

Beyond Excuse and Exemption

P. F. Strawson's "Freedom and Resentment" is deep and insightful. But it is also incomplete. Its central analysis of the pleas that inhibit resentment is incomplete.

Remedying this shortcoming reveals that the Strawsonian account of responsibility is more powerful and flexible than it is usually thought to be. In particular, it is well-equipped to account for our intuitive reactions to people like Robert Alton Harris¹ who commit terrible crimes, but do so in part because of the terrible circumstances in which they were raised.

My discussion thus strengthens the Strawsonian analysis of responsibility. But it also reveals that the central argument of "Freedom and Resentment" is incomplete. Strawson's argument that the truth of determinism could not undermine our responsibility practices relies on his incomplete analysis of resentment-undermining pleas. The expanded analysis I develop here opens the door for the truth of determinism to once again threaten a wide swath of our responsibility practices. Incompatibilists may, in fact, be able to leverage the strength of the Strawsonian framework to motivate their skepticism about our responsibility practices.

I. Four Types of Plea

The argument of "Freedom and Resentment" has three stages. In the first, Strawson identifies a group of familiar, interrelated human practices. Humans participate in engaged interpersonal relationships with each other. It is essential to these relationships that we make demands on the attitudes that other participants bear us.² Reactive attitudes like resentment and

¹ First discussed in Gary Watson's paper "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme" (Watson, 2004).

² I should say that I will use "demand" in a very loose, inclusive way in this paper. Others have used "expectation" to capture the same notion (see, e.g., R. Jay Wallace's discussion in (Wallace, 1998, p. 18 ff)).

indignation arise when these engaged interpersonal demands are violated. In the second stage, Strawson draws on this framework to identify two ways in which a special consideration or plea (e.g., an excuse) might inhibit resentment and other reactive attitudes. Finally, drawing on this analysis, Strawson argues that the truth of determinism could not be an inhibiting plea of either sort.

I focus on the second stage of Strawson's argument, his two-part analysis of the pleas or "special considerations" which "might be expected to modify or mollify this feeling [of resentment] or remove it altogether" (p. 77).

Strawson introduces his first type of consideration by citing expressions like "He didn't mean to" and "He couldn't help it—he was pushed" (p. 77). Such pleas change our beliefs about the attitudes with which someone acted. They show that, despite appearances to the contrary, the agent did not have attitudes that violate the relevant demand. Since resentment is a reaction to a violation of a demand, it will be undermined if it turns out that the demand was not violated. Following the terminology introduced in Gary Watson's influential paper "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme" (Watson, 2004), I will refer to these pleas as Strawsonian *excuses*.

Strawson introduces his second type of consideration by citing expressions like "He's only a child" and "He's a hopeless schizophrenic" (p. 78). Such pleas alter the way we relate to the agent in a profound way: they undermine the demand altogether. Since resentment is a reaction to a violation of a demand, it will be undermined if the demand itself is undermined. Again following Watson, I will refer to these pleas as *exemptions*—a natural term, as they motivate exempting an agent from demands that would otherwise be in place.

It is vital to my aims in this paper to recognize two facts about this analysis. First, Strawson has identified two *mechanisms* by which reactive attitudes might be undermined. And, second, Strawson's two-part account of these mechanisms is grounded in his account of the structure of our moral responsibility practices. Resentment and indignation are reactions to someone's attitudes, insofar as those attitudes *violate a demand*. A response to a violation of a demand will be undermined if it can be shown that the demand was in place but was not in fact violated, or if the demand itself is undermined. Pleas belonging to Strawson's first category show that the demand was not in fact violated; pleas belonging to his second category undermine the demand itself.

This two-type analysis is a common thread in the quality of will literature following Strawson. To name a few prominent examples: it can be found in Jonathan Bennett's influential paper "Accountability" (Bennett, 1979, p. 22); Gary Watson's "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme" (Watson, 2004, p. 224); R. Jay Wallace's *Responsibility and the Moral Sentiments* (Wallace, 1998, p. 119 and 154 ff); and Michael McKenna's recent *Conversation and Responsibility* (McKenna, 2012, p. 75 ff).

