injunction to 'practice what you preach' can be used either to discourage 'preaching' or to encourage practicing." (p. 413). And given the story we have just told about how norms are upheld, this would be an unfortunate outcome.

More importantly, since people often have limited resources with which to uphold norms, and co-ordinating is difficult, there is a non-trivial chance that focusing on N2 will take away from the enforcement of N1 (or some other norm). We need to be sensitive to the costs that come with the dispositions that *would* make S1 respond in the ways that S2 is implicitly trying to get them to develop. By consistently responding to the moral features they are cognizant of, agents risk falling prey to increased difficulty co-ordinating, as well as a higher rate of false positives (identifying something as a norm violation that wasn't, given one is relying less on consensus from others before being outraged).

One might question how limited these resources are. Surely most people could show much more outrage at violations than they currently do; they are not near the ceiling of where more outrage would start to negatively affect their lives, or impact on their other interests. While this is true, it is worth remembering that we want our outrage qua emotion to spur action, which does draw on resources that are more plausibly limited (such as those that pooled punishment systems rely upon), and that a central difficulty when upholding norms concerns how to organise. While outrage qua emotion can help people coordinate, this doesn't entail that more outrage or more frequent outrage generates better co-ordination. Agents that are too prone to expressing outrage when nothing can be done risk cheapening its reliability as a signal that others ought to attend to *this* matter and try to do something (cf. crying wolf). If Stacey spends all her time being outraged at e.g. the emissions of some obscure company several continents away, expecting members of her local rural town to do something about this, and her neighbours keep attending to her chosen issues only to realise there is simply nothing to be done, eventually they will start to simply tune her out.[13]

[13] There are also other risks. If sanctions are perceived to be disproportionate, or the norms illegitimate, this instead generates resentment, spite, and feuds can develop (Nikiforakis, Noussair, & Wilkening, 2012).

The second way to reply to our opponent is to question whether we should think that dispositions which result in selective outrage are objectionable, though the in-practice considerations described above remain relevant here. S1 arguably seems 'objectionable' because we implicitly compare her to an agent who responds equally well to N1 and N2, who seems more virtuous. But a fairer comparison needs to look at how communities of people who are worse at co-ordinating precisely because of their less objectionable motivations. When we keep in mind that contrastive 'unobjectionable' dispositions have costs we don't want to incur, the claim that S1 has 'objectionable' motives loses some of its bite.

Additionally, the argument that people should not get credit for their norm-upholding behaviour because they cannot consciously explain the rationale for it risks putting too much emphasis how well one can articulate their reasons over how well one *responds* to reasons. If S2 accepts the story we have told about the occasional benefits of selective outrage, but maintains that S1 is nevertheless behaving objectionably, they seem to be saying that S1 is someone who does the right thing for the wrong reasons. But this is an inapt description of many cases. It is not merely an *accident* that people exhibit this tendency. Rather, given the way that norms are upheld, and the importance and difficulty of co-ordinating with others, we think that, in general, agents who exhibit selective outrage because they respond to norms that are (e.g.) salient, well-accepted, and have a history of precedent around which some consensus has developed, count as being responsive to reasons.

If reasons are "considerations that count in favour of" something (Scanlon, 1998, p. 17), the fact that a norm violation is salient often is a good reason to think it is particularly important to address.[14] The fact that a norm is widely endorsed is also often a good reason to think it is worthwhile, or serves some purpose. The fact that others are outraged is a good reason to think

[14] Cf. Levy's (2019) argument that dispositions to e.g. favour the first listed name on a voting form are rational, even though listed order does not correlate with candidate quality, because people often list their implicit recommendations first, and it is generally rational to favour other people's recommendations.

that responding to this norm violation is more likely to generate some effect than one to which others are indifferent. To be sure, there are many exceptions to these characterisations, but it seems that in our assessments of objectionableness, we need to give some weight to people's dispositions given the range of circumstances they find themselves in, not merely focus on particular instances where those dispositions fail to respond as we'd hoped. The considerations which determine whether someone has objectionable motivations shouldn't only include whether particular token wrongs have features the agent is cognizant of. They should also include the ways our dispositions function more broadly, and the benefits they provide, across a range of circumstances.

