

Public Justification and the Reactive Attitudes*

Anthony Taylor, University of Oxford

Public justification requires, on a classic formulation, that the exercise of political power be acceptable from the point of view of each reasonable citizen. Numerous political philosophers have endorsed some version of this requirement, such as Charles Larmore, Thomas Nagel and John Rawls. While Rawls's view is no doubt the most discussed, Gerald Gaus has developed a sophisticated and interesting alternative. He defends the

Basic principle of public justification: A moral imperative " ϕ !" in context C , based on rule L , is an authoritative requirement of social morality only if each normal moral agent has sufficient reasons to (a) internalize rule L , (b) hold that L requires ϕ -type acts in circumstances C .² (2011: 263)

In order to know when this principle is satisfied, we need to know when we have sufficient reasons to internalise rules and hold that they require particular acts.³ Here Gaus defends

Having a reason: An agent has a sufficient reason R to hold that β is the thing to believe, or ϕ is the thing to do, if, and only if, (i) they have arrived at R by following the norms of good reasoning and (ii) if they engaged in a "respectable amount" of reasoning, they would not (or did not) discover defeaters for R . (2011: 249–250)

* This is the Accepted Manuscript version of an article published in *Politics, Philosophy & Economics* (published version: <https://doi.org/10.1177%2F1470594X17695070>). For their comments on previous versions of this article, I am very grateful to Paul Billingham, Simon Caney, Will Dahlgreen, Steve Hood, Joe Horton, Sam Kiss, David Miller, Asbjørn Schmidt, Tom Sinclair, two anonymous reviewers and the Editor of this journal.

² This formulation omits: "(c) moral agents generally conform to L ," which raises issues that I cannot consider here.

³ Internalise is used here as a technical term, defined by Allan Gibbard as "having a motivational tendency of a particular kind to act" on the pattern of behaviour prescribed by the norm, where the kind of tendency in question is related to "a purpose of coordination" (1993: 68–71).

Taking these claims together, we have an account of the justification of social-moral rules that takes seriously the evaluative standards of each individual. Social-moral rules are those “that require or prohibit action, and so ground moral imperatives that we direct to each other to engage in, or refrain from, certain lines of conduct.” (2011: 2) In order for such a rule to be one others can legitimately demand my compliance with, it must cohere with my perspective in the right way—it must be one I have sufficient reason to endorse.

Why should we accept the principle of public justification? The details of Gaus’s answer to this question will occupy us for the remainder of the paper, but in brief he argues as follows. As moral persons we are liable to feel the reactive attitudes of resentment and indignation toward those who we perceive to violate moral rules, and these attitudes have appropriateness conditions. Suppose that I believe you have intentionally charged into me causing me to spill my cup of coffee, and I feel resentment on account of your lack of regard for me. If I were to learn that you were in fact caught in a strong gust of wind, my resentment would no longer be appropriate. The object of my attitude, your lack of regard for me, is no longer present, and therefore it can no longer be rationally sustained. Gaus argues that the reactive attitudes are appropriate toward ostensible rule-violators only if they had sufficient reason to internalise the rule in question (2011: 218–224). He then defends the claim, following P.F. Strawson, that our propensity to the reactive attitudes is so deep a part of us that it sits beyond the reach of rational justification. Taking these claims together, he can reach the conclusion that there is a principle of public justification embedded in our social practice of issuing imperatives and following social-moral rules (192).

In spite of the foundational importance of this argument for Gaus’s defence of public reason liberalism, it has yet to receive sustained critical attention in the literature. In this paper I evaluate and ultimately reject this, as I shall call it, *reactive attitudes argument* for public justification. I contend that—even granting some generous

assumptions to his strategy—this argument fails. While, for all I argue here, there may be more to be said for the principle of public justification, I do not believe it can enjoy the kind of justification Gaus claims for it. It is not entailed by our commitment to rationally grounded reactive attitudes.

I proceed as follows. In section 1 I set out the reactive attitudes argument, paying particular attention to the nature of the commitment to public justification it aims to establish. Section 2 takes a closer look at Gaus's account of what it means to have a reason. Sections 3 and 4 examine the premises of the argument: the appropriateness conditions that Gaus takes to govern the reactive attitudes, and the Strawsonian strategy that seeks to establish the commitment in question. Section 5 sets out a potential alternative account of the appropriateness conditions for the reactive attitudes that, I argue, shows the scope of the Strawsonian strategy to be circumscribed. Finally, in section 6, I reconsider our commitment to public justification in light of this, arguing that that the strategy can no longer do the work that the reactive attitudes argument requires of it.

§1. The Reactive Attitudes Argument

The phrase 'reactive attitudes' is a term of art introduced by P.F. Strawson. These attitudes, Strawson tells us, are those that are responsive to our perception of the qualities of will that others express toward us as manifested in their actions or omissions (2008 [1962]: 5–7). In addition, these attitudes have 'vicarious analogues,' namely, the attitudes we adopt in response to our perception of the quality of another's will toward *others* (15). Resentment is a paradigmatic instance of a reactive attitude. When I believe you have intentionally charged into to me causing me to spill my cup of coffee, my resentment is a response to my perception that your action expresses ill will toward me. When I learn that you were actually caught in a strong gust of wind, given that gusts of wind are not willed by anyone, I learn that my resentment is no longer appropriate, as my belief that you expressed ill will toward

me is false. Perhaps some other attitude such as frustration would be appropriate, but not the distinctly *reactive* attitude of resentment.

