1. Introduction
Although Peter Strawson’s “Freedom and Resentment” was published 50 years ago and has been a widely discussed article, its main argument is still notoriously difficult to pin down. One significant reason for this, it seems to me, is that too often, it has been interpreted as a naturalistic attempt to justify the practices of moral responsibility—one grounded in mere biological, psychological, or sociological necessities. But this interpretation, I believe, is fundamentally mistaken.

Thus to rectify this recalcitrant mistake, in this paper I want to offer an alternative interpretation of Strawson’s argument, one that situates it firmly within the Kantian tradition. We’ll see this in Strawson’s objections to the extant compatibilist and incompatibilist positions, which are distinctively Kantian and anticipate Strawson’s subsequent endorsement of Kant’s methodological principle in The Bounds of Sense. But we’ll also see that rather than a naturalistic reduction of moral responsibility, as many have

For exceedingly helpful feedback on these ideas and on earlier drafts of this paper, I am very grateful to Maudemarie Clark, Christopher Franklin, Pierre Keller, Samantha Matherne, Michael Nelson, Martha C. Nussbaum, John Perry, Andrews Reath, Neal A. Tognazzini, and Gary Watson. And I am especially indebted to John Martin Fischer’s insight and encouragement in the writing of this paper.

2 For example, see John Fischer, “Peter Strawson and the Facts of Agency ,” unpublished manuscript [2012]. There, Fischer claims that “it is perhaps an unfortunate thing that this essay does not wear its meaning on its sleeve, so to speak; indeed I find it quite challenging” [Fischer 2012, 1]. And Fischer is hardly alone on this point. At a recent conference at the College of William & Mary in honor of “Freedom and Resentment,” a number of the invited speakers similarly acknowledged the difficulty of Strawson’s argument.
3 As a biographical point, this should hardly be surprising. After all, “Freedom and Resentment” was published between Individuals, Methuen & Co., Ltd., [1959] and The Bounds of Sense, Methuen & Co., Ltd., [1966]. In the former, Strawson offers a transcendental argument for the existence of other minds, and in the latter, he endorses key features of Kant’s critical philosophy.
suggested, Strawson offers a *transcendental argument* for the legitimacy of our moral responsibility practices. But getting Strawson right on this point isn’t just important for exegetical reasons; interpreting Strawson in this way provides us with a novel, yet powerful, argument for the compatibilism of moral responsibility and causal determinism.

---

Strawson begins “Freedom and Resentment” by outlining the conceptual landscape. He first distinguishes between two positions: *optimism* and *pessimism*. On the one hand, the optimist claims that the truth of causal determinism would not threaten the legitimacy of our responsibility practices. Consequently, the optimist is a *compatibilist* about moral responsibility and causal determinism. On the other hand, however, the pessimist worries that the truth of causal determinism would threaten the legitimacy of our responsibility practices. And so the pessimist is an *incompatibilist*. But rather than defending or rejecting one of these positions outright, Strawson interestingly takes his project to be that of reconciliation. In other words, he seeks to marry the virtues inherent to each of these views—a project which will require, as he puts it, “a formal withdrawal [from the optimist] in return for a substantial concession [from the pessimist]” [Strawson 1974, 2]. But before we can appreciate the “formal withdrawal” that the optimist owes us or the “substantial concession” that is necessary from the pessimist, we first need to appreciate those features of optimism and pessimism that Strawson seeks to retain. For despite their differences, Strawson thinks that each of these views is correct about some important feature of our responsibility practices. In fact, as we’ll see, Strawson’s own positive account of our practices’ justification relies on these aspects of optimism and of pessimism respectively.

Regarding these views, then, Strawson claims that while the optimist is right to think that the “facts as we know them” [Strawson 1974, 2] provide an adequate justification for our responsibility practices, the pessimist is right to think that our responsibility practices resist naïve utilitarian justifications. And by Strawson’s lights, we must respect each of these two insights if we are to have any hope in offering a plausible justification for our moral responsibility practices. So while he offers a diagnosis of the mistakes

---

4 Of course, as we’ll see, in reconciling optimism and pessimism, Strawson ultimately endorses a “sufficiently, that is *radically*, modified” [Strawson 1974, 25] optimism.
inherent to traditional statements of optimism and pessimism, in “Freedom and Resentment,” Strawson also argues that the insights of optimism and pessimism can be reconciled in a way that renders our responsibility practices justified even if causal determinism obtains.

Since its publication, Strawson’s argument has most commonly been teased out as a naturalistic argument. On this reading, our responsibility practices are justified (even if causal determinism obtains) because it would be psychologically impossible—or nearly psychologically impossible—that we abandon praise and blame. But as it has been frequently suggested, this is implausible as a justification. However, it seems to me that this interpretation of Strawson’s positive argument is mistaken. For as we’ll see, Strawson isn’t claiming that it would be psychologically impossible for us abandon our responsibility practices. Instead, he is claiming that it would be practically impossible for us to abandon these practices given our antecedent practical commitment to participation in ordinary human relationships. In other words, for Strawson, our responsibility practices are justified because it is impossible, from the first-person point of view of relational agency, to fail to regard yourself and others as, in general, morally responsible agents. Strawson thus offers a transcendental argument for the legitimacy of our practices.