None of these considerations are decisive, and it is difficult to know the extent to which they apply to any given instance of selective outrage. But we do think they significantly lessen the impetus to call out selective outrage simply on the basis that N1 and N2 are relevantly similar and S1 can't point to a clear difference.

As some conciliation, we think it is possible to give people who find selective outrage objectionable what they want, in a way that doesn't risk disincentivising its benefits. The cases where selective outrage seems the most objectionable are cases with quite good reasons to be outraged about N2 independent of N1. Recall that in *Terrorists*, the point of objecting was to get people to care about the terrorist act in Turkey. But terrorist attacks are something that there is plenty of reason to be outraged about, which most people would recognise upon reflection, and which could be prompted without needing to emphasise S1's selectivity. To put it another way, using S1's response to N1 as grounds for criticism of their response to N2, *in order* to spur a different response to N2, does not seem significantly more likely to succeed compared to simply

emphasising the outrageousness of N2 (though it might be worthwhile to point to S1's response to N1 as *evidence* about the outrageousness of N2).[15]

7. Conclusion

When we see people exhibiting selective outrage, it can be tempting to attribute criticisable motives to them, and criticise them on that basis. However, many instances of selective outrage can lead to important benefits. While these benefits don't entail that people's motivations are as sensitive to moral reasons as we'd sometimes like, they are something we need to keep in mind when it comes to thinking about our practices of criticising, objecting, and expressing outrage.

Constantly objecting to selective outrage risks undermining these benefits. However, abstaining from objecting to selective outrage entirely risks allowing people with criticisable motives to use their outrage for more self-serving ends. To avoid either of these two extremes, our recommendations for responding to what seems to be selective outrage are as follows: if our concern is for norms, then we can simply ask what expressions are warranted on each norm's merits. Ask whether N1 is a norm worth upholding (at this time, in this way, to this degree etc.). If it is, then we should not dissuade S1 from objecting to violations of it. If N2 is worth upholding, then one should encourage S1 to feel outraged about it on that case's own merits.

[15] One last factor bearing on objectionableness is worth noting. We have said that norms are upheld by agents in communities. But there is an important question of which is the relevant community and how to determine its boundaries. On the one hand, many people find it plausible to think that all agents are members of the moral community, and our responses ought to be sensitive to this. At least, we should not draw the boundaries of our communities in objectionable ways; a leading explanation for why some Western citizens expressed more outrage to the terrorists in France, for instance, is that French citizens were seen as having more 'culturally proximity' than Turkish citizens. On the other hand, norms are upheld by agents who actually interact with one another (or have systems which are more or less effective at doing so), and it is simply infeasible to expect the entire moral community to coordinate in the way that e.g. local towns do. We suspect people's intuitions on this point will be influenced by how (un)sympathetic they are to cosmopolitanism or its objections. We did not have the space to engage with this very important question and so had to set it aside, but we acknowledge that in many cases this factor will justifiably underlie many S2's objections to S1.

If, in contrast, our concern is inherently for the agent's motivations, then we can ask whether there is strong independent evidence that the agent has objectionable motives. If there is (and the risks of reducing the enforcement of N1 are outweighed by the importance of objecting) then object to S1 on *that* basis.[16] If there is not any independent evidence, then do not object.[17]

This should provide a lot of agreement between those who are against selective outrage, and those who, like us, are wary of undermining valuable norm-enforcement. The primary disagreement concerns cases where the selectivity is the primary evidence of criticisable motives. We think this is as things should be: we are prone to inferring our political opponents have criticisable motives, and requiring people to have more substantive evidence can help correct this tendency. This need not come at the cost of letting many other violations of valuable norms slide, because there are independent grounds on which to encourage everyone to respond.

[16] To use *Holiday* as an example, we can point out that Hollie is hypocritical by simply citing the fact that she has adequate opportunity to live up to her pronouncements (e.g. holiday locally), but does not take them. We do not need to include any reference to the fact that she had outrage, which was selective, to make our point.

[17] The sentiments in this paragraph echo some of Dover's (2019) conclusions about how to think about hypocritical criticism, given she argues that we should not have strong norms against it. However, her path to these recommendations are considerably different; her argument is that criticism is not punitive, or educative, but dialogical, and that once we control for various kinds of criticisable conduct that are comorbid with hypocrisy, there is nothing left to explain. We, in contrast, maintain that outrage is (or tends to lead to treatments that are) somewhat punitive, and make the positive claim that selective forms of it provide distinct benefits.