There is some disagreement both about the range of attitudes that are reactive in Strawson's sense, and about what hangs them together as a class.⁴ However, settling this issue is not essential for the reactive attitudes argument, which can be reconstructed as follows.

- (1) The reactive attitudes of resentment and indignation are appropriate only toward rule-violators who are i) capable of caring for moral rules, and ii) have sufficient reason to internalise the applicable moral rule.
- (2) Authoritative moral imperatives are appropriately addressed only to those to whom the reactive attitudes would be appropriate.
- (3) Therefore, authoritative moral imperatives are appropriately addressed only to those who i) are capable of caring about moral rules, and ii) have sufficient reason to internalise the applicable moral rule (Gaus, 2011: 205–258).

The first premise of the argument provides an account of when resentment and indignation are appropriate. As we saw in the spilt coffee example, certain beliefs are necessary to rationally sustain these attitudes. Discerning when resentment and indignation are appropriate is a matter of working out what the propositional content of these attitudes is. That is, answering the question: what must we believe about their object in order to sustain the attitude? P1 expresses Gaus's answer to this question.

I will discuss P1 in more detail in §3. Setting it aside for now, the controversial claim here is P2. Why should we believe that the fact that our reactive attitudes presuppose certain claims is sufficient to establish that those claims are true? After all, an objector

⁴ See, for example (O'Neill, 2008: 77-83) and (Wallace, 1996: chapter 2)

will be inclined to think, we can reflect on our attitudes and modify them—or perhaps even reject them completely—if we come to conclude they are unjustified. However, Gaus argues, following a strategy originally developed by Strawson, that rational justification of these attitudes is beside the point. The following passage most clearly expresses that strategy.

We are embedded in certain sorts of practices, with certain beliefs and emotions. They form part of the reasons from which we must judge, criticize, and propose changes. A practice such as social morality is deeply embedded in our view of the world; it affects our understanding of interpersonal relations, including love and friendship, and so of what sort of life is worth living. If the presuppositions of our moral practices are so deep a part of the way we see the world, then to renounce the practice would be to renounce most of what we care for and value. But how can we have reason to do *that*? How could we survey all that matters to us and come to the conclusion that our reasons lead us to give it up, by renouncing the view of the world on which our reasons depends? Where would *that* reason come from? It is, I think, as difficult to argue a moral person out of her moral practices as it would be to argue the psychopath into them; given who they are, they do not have reasons to change their view of the world (2011: 192).

The inevitability of resentment and indignation for those bound up in the practice of experiencing them leaves them, in a sense, outside the reach of rational justification. If that is true, and we agree with Gaus about what the appropriateness conditions for these attitudes are, then we reach the conclusion. What is important here is the kind of conclusion we reach, and thus the kind of justification Gaus can say the principle of public justification enjoys. It is not “an exogenous (external) demand on an acceptable social morality based on some foundational moral intuition but a deep presupposition of our social morality with rational reactive attitudes” (2011: 223).

On introducing the conclusion that moral prescriptions are appropriately addressed only toward those who are capable of caring about moral rules and have sufficient reasons to endorse them, Gaus refers to it as *The Principle of Moral Autonomy* (2001: 211). Its relevance for public justification should be relatively clear. If imperatives that

demand we follow particular social moral rules should only be addressed toward those who have sufficient reason to internalise the rule in question, then we will want to know what rules (within the relevant jurisdiction) everyone has sufficient reason to internalise. This is what the principle of public justification tells us.

What does it mean to say that public justification is a deep presupposition of our social morality with rational reactive attitudes? Clearly Gaus does not think the principle of public justification is a conclusion we reach after extensive moral reflection, considering its implications and its coherence with our considered moral judgements about particular cases. Rather, it seems, he thinks it has normative force in the following way. We are, in virtue of the kind of beings we are—*moral persons*—committed to a certain kind of view of the world, including a certain framework of attitudes. To give up this framework is, to import some language from Strawson, practically inconceivable for us. Since, as P1 tells us, a presupposition of this framework of attitudes is *The Principle of Moral Autonomy*, we are committed to the idea of a publicly justified social morality in virtue of our commitment to this framework. It is the fact that our commitment to this framework is not something we can rationally reject that gives the conclusion its normative force.

There is evidence that confirms this understanding in the following passage from the preface to Gaus's earlier *Justificatory Liberalism* (1996):

We understand ourselves and others as capable of putting aside personal valuings, and of acting on norms that can be justified to all. To rid ourselves of this conception [...] would undermine our understandings of ourselves, and others and social life, leaving us without rational grounding for most of what we hold dear. *Given who we are, [...] we are committed to the idea of a publicly justified morality* (vii).⁵

⁵ This quote summarizes an earlier formulation of the reactive attitudes argument from Gaus's *Value and Justification* (1990).