But before we can get to this reinterpretation of Strawson’s positive argument, I first want to consider his objections to optimism and pessimism, which I’ll do in §2. Then, in §§3-4 I will reconstruct Strawson’s positive argument, and I conclude by arguing that it is only when we understand Strawson as offering a transcendental argument, that we can hope to appreciate its subtlety as an argument for compatibilism.

2. Diagnosing the Problem

According to Strawson, both the optimist and pessimist justifications for moral responsibility face significant challenges. As we’ll see, however, at a fundamental level, these attempts to justify our responsibility practices are mistaken for similar reasons.

---

5 On some interpretations, our commitment to responsibility practices is a biological or sociological necessity rather than a psychological necessity.
The optimist identifies the justification for our responsibility practices with their instrumental role in creating a happy social order. That is, the optimist seeks to justify our responsibility practices on purely consequentialist grounds, by appealing to their usefulness in securing a well-ordered society—“optimists about determinism point to the efficacy of the practices of punishment, and of moral condemnation and approval, in regulating behavior in socially desirable ways,” [Strawson 1974, 2]. But this is quite implausible. Indeed, because they appeal to the efficacy of our responsibility practices in regulating behavior as a justification for the practices, the optimists miss something vital; as Strawson says, “[utility] is not a sufficient basis, it is not even the right sort of basis, for these practices as we understand them,” [Strawson 1974, 4]. As Stephen Darwall puts Strawson’s point here, the “desirability [of a particular social order] is a reason of the wrong kind to warrant the attitudes and action in which holding someone responsible consists in their own terms,” [Darwall 2006, 15]. In other words, according to Strawson, the optimist appeals to the wrong set of facts—i.e., the wrong kind of reasons—in her attempt to justify our responsibility practices.

Thus, Strawson accuses the optimist of over-intellectualizing our responsibility practices, since she seeks to justify those practices with a calculating appeal to their instrumental value in bringing about a well-regulated social order. As a result, the optimist fails to take into account all of the facts as we know them—most notably, she ignores the fact “our practices do not merely exploit our natures [for the good of social regulation], but express them,” [Strawson 1974, 25]. And in so doing, she fails to appreciate the significance of our responsibility practices. Indeed, understanding our practices in this way cheapens both the practices themselves as well as the status of our interpersonal relationships, which Strawson takes to be intimately tied to “non-detached” responsibility-entailing attitudes like resentment and gratitude, esteem and indignation, etc. For if the justificatory status of our responsibility practices rests on their utility, then we cannot understand ourselves as legitimately involved in directly interpersonal relationships, since if the optimist is correct, the legitimacy for such relationships would depend on some utility calculus. But of course, it’s absurd to think that it would be appropriate to be grateful to or to love another merely because Utility(love) > Utility(~love). It can be appropriate

---

for us to love or care about another simply because they are our friend, our brother or sister, our mentor, or that they are just simply lovely.\(^8\)

By contrast, the pessimist rejects the optimist’s nakedly consequentialist justification for our practices. Indeed, she “recoils”—and rightly so, according to Strawson—at the notion that our responsibility practices are mere instruments for social regulation, to be justified if and only if they efficiently bring about certain forms of behavior. Thus, the pessimist position can be seen as an attempt to take seriously the idea that only backward-looking considerations can serve to justify our practices. And specifically, for the pessimist, it is only freedom of the will and basic desert (of the sort that would entail the falsity of causal determinism) that can serve to justify our practices.\(^9\) However, the falsity of causal determinism is not among the “facts as we know them,” and so the pessimist wrongly, at least according to Strawson, takes these facts to be inadequate as a justificatory basis for our practices. And by stepping outside of the facts as we know them in her attempt to justify our practices, Strawson claims that the pessimist fails to adequately recognize and appreciate her “own humanity” [Strawson 1974, 24], which as we’ll see, Strawson takes to be a sufficient basis upon which to ground our responsibility practices.

So as it turns out, like the optimist, the pessimist also over-intellectualizes our responsibility practices, since she seeks to explain their significance by going beyond “the facts as we know them.” According to the pessimist, the vital thing that the optimist justification ignores is not the essentially relational nature of our humanity,\(^10\) but is instead some further

---

\(^8\) Strawson’s criticisms of the optimist, at least as the optimist is construed at the outset of “Freedom and Resentment,” are of a piece with many of the traditional objections to utilitarianism as a general normative theory.