But the two-part analysis is too limited. Erin Kelly's recent work on excuses makes that clear. In "What is an Excuse?," Kelly describes an important range of cases:

Examples extend from the ordinary to extraordinary: the parent under emotional strain who yells at her kid, the child who is mistreated by her parent and in turn bullies another child, the inmate who brutalizes another to avoid appearing weak, the soldier in a field of battle who shoots a civilian on orders from a superior, the alcoholic who drives after drinking, the compulsive who tells a lie... The circumstances or psychological vulnerabilities of these agents fit uncomfortably with the presumptions of our blaming attitudes and judgments. (Kelly, 2013, p. 248)

The two-part analysis does not do justice to these cases. To see that, consider the parent under emotional strain. If the parent's emotional strain were an *exemption*, it would render her incapacitated for normal interpersonal relationships, and thereby exempt her from any

interpersonal, blame-related demands. But being under emotional strain does not—not even locally, or temporarily—remove a parent from such relations altogether. *Some* demands still apply to a parent, even when she is under emotional strain.

What about Strawson's other category of pleas, those which prompt us to revise our judgments about the attitudes with which someone acted? Sometimes emotional strain does function as a plea of this sort. If I discover that a parent yelling at her child is under severe emotional strain, I likely will not think that she is doing so out of cold, deliberate malice. Her attitudes may still be problematic in some way, but they are not nearly so problematic as they would otherwise seem.

But even after we have adjusted our understanding of the parent's attitudes in this way, the fact that the parent was under emotional strain seems to have some further relevance to our reactions to those attitudes. To see this, consider one possible elaboration on the parent case. Suppose that the parent does not yell at her child out of cold, deliberate malice; instead—because of her emotional state—she responds with disproportionate irritation and frustration to her child's ordinary, if frustrating, behavior. This parent's disproportionate irritation and frustration is less problematic than cold malice would be. But even taking that into consideration, the fact that the parent is under emotional strain can have some additional impact on our reactions. If she were *not* under emotional strain and responded to her child's behavior with the very same level of disproportionate anger and frustration, our responses would be different.

Being under emotional strain thus has some further impact on our moral responses. Kelly describes one way this further impact might work: the parent's emotional strain might function “by undermining what normally are reasonable expectations about how a person should be motivated” (p. 256). That is not to say that these considerations provide an all-out exemption

from all expectations; rather, they might motivate us to “relax our expectations” (p. 256). (Now, you might think that Strawson can just assimilate these considerations to his “exemption” category anyway—as a sort of partial exemption. I will argue shortly that he cannot.) The fact that the parent was under emotional strain weakens the demands we make of her, without undermining them altogether. The *content* of those demands is altered, so that they are easier to satisfy. Our ordinary reactions to the parent’s actions are undermined because—considering the emotional strain she was under—it would not be fair to demand that she exercise the usual degree of self-control.

The phenomenon Kelly describes involves altering the content of the demand made of the parent under emotional strain, to the point that the parent in fact satisfies the demand. Then our reaction to a violation of a demand is undermined, because the parent in fact satisfies the (weakened) demand that is made of her.

There is one more way in which our response to the parent may be affected. Consideration of the emotional strain a parent is under may leave in place the demand that the parent avoid disproportionate anger and frustration, but alter *how we respond* to the parent’s demand-violating attitudes. Elsewhere I have argued (see Goldman, 2014) that a very wide range of responses, ranging from angry resentment to wholly nonantagonistic responses (for instance, a moralized form of disappointment) are compatible with making a demand within an engaged interpersonal relationship, and seeing that demand violated. Thus consideration of the emotional strain a parent is under might prompt us to respond to violations in a milder, nonantagonistic way, rather than with anger or indignation.

These two additional types of resentment-undermining pleas combine with the two types identified by Strawson, to give us a four-part analysis:

- A. First there are the exemptions that Strawson identifies. Such pleas exempt an agent from engaged interpersonal relationships, and their concomitant demands, altogether. To exempt someone is, as Strawson sometimes puts it, to adopt the objective attitude towards her. (These exemptions can operate permanently, as “She’s a hopeless schizophrenic” might; they can also operate temporarily.) Since these pleas leave no demands in place, there can be no violations of the demand to which we might respond with resentment.
- B. Second there are pleas that shift the *content* of the demand. If the content of the demand is relaxed to the point that the agent actually did satisfy the demand, then there was no violation of the demand to which we might respond with resentment.
- C. Third there are pleas of Strawson’s first type, the Strawsonian excuses. These pleas shift our understanding of the attitudes with which the agent acted. If the agent’s attitudes turn out to have been in line with what was demanded of her, then there was no violation of the demand to which we might respond with resentment.
- D. Fourth, a plea may shift the way one reacts to a violation of a demand. Without removing someone from participation in an engaged interpersonal relationship, without altering the content of the demands made of her, and without altering the fact that she had attitudes that violated the demand, a plea might prompt a shift away from resentful, angry, or antagonistic reactions to that violation

This is an exhaustive analysis: every consideration that bears on how we do or should respond to moral violations falls into one of these four categories.