This completes my account of the reactive attitudes argument. It is undoubtedly an interesting argument for a principle of public justification, and considering whether it is successful will require taking a closer look at what Gaus and Strawson have to say about this strategy for defending P2, which I do in §4. First, however, I will briefly consider the idea of *Having a reason* (§2) and what Gaus says in support of P1 (§3).

§2. Having Reasons

Part of what makes Gaus's conception of public justification distinctive is his account of what it means to have a reason. Gaus presents us with an internalist, as opposed to an externalist, account of that relation. An externalist account would say that you have a reason *R* if and only if *R* is an external reason that applies to you (2011: 233). Gaus rejects such views in the following passage:

I believe the Externalist View of Having a Reason to be implausible: it misconstrues the relation between having a reason and being a rational agent. First consider the Externalist View of Having a Reason applied to theoretical reasoning. It implies that Aristotle, when writing on physics, had – possessed – a reason to embrace particle physics, because particle physics is true. But surely he did not have any such reason; to see *R* as a reason is to see it as justificatory, but Aristotle simply could not employ his rationality in a way that could lead him to see the facts supporting particle physics as justificatory (he could not even understand these facts). Only by not following the conclusions of his rational deliberation – being irrational – could Aristotle endorse such a “reason,” and he could never see it as justifying a belief (233–234).

What does it mean to have a reason, then? We might suppose we stand in this having-relation to a reason *R* when it is a reason that applies to us, and we can be said to have or possess it.⁶ However, consider a case of false belief, such as Bernard Williams's

Gin and tonic: Alice orders a gin and tonic, but unbeknownst to her the bartender hands her a glass of petrol (1981: 102).

⁶ See, for example, (Schroeder, 2008). This is the factoring account of having a reason that Schroeder rejects.

Does Alice have a reason to take a sip in this case? If the glass had contained gin and tonic, we would naturally say that there is a reason for Alice to take a sip. She wants to drink a gin and tonic and it is a gin and tonic, so there is a reason for her to drink it. Further, she might sensibly be said to have this reason, as she quite justifiably believes it is a gin and tonic, given that that is what she ordered. In the case where the glass contains petrol, however, Alice may still have a warranted belief that it is a gin and tonic, but because this belief is false we would no longer say that there *is* a reason to drink it. Therefore, if we think we stand in this having-relation to a reason when it is both a reason that applies to us and one that we can be said to have or possess, we will say Alice does not have a reason to take a sip, as the reason in question does not exist.

On Gaus's view, however, whether Alice has a reason to take a sip does not depend on whether there is such a reason. In deciding whether we have a reason *R*, only our justified beliefs about whether *R* is a reason that applies to us are relevant, not whether there is in fact a reason *R* (2011: 247). In considering *Having a reason*, then, it is important to be clear that what we are talking about is simply our warranted beliefs about the reasons that apply to us. With this in mind, nothing related to public justification follows directly from this having-relation. That an agent has a warranted belief that they should ϕ does not tell us whether we may permissibly stop them, or issue a morally authoritative imperative that they not take a sip. We need an explanation of why our warranted beliefs about our reasons have normative import for such questions. This is what the conclusion of the reactive attitudes argument aims to provide, so let us now take a look at the premises of that argument.

§3. The Appropriateness of Resentment and Indignation

As I noted in §1, we discern when resentment and indignation are appropriate by working out what the propositional content of those attitudes is.⁷ We also saw that Strawson claimed these reactive attitudes are a response to the quality of other's wills toward us, but this suggestion alone will not take us far. The pertinent question is: how should we specify the conditions that need to be met for an agent to count as expressing a relevantly malevolent will?

Gaus answers this question by defending two conditions, which I will now set out in turn. Resentment is inappropriate toward those that are incapable of caring about moral rules when they do not promote their wants, ends, or goals (2011: 211). They may care about rules when they correlate with their ends, such as when they know they can avoid an unpleasant punishment by conforming with the rule, but they lack the ability to understand that a moral demand can be authoritative, that it can require them to set aside what they care about and act in line with it. What they are lacking is in part an affective capacity. Not simply the ability to understand moral rules or reasons and their importance in cooperative social life, but the capacity to care about such rules.

This answer explains part of our everyday practice of holding people morally responsible. One thing we tend to take into account when deciding whether someone is responsible, and thus appropriately the object of resentment or indignation, is whether they possess certain capacities. It is not implausible, *prima facie*, to describe the relevant capacity as the capacity to care about and internalise moral rules. Young children, psychopaths, and gusts of wind do not have this capacity, but the kind of

⁷ Note that saying that these attitudes have propositional content is not to assume moral cognitivism. Rather, it is simply to say that these emotions are *about* something: that they have an object. The claim that resentment and indignation have a form of implicit demand as their object is common in discussions of the moral reactive attitudes. See, for example (Darwall, 2006). However, the claim that there is a strong link between the object of resentment and an action being wrong is controversial.

agents we generally take to be morally responsible do. I will refer to this first condition as the *capacity condition*.