References

Amit, E., & Greene, J. D. (2012). You see, the ends don't justify the means: Visual imagery and moral judgment. *Psychological Science, 23*(8), 861-868.

Batson, C. D., Kennedy, C. L., Nord, L. A., Stocks, E. L., Fleming, D. Y. A., Marzette, C. M., ... & Zerger, T. (2007). Anger at unfairness: Is it moral outrage?. *European Journal of Social Psychology*, *37*(6), 1272-1285.

Bell, M. (2013). The Standing to Blame: A Critique. In J. D. Coates & N. A. Tognazzini (Eds.), *Blame: Its Nature and Norms* (pp. 141–161). New York: Oxford University Press.

Brady, W. J., Gantman, A. P., & Van Bavel, J. J. (2020). Attentional capture helps explain why moral and emotional content go viral. *Journal of Experimental Psychology: General, 149*(4), 746.

Caplan, B. (2017). The Unbearable Arbitrariness of Deploring. Retrieved from https://www.econlib.org/archives/2017/12/the_unbearable.html

Case, S. (2017, 21 September). The Rage Against Selective Outrage. *Quillette*. Retrieved from https://quillette.com/2017/09/21/rage-selective-outrage/

Chudek, M., & Henrich, J. (2011). Culture–gene coevolution, norm-psychology and the emergence of human prosociality. *Trends in cognitive sciences, 15*(5), 218-226.

Chudek, M., Zhao, W., & Henrich, J. (2013). Culture-gene coevolution, large-scale cooperation, and the shaping of human social psychology.

Curry, O. S., Mullins, D. A., & Whitehouse, H. (2019). Is it good to cooperate? Testing the theory of morality-as-cooperation in 60 societies. *Current Anthropology, 60*(1), 47-69.

Dover, D. (2019). The Walk and the Talk. *Philosophical Review, 128*(4), 387-422.

Effron, D. A. (2022). The moral repetition effect: Bad deeds seem less unethical when repeatedly encountered. *Journal of Experimental Psychology: General, 151*(10), 2562.

Fehr, E., & Gachter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review, 90*(4), 980-994.

Fritz, K. G., & Miller, D. (2018). Hypocrisy and the Standing to Blame. *Pacific Philosophical Quarterly, 99*(1), 118–139.

Gaus, G. (2011). Retributive Justice and Social Cooperation. in *Retributivism: Essays on Theory and Practice*, ed. White, Mark D. Oxford: Oxford University Press.

Gavrilets, S., & Richerson, P. J. (2017). Collective action and the evolution of social norm internalization. *Proceedings of the National Academy of Sciences, 114*(23), 6068-6073.

Gelfand, M. J., Harrington, J. R., & Jackson, J. C. (2017). The strength of social norms across human groups. *Perspectives on Psychological Science, 12*(5), 800-809.

Gillepsie, E. (2011, 19 April). Hypocrisy of champagne environmentalists is deceitful and distracting. *The Guardian*. Retrieved from https://www.theguardian.com/environment/green-living-blog/2011/apr/19/champagne-environmentalists-damaging-climate-change

Ginther, M. R., Hartsough, L. E., & Marois, R. (2022). Moral outrage drives the interaction of harm and culpable intent in third-party punishment decisions. *Emotion, 22*(4), 795.

Haidt, J. (2003). The moral emotions. *Handbook of Affective Sciences*, 11, 852-870.

Henrich, J., & Ensminger, J. (2014). Theoretical foundations: The coevolution of social norms, intrinsic motivation, markets, and the institutions of complex societies.

Hechler, S., & Kessler, T. (2018). On the difference between moral outrage and empathic anger: Anger about wrongful deeds or harmful consequences. *Journal of Experimental Social Psychology*, *76*, 270-282.

Levs, J. (2013, 28 August). Syria 'Red line' debate: Are chemical weapons in Syria worse than conventional attacks? *CNN*.

Levy, N. (2019). Nudge, nudge, wink, wink: Nudging is giving reasons. *Ergo, 6*.