This analysis is exhaustive, but it is also extremely schematic. The fourth category in particular is a catch-all: any consideration that does not bear on the first three questions but has

an impact on moral response falls into the fourth category. I do not deny that all of these categories are susceptible to much further analysis and investigation.

Despite that, this analysis provides a useful framework for thinking about the ways that moral response may be undermined. That is because it is a *functional* analysis: we have separated out different types of pleas in terms of the *mechanism* by which they undermine the reactive attitudes. Clearly describing these mechanisms makes it easier to understand complex or difficult cases; it also enables us to reason at a more general level about the prospects for undermining or radically revising our existing practices of moral response.

II. Assimilation Objections

Next I want to consider an important objection to my claim that I am expanding Strawson's two-part analysis. Strawson does not discuss type-B considerations separately; but perhaps the categories of plea that he does consider do in fact include them, or can be extended to include them. In discussing type-A exemptions, for instance, Strawson does identify a "far less important" subset of exemptions which operate "temporarily." These include the consideration—reminiscent of Kelly's example of a parent under emotional strain—"He has been under very great strain recently" (p. 78). Perhaps we can understand this subset to involve, not an exemption from *all* demands (even temporarily), but an exemption from *some* demands, while other demands remain in place. On this analysis a difficult-to-satisfy demand is eliminated, and an easier-to-satisfy demand remains. The effect is to shift the content of what is demanded of someone. This subset, then, turns out to be equivalent to the type-B pleas I have identified.

But if one assimilates type-B pleas to type-A exempting pleas in this way, Strawson's argument that determinism could not be a type-A exemption is undermined. This argument turns on the crucial fact that issuing a type-A exemption means ceasing to make demands of someone,

and thereby exiting one's engaged interpersonal relationship with her. Because determinism is a universal thesis (if it is true, it is true of everyone, all the time), if determinism motivates a type-A exemption, it thereby motivates us to exit *all* of our engaged interpersonal relationships. And that, Strawson thinks, is impossible.³

If exemptions also include type-B pleas, however, Strawson's crucial fact is false. A type-B plea alters the content of the demand *while leaving some demand in place*. If exemptions include type-B pleas, it is possible to issue an exemption while leaving the engaged, demanding interpersonal relationship in place. Then a universal exemption would not mean exiting all our engaged human relationships.

It might be more natural to assimilate type-B pleas to type C. Type-B pleas adjust the content of the demand being made of the agent; such a plea undermines reactive attitudes by adjusting the demand to the point that the agent satisfied the demand rather than violating it. A type-C plea alters our understanding of the agent's actual quality of will, thereby showing that the agent's actual quality of will satisfied the demand rather than violating it. Both types of plea are, then, relevant to whether the demand was violated; they just approach the violation of the demand from opposite directions, so to speak.⁴

Both Watson and McKenna identify a lacuna in Strawson's discussion at just this point. Alongside type-C pleas they identify a category of pleas that they label *justifications*. Perhaps these justifications work like the second type of plea I have identified, by weakening the content of the demand. Watson, for instance, says that type-C pleas function by "denying the appearance that the other failed to fulfill the basic demand," adding, without much elaboration, that "of

³ Strawson's reasons for thinking this are the subject of much dispute. I set those disputes aside.

⁴ Wallace sees this. See (Wallace, 1998, pp. 143-147). But he fails to mark this distinction when he argues that determinism could not serve as an excuse; see (Wallace, 1998, pp. 147-153).

course, justification does this as well, but in a different way. ‘He realized what he was doing, but it was an emergency’” (Watson, 2004, p. 224).

Watson does not say much more to shed light on how he thinks about justifications. But reflecting on his example suggests that justifications work in just the same way as ordinary type-C pleas. “It was an emergency” is relevant because it indicates that this person acted in response to the emergency, rather than out of, say, malice or contemptuous disregard for my interests. Similarly, McKenna’s example of a justification reveals that an apparently racist remark is in fact intended to parody and criticize an actual racist.⁵ Such considerations prompt us to reassess our beliefs about the specific attitudes with which someone acted, in just the way other type-C pleas do.