This is not the only condition Gaus claims governs the appropriateness of resentment and indignation. He asks us to imagine we are issuing an imperative to someone unable to grasp its authority, in spite of the fact that they meet the capacity condition.

I know that she does not see that it is the moral thing to do, and suppose I think her lack of appreciation is quite genuine. I demand “ ϕ !” and she does not see why she is obligated to ϕ . She is puzzled that anyone would think ϕ is obligatory. If I think this, then again I cannot reasonably feel resentment or indignation that she fails to ϕ , anymore than I can feel indignation at a four-year-old who is unable to detach himself from what he most wants to do and so steals some favourite candy. She just cannot see how “ ϕ !” has any internal authority over her (2011: 219).

He goes on to argue that if we offered this person a reason to adopt the rule that requires she ϕ s, and she were unable to grasp the force of *that* reason, then resentment would still be inappropriate. She cannot grasp that ϕ -ing is the thing to do, because she cannot grasp my reason for holding that ϕ -ing is the thing to do. Therefore, I cannot rationally resent her failure to act (220). I will refer to this second condition as the *sufficient reason condition*.

On this view, the capacity condition is what we can call a global condition. Where it is not met it tells us that due to some fact about the agent it is not appropriate to hold them responsible *in general*, and therefore that it would be inappropriate to resent them. The sufficient reason condition, by contrast, is what we can call a local condition. It tells us that due to some fact about the relationship between the agent and this particular event, it is not appropriate to hold them responsible for it, and therefore that it would be inappropriate to resent them. If you do not meet the sufficient reason condition with regard to some particular rule, you are still considered to be a responsible moral agent in general, who might be appropriately

resented in other instances. If you fail to meet the capacity condition, however, you are not considered to be a responsible moral agent at all.

Taken together, these two conditions have a certain unity. The explanation of why some people can be the object of resentment and indignation is continuous with the explanation of why some specific actions or omissions can be appropriately resented. The explanatory factor in both the global case and the local case is the idea of having sufficient reason to internalise and comply with social-moral rules. If we accept this picture, the propositional content of these attitudes—*what* we resent, or feel indignant about, when we do so rationally—is the attitude of the violator’s failure to follow a rule that they had sufficient reason to. When we feel indignant because we believe someone has violated a rule, global conditions such as insanity, or being a young child, and local conditions, such as being caught in a strong gust of wind, function to tell us that this attitude was not in fact present, and so our indignation cannot be rationally maintained.

However, it is far from obvious that resentment or indignation are never appropriate responses to those who do not meet the sufficient reason condition with respect to some rule. Intuitively, conscientious SS officers often acted wrongly and were thus appropriately resented, even if they would believe after sufficient amount of good reasoning that any rules prohibiting their acts were ones they did not have sufficient reason to follow.⁸ We might think, therefore, that we should accept an alternative view about when resentment and indignation are appropriate that does not have this implication. This is simply to doubt the normative significance of Gaus’s notion of *Having a reason* for the question of when the reactive attitudes are appropriate.

⁸ I adapt this example from (Parfit, 2011: 158).

One way to respond to this doubt would be to appeal to a deeper explanation of why it is inappropriate to resent those who do not meet the sufficient reason condition. However, as we shall see, Gaus's Strawsonian strategy precludes his providing such an explanation. That strategy is to claim that renunciation of the framework of reactive attitudes is practically inconceivable, and even if it were conceivable, it would be irrational. It tells that these attitudes lie beyond the scope of rational justification—so a deeper explanation of the conditions force is neither necessary nor possible.

§4. The Strawsonian Strategy

The preceding section set out what Gaus says in support of (1)—his account of when resentment and indignation are appropriate. In spite of the doubt I have just raised, for the sake of argument I will assume for the time being that, as a description of our current practice of holding one another responsible with its attendant liability to the reactive attitudes, this account is correct. That is, I will grant Gaus P1 of the reactive attitudes argument.

Attending to the second premise of that argument, recall that it states

- (2) Authoritative moral imperatives are appropriately addressed only to those to whom the reactive attitudes would be appropriate.

The Strawsonian strategy that supports this premise is a much-discussed argument in debates surrounding the compatibility of moral responsibility and determinism. The stated aim of Strawson's "Freedom and Resentment," the paper that introduced this strategy, is reconciliation between the positions of the "optimist" and the "pessimist" (2008 [1962]: 1–2). The pessimist holds that the truth of determinism would undermine our everyday practice of holding people morally responsible. Here is an example of how that practice might be undermined. One explanation as to why we let people off the moral responsibility hook, and thus do not feel resentment on account of their actions or omissions that would otherwise be resentment-worthy is that they could

not have acted otherwise. On this explanation, known as the principle of alternate possibilities, the fact that you were caught in a strong gust of wind gets you off the moral responsibility hook because it tells me you could not have acted other than to make me spill my coffee (Frankfurt, 1969: 829). However, if this were the best explanation of why you ought to be let off the hook in this instance, the truth of determinism would render the practice of moral responsibility unjustified across the board. That determinism is true tells me that *everyone* is relevantly like you, unable to do otherwise.⁹ Strawson's optimist, by contrast, holds that our practice of holding people morally responsible is justified by its effectiveness at regulating social behaviour in ways that produce good consequences. According to this optimist then, the truth of determinism does not show our practice to be unjustified, since it does not tell us that our practice has negative consequences.