Marton-Alper, I., Sobeh, A., & Shamay-Tsoory, S. (2022). The effects of individual moral inclinations on group moral conformity. *Current Research in Behavioral Sciences, 3*, 100078.

McMahan, J. (2016). Philosophical Critiques of Effective Altruism. *The Philosophers' Magazine* 73:92-99.

Molho, C., Tybur, J. M., Güler, E., Balliet, D., & Hofmann, W. (2017). Disgust and anger relate to different aggressive responses to moral violations. *Psychological science*, *28*(5), 609-619.

Nguyen, C. T. & Williams, B. (2020). Moral outrage porn. *Journal of Ethics and Social Philosophy* 18 (2):147-72.

Nikiforakis, N., Noussair, C. N., & Wilkening, T. (2012). Normative conflict and feuds: The limits of self-enforcement. *Journal of Public Economics, 96*(9-10), 797-807.

O'Brien, M., & Whelan, A. (forthcoming). Hypocrisy in Politics. *Ergo: An Open Access Journal of Philosophy*.

Perc, M., & Szolnoki, A. (2012). Self-organization of punishment in structured populations. *New Journal of Physics, 14*(4), 043013.

Piovarchy, A. (forthcoming). Does Being 'Bad Feminist' Make Me a Hypocrite? Politics, Commitments and Moral Consistency. *Philosophical Studies*:1-22.

Piovarchy, A. (2020). Hypocrisy, Standing to Blame and Second-Personal Authority. *Pacific Philosophical Quarterly, 101*(4), 603-627.

Piovarchy, Adam (2023). Situationism, subjunctive hypocrisy and standing to blame. *Inquiry: An Interdisciplinary Journal of Philosophy* 66 (4):514-538.

Piovarchy, A. (manuscript). *Signalling, Sanctioning, and Sensitising: How to Uphold Norms With Blame*.

Rossi, B. (2020). Hypocrisy is Vicious, Value-Expressing Inconsistency. *The Journal of Ethics, 25*(1), 57-80.

Royzman, E., Atanasov, P., Landy, J. F., Parks, A., & Gepty, A. (2014). CAD or MAD? Anger (not disgust) as the predominant response to pathogen-free violations of the divinity code. *Emotion*, *14*(5), 892.

Sandis, C. (2015). Verbal Reports and 'Real' Reasons: Confabulation and Conflation. *Ethical Theory and Moral Practice, 18*(2), 267–280.

Saul, B. (2015, 24 November). Solidarity after Paris means being more attentive to suffering elsewhere. *The Conversation*. Retrieved from https://theconversation.com/solidarity-after-paris-means-being-more-attentive-to-suffering-elsewhere-51108

Scanlon, T. M. (1998). *What We Owe to Each Other*. Cambridge, Massachusetts: Belknap Press of Harvard University Press.

Shoemaker, D. & Vargas, M. (2021). Moral torch fishing: A signaling theory of blame. *Noûs* 55(3), 581–602.

Singer, P. (1972). Famine, affluence, and morality. *Philosophy and Public Affairs, 1*(3), 229–243.

Singer, P. (2015). *The most good you can do*: Yale University Press.

Sobel, D. (2007). The impotence of the demandingness objection. *Philosophers' Imprint, 7*, 1-17.

Hirji, S. (2022). Outrage and the Bounds of Empathy. *Philosophers' Imprint* 22 (16).

Telech, D., & Tierney, H. (2019). The Comparative Nonarbitrariness Norm of Blame. *Journal of Ethics and Social Philosophy, 16*(1).

Thulin, E. W. & Bicchieri, C. (2016). I'm so angry I could help you: Moral outrage as a driver of victim compensation. *Social Philosophy and Policy* 32 (2):146-160.

Tierney, H. (2021). Guilty Confessions. In D. Shoemaker (Ed.), *Oxford Studies in Agency and Responsibility* (Vol. 7). New York: Oxford University Press.

Todd, P. (2019). A Unified Account of the Moral Standing to Blame. *Noûs, 53*(2), 347–374. doi:10.1111/nous.12215

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

Traulsen, A., Röhl, T., & Milinski, M. (2012). An economic experiment reveals that humans prefer pool punishment to maintain the commons. *Proceedings of the Royal Society B: Biological Sciences, 279*(1743), 3716-3721.