The type-B pleas that I have identified—the ones that alter the *content* of our demands—should not similarly be assimilated to type C. That is because doing so would, once again, undermine Strawson’s argument about determinism. His argument that determinism could not function as a type-C plea turns on the claim that “it has never been claimed as a consequence of the truth of determinism that... anyone who caused an injury *either* was quite simply ignorant of causing it *or* had acceptably overriding reasons for acquiescing reluctantly in causing it *or* ..., etc. The prevalence of this happy state of affairs would not be a consequence of the reign of universal determinism, but of the reign of universal goodwill” (p. 80). It is obviously true that determinism would not generally affect our judgments about the actual attitudes with which people act, in the specific way required to undermine resentment. But type-B pleas do not change our judgments about people’s attitudes; rather they change the content of our demands. That is an adjustment to a normative judgment about what one ought to demand of people, not an adjustment to a judgment about what someone’s psychology was like. The truth of determinism

⁵ See (McKenna, 2012, p. 95).

is not obviously irrelevant to the former in the way it is obviously irrelevant to the latter. If type-B pleas were assimilated to type C then, once again, Strawson's argument would be unsound.

III. Applications

I conclude that, within the framework Strawson describes, there are four distinct categories of resentment-undermining plea. Type-A pleas exempt an agent from engaged interpersonal relationships, and their concomitant demands, altogether. Such pleas undermine resentment by undermining engaged demands altogether. Type-B pleas alter the *content* of the demands made of an agent. Such pleas undermine resentment if they undermine the demand that the agent apparently violated. Type-C pleas alter our understanding of the attitudes with which an agent acted. Such pleas undermine resentment if they indicate that the agent's actual attitudes did not violate a demand in the first place. Type A, B, and C pleas all undermine the judgment that a demand was violated. Type-D pleas alter the response we have to a violation of a demand. Such pleas leave the violation in place, but undermine resentment in particular as a response to that violation.

Now I want to explore some of the consequences of recognizing all four types of plea. First consider a case that has seemed to many to be difficult to square with the Strawsonian quality of will approach. In "Responsibility and the Limits of Evil," Gary Watson discusses a brutal killer, Robert Alton Harris, who was also the victim of terrible abuse as a child. As Watson makes clear, the standard Strawsonian options are really shockingly limiting here. Watson first rules out viewing Harris's upbringing as a Strawsonian excuse (what I have labeled a type-C consideration)—one that shifts our understanding of the attitudes with which an agent acted—by pointing out that Harris's background "in no way undermines the judgments that he is brutal, vicious, heartless, mean" (Watson, 2004, p. 242). Moreover, Watson argues, it does not

seem right to issue an exemption (what I have labeled a type-A consideration) in Harris's case. Doing so would mean exempting Harris from engaged, demanding interpersonal relationships altogether. We *might* do that on the basis of the apparent extreme intractability of Harris's evil; but that, Watson argues, would be dissatisfying when extended to other implacably evil agents (Watson, 2004, pp. 238-9), and in any case it does not explain why Harris's *specific history of abuse* makes us uneasy blaming him (Watson, 2004, pp. 242-3). Alternatively, we might exempt Harris on the basis of his specific history; but it is difficult to explain why historical considerations—having to do with how Harris got to be how he is, rather than how he is *now*—would prevent us from participating in any engaged relationships with him *now*.

There are, we can now see, additional options available to the quality of will theorist: bad formative circumstances might function as a type-B consideration, one that prompts us to demand less of Harris (without disengaging from him altogether); or they might function as a type-D consideration, one that prompts us to respond differently to his violations of the engaged demands we do make of him. The former option might be motivated by consideration of what it is fair to expect of Harris: it might not be fair to expect someone to endure the abusive environment that Harris grew up in without developing a coldhearted disregard for the lives of others. (If society or the state were somehow implicated in Harris's childhood—if, for instance, the state could have removed Harris from his terrible family environment, but failed to do so—this reasoning might seem especially attractive.) Or we might keep in place the demand that Harris show due regard for the lives of others, while modifying the responses that we think are appropriate to his violations. We might think that, in light of Harris's terrible upbringing, angry or retributive reactions should give way to, or be complemented by, a focus on giving Harris the support in becoming a morally decent person that he never had as a child.

I do not aim here to establish any specific conclusions about how we *should* respond to Harris. I only want to point out that recognizing these additional possibilities within the quality of will framework provides further, and far more palatable, options for responding to Harris's terrible upbringing. Quality of will theorists are not, as Watson worried, limited to a choice between concluding that Harris is less terrible than he seems, or else disengaging altogether.

Next I want to consider an old pest: the truth of a thesis of determinism. My aim here is modest. I want to identify two ways of arguing for incompatibilism that are left open by Strawson's argument in "Freedom and Resentment." That is not to say that I think such arguments succeed; perhaps they fail. But additional argument is needed to show that. The arguments of "Freedom and Resentment" (recapitulated by many prominent quality of will theorists; see, e.g., (Wallace, 1998, pp. 147-153)) will not suffice. Further, I want to suggest some affinity between the sort of incompatibilism that these arguments would support and traditional incompatibilist positions.