Strawson aims to reconcile these positions by showing they are both wrong about the grounds of the practice of moral responsibility. On his view, while the grounds of the practice are not undermined by the truth of determinism as the pessimist thinks, the practice also is not justified by its good consequences, as the optimist thinks. His argument rests on an account of the reactive attitudes, and the conditions under which we consider it appropriate to suspend them.

One instance where we deem it appropriate to suspend the reactive attitudes, Strawson tells us, is where we learn that we were mistaken about the quality of will in question (2008 [1962]: 7–8). Again, the spilt coffee case is an example of this. When I learn that you were caught in a strong gust of wind, I learn that I was mistaken to hold that your spilling of my coffee expressed ill will toward me—it expressed no

⁹ Note that I do not say what the thesis of determinism says here, as Strawson's argument proceeds without a precise definition. He instead proceeds on the basis that "if there is a coherent thesis of determinism, then there must be a sense of 'determined' such that, if that thesis is true, then all behavior whatever is determined in that sense." (2008 [1962]: 11)

such will at all. This is what I called a local condition above. A second instance where we deem it appropriate to suspend the reactive attitudes is where we learn that it would be a mistake to resent this person due to a more general fact about them—their insanity, for example, or their being a young child (8 – 9). This is what I called a global condition above.

Finally, beyond these two kinds of case, Strawson notes a further kind. We can, it seems, suspend the reactive attitudes not because, for one of the two preceding reasons, we believe it would be *inappropriate* to cling on to them, but rather because we have the ability to suspend them, which we can employ “as a refuge, say, from the strains of involvement; or as an aid to policy; or simply out of intellectual curiosity” (10).

Presented with this account of our practice of moral responsibility we will want to know what role it can play in answering the normative question of when we *should* suspend the reactive attitudes. After all, Strawson’s pessimist presumably holds that regardless of when we *in fact* suspend the reactive attitudes, we ought to do so permanently, across the board, as the truth of determinism renders them unjustified.

To see Strawson’s answer to this question we should first note that he holds that the alternative to the reactive attitudes is what he calls objectivity of attitude.

To adopt the objective attitude to another human being is to see him, perhaps, as an object of social policy; as a subject for what, in a wide range of sense, might be called treatment; as something certainly to be taken account, perhaps precautionary account, of; to be managed or handled or cured or trained; perhaps simply to be avoided, though this gerundive is not peculiar to cases of objectivity of attitude (2008 [1962]: 9).

Importantly, this objectivity of attitude is fundamentally opposed to human interpersonal relations. To be involved in such relations precisely is, Strawson tells us,

to be liable to the reactive attitudes (10). The problem, then, is that that the pessimistic view just described would recommend we take up the objective stance permanently, as it tells us that our practice of holding people morally responsible is *always* unjustified. In taking up this stance permanently we would thus be sacrificing all interpersonal relations, and it is this that is

for us as we are, *practically inconceivable*. The human commitment to participation in ordinary inter-personal relationships is, I think, too thoroughgoing and deeply rooted for us to take seriously the thought that a general theoretical conviction might so change our world that, in it, there were no longer any such things as inter-personal relationships as we normally understand them; and being involved in inter-personal relationships as we normally understand them precisely is being exposed to the range of reactive attitudes and feelings that is in question (12).

Moreover, Strawson argues, even if we could conceive of having a choice in this matter, it would be irrational to choose to permanently adopt the objective stance. In making such a choice “we could choose rationally only in the light of an assessment of the gains or losses to human life; its enrichment or impoverishment; and the truth or falsity of determinism would not bear on the rationality of *this* choice” (14). If he is right that the decision must be made on the basis of these considerations, then, taking into account the claim that the objective stance is fundamentally opposed to interpersonal relations, it is only a short step to the conclusion that permanent adoption of the objective stance is irrational.

So Strawson’s answer to the question of what bearing an account of when we in fact suspend the reactive attitudes has on the normative question of when we ought to suspend those attitudes turns on what I will call the

Commitment claim: Given our deep commitment to the reactive attitudes, it is practically inconceivable that we could abandon them; and further, if we could abandon them, it would be irrational to do so.

To be sure, this claim raises a number of further questions that would need to be answered in order to establish that it can provide a satisfactory answer to the normative question. As I noted above, this argument is now the subject of a large literature in debates about moral responsibility. I will not be able to offer anything close to an adequate analysis of it with respect to those debates here. What I will do instead is grant, as a second assumption, that the commitment claim is true. Moreover, I want to grant that it provides a satisfactory answer to the pessimist's normative question. We *ought not* renounce the reactive attitudes in light of the truth of determinism, because our doing so is practically inconceivable in light of our commitment to interpersonal relationships. This is a controversial assumption. However, I will argue in the remainder of the paper that even with it in place Gaus's argument for the principle of public justification is unsuccessful. Those who find the Strawsonian approach wanting more generally will simply have a further reason to reject that argument.