Many incompatibilist arguments begin with an intuition that if determinism is true there is something deeply unfair about our actual responsibility practices. Strawson argues that we cannot cash this intuition out in terms of type-A pleas: determinism could not threaten our participation in engaged, demanding interpersonal relationships across the board. What's more, Strawson points out, it does not seem likely that determinism will operate as a resentment-undermining type-C plea, one that prompts us to revise our judgments about the specific attitudes with which people act. As I have argued in section II, these arguments cannot be extended to type-B or type-D pleas.

Strawson's arguments thus leave open the possibility that determinism might operate as a type-B plea. That is, determinism might generally weaken the demands we make of each other—not undermining them altogether, but making them weak enough that everyone does actually satisfy them. (Although I cannot pursue the matter further here, there is at least some *prima facie* plausibility to this thought. It can seem perverse and deeply unfair to demand of someone something that she cannot deliver. And if determinism is true, then there is a sense in which nobody can do otherwise than they actually do. Perhaps by this route determinism would, or should, make us less demanding of each other.)

Strawson's arguments also leave open the possibility that determinism might operate as a type-D plea, ones that systematically alters our responses to demand violations (I have argued for this conclusion in Goldman, 2014). Thus—for all that Strawson has argued—it is possible that the truth of determinism would, or should, motivate an across-the-board elimination of antagonistic response in favor of moral disappointment or some other nonantagonistic moral response.

Neither of these two possibilities involves undermining our basic engaged, demanding interpersonal relationships. (They are not type-A exempting pleas.) And so traditional Strawsonians would deny that they involve undermining moral responsibility. But each does involve a profound alteration of our current, antagonistic practices of responding to wrongdoing. What's more, these alterations do seem to capture what some skeptics about moral responsibility actually talk about.

Take (P. F. Strawson's son) Galen Strawson, for instance. The younger Strawson believes that nobody is “truly or ultimately” morally responsible. “As I understand it,” he writes, “true

moral responsibility is responsibility of such a kind that, if we have it, then it *makes sense*, at least, to suppose that it could be just to punish some of us with (eternal) torment in hell and reward others with (eternal) bliss in heaven” (G. Strawson, 1994, p. 216). “True moral responsibility” involves susceptibility to specific, punitive responses to wrongdoing.

Undermining harsh, antagonistic responses—while perhaps leaving in place some milder responses—is precisely what determinism would accomplish, if it were a type-D consideration.

Attacks on specific aspects of moral response may, then, have a home *within* the quality of will framework as I have developed it. And framing such attacks in this way is quite useful. The attack is now extremely narrowly targeted: it is focused on a very narrow moral question about the fairness of the content of some specific standards. It is thus distinguished from questions about the fairness of imposing particular penalties for violations of those standards (a type-D question), and from questions about the fairness or the viability of the general practice of holding people accountable (a type-A question).

Importantly, this sort of attack is compatible with the continued existence of the general framework that Strawson identifies, the framework of engaged interpersonal relationships and demands for goodwill. That is not to deny the significance of such attacks; establishing that nobody ever violates a legitimate demand, or undermining all antagonistic responses to violations, would involve deep and far-reaching changes to our practices. But the framework of engaged, demanding interpersonal relationships, which the quality of will theorist takes to be essential to our moral responsibility practices, remains intact.

Bennett, J. (1979). Accountability. In *Philosophical Topics* (pp. 14–47). Oxford.

Goldman, D. (2014). Modification of the Reactive Attitudes. *Pacific Philosophical Quarterly*, 95, 1–22. doi:10.1111/papq.12014

Kelly, E. (2013). What Is an Excuse? In D. J. Coates & N. A. Tognazzini, *Blame: Its Nature and*

- Norms* (pp. 244–262). Oxford: Oxford University Press.
- McKenna, M. (2012). *Conversation and Responsibility*. Oxford: Oxford University Press.
doi:10.1093/acprof:oso/9780199740031.001.0001
- Strawson, G. (1994). The Impossibility of Moral Responsibility. *Philosophical Studies*, 75(1/2), 5–24.
- Strawson, P. F. (2003). Freedom and Resentment. In G. Watson, *Free Will* (2nd ed., pp. 72–93). Oxford, UK: Free Will.
- Wallace, R. J. (1998). *Responsibility and the Moral Sentiments*. Harvard University Press.
- Watson, G. (2004). Responsibility and the Limits of Evil: Variations on a Strawsonian Theme. In *Agency and Answerability : Selected Essays* (pp. 1–21). Agency and Answerability.