However, in granting the commitment claim as an assumption, I want to keep in the fore why Strawson thinks this claim is true. It is because he holds that being involved in interpersonal relationships *precisely is* being liable to the reactive attitudes that he holds it to be practically inconceivable for us to give them up.

§5. An Alternative

With this thought in mind, I will set out an alternative view of when the reactive attitudes are appropriate. Consideration of this alternative will, I believe, show that the scope of the commitment claim is limited.

As we have seen from the discussion so far, there are numerous cases where someone meets the capacity condition and ostensibly violates a moral rule, and yet we are inclined to think that it is not appropriate to feel resentment or indignation toward them, as in the split coffee example. As I noted above, if we think that the underlying

principle that tells us when resentment or indignation are appropriate is the principle of alternate possibilities, then the compatibility of moral responsibility and determinism will be threatened. Since the truth of determinism will tell us that none of us could ever have acted otherwise, the principle will tell us that no one is ever the appropriate target for resentment or indignation.

An alternative principle, expounded by R.J. Wallace, is that of “no blameworthiness without fault” (1994: 135). On his view, our resentment or indignation can be rationally sustained where we believe a person has violated a moral obligation that we accept and hold them to. At first glance this may seem to have a limited ability to account for local conditions, because that I accept and hold you to a moral obligation seems like it would be unaffected by facts about your relationship to a particular act or omission. That is, it seems that if you are under a moral obligation to refrain from violating my bodily integrity by knocking into me, for example, the fact that you were caught in a strong gust of wind will be irrelevant to whether or not you have violated this obligation. However, Wallace suggests that because holding someone to an obligation we accept involves a commitment to the existence of reasons that support that obligation, moral obligations must be focused on states of affairs that are directly susceptible to the influence of reasons. Therefore, an act that was not potentially susceptible to the influence of reasons cannot constitute the violation of moral obligation, and thus cannot make a person the object of rationally appropriate resentment (118–147). This insight shows that the view can account for local conditions. When I learn you were caught in the gust of wind, what I learn is that your bodily movement was not the kind of thing that is susceptible to the influence of reasons, and therefore that you did not violate a moral obligation at all. Further, Wallace also supports a condition much like Gaus’s capacity condition. It is not appropriate to hold someone responsible in general if they do not have the power of reflective self-control. That is, the power to grasp and apply moral reasons and regulate their behaviour in light of them (154–155). The underlying normative

principle here, then, is that it is unfair to hold those who lack such powers responsible. Cases where we tend to let those with such powers off the hook are all cases where the agent turns out not to have performed that action intentionally at all. Importantly for Wallace, the compatibility of moral responsibility with determinism is not threatened on this view, as even in a deterministic world we would have no reason to think that no one ever has the power of reflective self-control.

I will call Wallace's alternative condition governing when we ought to let the generally responsible off the hook the *susceptibility to reason condition*. It should be clear that it is quite different from the sufficient reason condition. Provided Alf meets the capacity condition, he can act on his warranted beliefs about the reasons he has and still violate a moral obligation Betty holds him to, making her resentment appropriate. That is, even though he believes after a sufficient amount of good reasoning that he has a moral permission to ϕ , if he ϕ 's, where ϕ is both an act susceptible to the influence of reasons, and an act Betty holds him to an obligation not to perform, her resentment is rationally appropriate. On this view the propositional content of resentment is the belief that someone has violated an obligation we hold them to, where that requires, necessarily, that their act or omission was susceptible to the influence of reasons.

I have assumed that Gaus's account of when the reactive attitudes are appropriate in our current practice of moral responsibility (1) is correct. I have granted, then, that within this practice local conditions are governed by the sufficient reason condition, not the susceptibility to reason condition. But supposing that is right, and we are currently participants in a practice governed in part by the sufficient reason condition, might we abandon that condition in favour of the susceptibility to reason condition? That would involve ceasing to suspend our resentment or indignation in the cases in which Gaus thinks we should, those in which we learn that the person in question did not have sufficient reason to internalise the rule we believed they had violated.

Instead, we would suspend our resentment or indignation only in cases where we learn that the person in question does not meet the capacity condition or their act or omission did not meet the susceptibility to reason condition.

In answering this question about the possibility of modifying our practice it seems that Gaus would want to appeal to the commitment claim. The sufficient reason condition is a presupposition of the reactive attitudes as we find them when examine our current practice. Given that the reactive attitudes are constitutive of interpersonal relationships, and we are committed to such relationships, we are committed to the reactive attitudes *as we find them*. However, this answer to our question is unsatisfactory. While it follows from our commitment to interpersonal relationships that wholesale repudiation of the reactive attitudes is practically inconceivable; it does not clearly follow that revising the conditions under which we deem it appropriate to suspend them is practically inconceivable. To defend that claim we would need to argue that this revision would amount to sacrificing all interpersonal relations, as Strawson thinks would be the case with permanent adoption of the objective stance.

Is there any reason to suppose that revising the conditions under which we deem it appropriate to suspend the reactive attitudes would have such severe consequences? Gaus appears to think this is the case. As I pointed out above, he tells us that to reject the understanding of ourselves and others as capable of acting on rules that can be publicly justified would be to “undermine our understandings of ourselves, and others and social life, leaving us without rational grounding for most of what we hold dear” (1996: vii). Further, he tells us: “If the presuppositions of our moral practices are so deep a part of the way we see the world, then to renounce the practice would be to renounce most of what we care for and value” (2011: 192). These claims echo Strawson’s point about changing our social world so that there are no longer such things as the reactive attitudes, and therefore no longer such things as interpersonal

relationships. However, it is difficult to see how, in Gaus's case, these claims could be true.

When the conditions under which we deem it appropriate to suspend the reactive attitudes are governed by the capacity condition and the sufficient reason condition, we have an expectation of others that they follow the rules that they have sufficient reason to internalise and comply with. When we feel indignant on account of their failure to do so, the object of our indignation is a fact about the relationship between their capacities and their act or omission. The object of our indignation is: "This person has the ability to care about and internalise social-moral rules, and they have sufficient reason to internalise and comply with this rule in particular—but they have violated it nonetheless!" By contrast, when the conditions under which we deem it appropriate to suspend the reactive attitudes are governed by the capacity condition and the susceptibility to reason condition, we have an expectation of others that they not violate moral obligations we accept and hold them to. The object our indignation, when we feel indignant on account of their failure to do so, is again a fact about the relationship between their capacities and the act or omission. It is: "This person's action, which violated a moral obligation, was susceptible to the influence of reasons and so can properly be seen to express a judgement on behalf of their object. They have *chosen* to ϕ , and given that ϕ violates an obligation I hold them to, I resent their choice to ϕ ."

Both of these accounts involve demands and expectations. Both have an object that reflects a fact about the relationship between the person's capacities and the act or omission we might take to be worthy of say resentment, indignation, or gratitude. There might be numerous theoretical reasons to prefer one account to the other (or indeed to reject both) either as an account of what our practice is or ought to be. However, what seems clearly false is that Gaus can respond to the question of whether we ought to revise our practice from one to the other by appealing to the commitment

claim to say that we should not, because that change would amount to renouncing all we care for and value. With this in mind we can see that, even granting the commitment claim, its scope is limited—it does not tell us that renouncing the sufficient reason condition in favour of an alternative is practically inconceivable for us.

Before moving on to reconsider the reactive attitudes argument in light of this, it is worth pausing to consider one way in which Gaus might reply to this argument. The susceptibility to reason condition seems to fit more intuitively with an externalist account of reasons and rationality. What makes that condition distinct from Gaus's sufficient reason condition is that it does not take the question of whether resentment is appropriate to hang on whether or not their target had sufficient reason, in his internalist sense, to endorse the rule in question. It takes the fact that there is a moral obligation not to ϕ , alongside the fact that the agent possesses the relevant capacity, as sufficient to ground rational resentment. This only makes sense, we might think, if we take external reasons to be normatively significant, such that an agent is guilty of a rational flaw for failing to respond appropriately to a reason that applies to them, even if they would not see it as a reason after a sufficient amount of good reasoning. Otherwise it would imply, somewhat counterintuitively, that we can rationally resent someone who is guilty of no rational failing at all. Since, as we have seen, Gaus rejects the externalist account of having a reason, we might think that he could thus reject any conditions that draw on externalism on the same grounds.

However, while Gaus rejects externalism as an account of the reasons that we *have*, he explicitly does not deny externalism about the reasons that there *are* (2011: 232–233). We can happily accept that he is right to think that his internalism is the right account of the having-relation. The important question is whether or not that relation is of normative significance. As I noted at the end of §2, Gaus thinks this relation is significant because of the reactive attitudes argument. It is because we only resent

those who have sufficient internal reasons to endorse the rules they have violated—and we are committed to doing so in the sense implied by the commitment claim—that the reasons we have are significant for the question of when resentment is appropriate. Since in presenting the susceptibility to reason condition my aim has been precisely to call into question the nature of this commitment, Gaus cannot respond by simply asserting it.

§6. Reconsidering our Commitment to Public Justification

What are the implications of this argument about the scope of the commitment claim? It shows that the claim does not pick out the attitude of resentment with the sufficient reason condition as one of its appropriateness conditions in particular. Rather, it picks out a set of appropriateness conditions for resentment.¹⁰ However, the reactive attitudes argument cannot support the principle of public justification unless the commitment claim picks out resentment with the sufficient reason condition as one of its appropriateness conditions in particular. To show this I will now reconstruct that argument taking the proper scope of the claim into account.

(1) stated that resentment and indignation are only appropriate toward rule-violators who are (i) capable of caring for moral rules, and (ii) have sufficient reason to internalise the applicable moral rule. Our confidence in this premise will surely drop when we realise that the commitment claim does not provide support for it. We might therefore be inclined to think that (1) is not an accurate description of our practice of holding people responsible. If (1) is false, then Gaus's principle of public justification will not follow from the argument. However, I will now show that the principle does not follow even if we continue to grant this premise as an assumption.

¹⁰ Some may prefer to refer to this alternative as an alternative attitude to resentment, rather than an alternative appropriateness condition for resentment. Nothing in the following argument hangs on this choice, but I refer to it as an alternative appropriateness condition for simplicity.

We can begin by noting that the normative premise (2) must be modified. Originally, it stated that authoritative moral imperatives are appropriately addressed only toward those to whom the reactive attitudes are appropriate. But when we ask why this is true Gaus's response is to appeal to the commitment claim. As I have shown, this claim is only plausible with regard to a set of presuppositions of resentment, and at least one member of this set does not have the sufficient reason condition as a presupposition. The argument for this is simply the argument of the previous section: there is another appropriateness condition, the susceptibility to reason condition, which is capable of playing the role of rationally grounding valuable interpersonal relationships. The availability of such an alternative shows that the commitment claim does not show that the sufficient reason condition is beyond the reach of rational justification. Rather, it shows that it would be practically inconceivable, and irrational if conceivable, to not endorse a member of the set of appropriateness conditions that is capable of furnishing rational grounds for human interpersonal relationships. Moreover, the commitment claim has nothing to say about how we choose from within this set—it only tells us that we cannot rationally choose to not to endorse any member of it.

We must therefore modify (2) to state: authoritative moral imperatives are appropriately addressed only toward those to whom a member of the set of attitudes we cannot rationally reject would be appropriate. The scope of the set of appropriate moral imperatives, according to this premise, is dependent on the scope of the set of appropriateness conditions that we cannot rationally reject. All I have claimed about this latter set is that it contains a single further condition other than the sufficient reason condition. Perhaps further analysis would show that this set contains more conditions, but I will not investigate that possibility further because the existence of a single further condition is sufficient to establish that the reactive attitudes argument for the principle of public justification fails. Once (P2) is modified to refer to a set of

attitudes of which one member does not include the sufficient reason condition as a presupposition, the conclusion no longer follows.

The conclusion, recall, was that authoritative moral imperatives are appropriately addressed only toward those who are (i) capable of caring for moral rules, and (ii) have sufficient reason to internalise the applicable moral rule. This conclusion shows, Gaus claims, that public justification is a “deep presumption of our social morality with rational reactive attitudes” (223). However, the way in which (2) gives this conclusion its normative force in the original argument is, we saw, via the commitment claim. Since that claim does not tell us that we cannot modify the practice of social morality such that the sufficient reason condition is abandoned in favour of the susceptibility to reason condition, it does not give the conclusion its normative force in the modified argument. For all that has been said, we could adopt the susceptibility to reason condition, modify our practice as it recommends, and still enjoy rationally grounded interpersonal relationships without a commitment to a public justified social morality. Note that I need not even argue that we ought to do this. The mere availability of an alternative that is not ruled out by the claim strips the public justification principle of its normative force. On this reconstructed version of the argument, we are not committed to public justification in virtue of being the kind of beings that we are. Rather we are committed to rationally grounded interpersonal relationships. Public justification is one way, among others, that we can furnish rational grounds for these relationships. Therefore, public justification is not, as Gaus claims, a deep presupposition of our social morality with rationally grounded reactive attitudes, something that we cannot rationally reject given the kinds of beings that we are.

Conclusion

I will now summarise the argument I have developed over the course of this paper. Gaus’s argument for the principle of public justification appeals to the

presuppositions of our framework of reactive attitudes. That argument aims to show us that we are committed to public justification by virtue of our commitment to this social-moral framework of attitudes. Some may find Gaus's description of our social-moral practice unconvincing, and still more may find his Strawsonian strategy wanting. However, I have argued that even assuming that Gaus is right on both of these points, the reactive attitudes argument fails. When we properly delineate the scope of the commitment claim we see that our commitment is not to public justification but to interpersonal relationships, for which we can furnish alternative rational grounds. Perhaps other arguments can be marshalled in defence of the principle of public justification, but it is hard to see how they could establish that the principle enjoys the kind of justification Gaus claims for it.

References

- Darwall S (2006) *The Second-person Standpoint: Morality, Respect, and Accountability*. Cambridge, MA: Harvard University Press.
- Frankfurt H (1969) Alternate Possibilities and Moral Responsibility. *The Journal of Philosophy*, 66: 829–839.
- Gaus G (1996) *Justificatory Liberalism: An Essay on Epistemology and Political Theory*. New York: Oxford University Press.
- Gaus G (1990) *Value and Justification: The Foundations of Liberal Theory*. New York: Cambridge University Press.
- Gaus G (2011) *The Order of Public Reason*. New York: Cambridge University Press.
- Gibbard A (1993) *Wise Choices, Apt Feelings: A Theory of Normative Judgement*. Oxford: Clarendon Press.
- O'Neill M (2008) *Freedom, Fairness and Responsibility*. Harvard University PhD Dissertation.
- Strawson PF (2008) *Freedom and Resentment and Other Essays*. Oxford: Routledge.
- Schroeder M (2008) Having Reasons. *Philosophical Studies*, 139: 57–71.

Wallace RJ (1994) *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard University Press.

Williams B (1981) Internal and External Reasons. *Moral Luck: Philosophical Papers 1973-1980*. Cambridge: Cambridge University Press.