\(^9\) As I understand it, “basic desert” is a conception of desert according to which \(S\) deserves \(X\) cannot be analyzed in terms of consequentialist or contractualist considerations. I have to admit however, that although I can sympathize with the pessimists’ tendency to identify desert as somehow essentially implicated in the concept of responsibility, it is not clear to me why the relevant notion of desert must be unanalyzable. For more on basic desert, see Michael McKenna, Responsibility and Conversation, Oxford University Press [2012] and Derk Pereboom, “Free Will Skepticism, Blame, and Obligation,” Blame: Its Nature and Norms, eds. D. Justin Coates and Neal A. Tognazzini, Oxford University Press [forthcoming] pgs. 189-206.

\(^10\) Cf. Strawson’s criticism of the optimist’s naïve consequentialism in terms of a failure to appreciate that our responsibility practices do not merely exploit our relational natures, they express them [Strawson 1974, 24].
metaphysical fact (e.g., that we possess “freedom of the will”). And for Strawson, this further metaphysical fact is nothing more than “a pitiful intellectualist trinket” that we might “wear as a charm against the recognition of [our] own humanity” [Strawson 1974, 24].

All joking aside, though, it’s important to recognize that behind Strawson’s critical rhetoric, there’s an important point lurking. For whereas “the optimist’s style of over-intellectualizing the facts is that of a characteristically incomplete empiricism, a one-eyed utilitarianism” [Strawson 1974, 23], the pessimist’s style of over-intellectualizing the facts is that of (to borrow from a line from elsewhere in Strawson’s oeuvre) a “dogmatic rationalism [that] exceeds the upper bounds of sense” [Strawson 1966, 12]. Pessimists make this latter mistake because they fail to recognize that the agential capacities which underwrite our participation in meaningful interpersonal relationships are sufficient to underwrite our status as morally responsible agents.

The pessimist’s mistake comes into even sharper focus when we consider Strawson’s subsequent endorsement of Kant’s so called “Principle of Significance” in The Bounds of Sense. According to the Principle of Significance, “if we wish to use a concept in a certain way, but are unable to specify the kind of experience-situation to which the concept, used in that way, would apply, then we are not really envisaging any legitimate use of that concept at all” [Strawson 1966, 16]. And an important consequence of the Principle of Significance—at least as it was wielded by Kant—is that so called “transcendent” metaphysical concepts (e.g., freedom of the will, God, the world-whole), which abstract away from the intuitive contexts that give concepts their significance, have no justification. Following Kant, it seems that Strawson also wields the Principle of Significance against the “panicky metaphysics” of pessimism. Inflated metaphysical notions of freedom and responsibility, which are not grounded in our experiences as participants in ordinary human relationships, cannot justify our practices of praising and blaming because they too, like God and the world-whole, abstract away

11 In a companion piece, “Responsibility Without (Panicky) Metaphysics,” I explore in greater detail the contours of moral responsibility in a way that respects the methodological constraint that we not extend our theorizing beyond the facts as we know them to be (as we’ll see, like Strawson, I think this is fundamentally, a Kantian point).

12 Like transcendental freedom, God and the world-whole were products of transcendent metaphysical speculation that is not grounded in the conditions of possible experience.
from the context that gives them significance. On the basis of the Principle of Significance, then, Strawson rejects pessimism.  

If this is correct, then we can understand Strawson’s diagnosis of the problems of optimism and pessimism to be roughly analogous to Kant’s diagnosis of the problems found in Empiricism and Rationalism. For Kant argues that the Empiricists, who, in the Third Antinomy, defend the “Antithesis”—i.e., the view that “there is no freedom; everything in the world takes place solely in accordance with laws of nature” [CPR A445/B473]—fail to take seriously our humanity because their view cannot accommodate the practical perspective, which Kant takes us to be committed to qua agents. Further, Kant similarly criticizes the Rationalists, who defend the “Thesis”—the view that “causality in accordance with laws of nature is not the only causality… [and that] it is necessary to assume that there is also another causality, that of freedom” [CPR A444/B472]. Thus, (if Strawson’s interpretation of Kant is correct) the Rationalists also fail, in some sense, to take seriously our humanity, since in their zeal to posit transcendental freedom as an independent species of causation, they abstract away from the conditions required for possible experience—the very conditions that give our status as persons in a causal order its significance. In short, both the Empiricist and the Rationalist over-intellectualize the facts. 

Moreover, in discussing the third antinomy Strawson explicitly claims of the Third Antinomy that “it seems obvious what the correct “critical” solution of this conflict should be. … Every member of the series [of causes] which is actually “met with” in experience, however, may and must, be taken to have an antecedent cause. The thesis, then, is false, the antithesis true” [Strawson 1966, 209]. In other words, according to Strawson, although there are problems with both thesis and antithesis, the right way to resolve the antinomy is not by appeal to transcendental idealism, but simply to accept (a suitably modified) form of the antithesis (i.e., empiricism). And this is precisely parallel to the move that Strawson makes with respect to the debate between pessimist and optimist, ultimately siding with a suitably modified version of optimism. This tight parallel with his own solution to the Third Antinomy further suggests that Strawson has set up the dialectic between optimism and pessimism in a way that parallels Kant. We should expect the solution that Strawson offers, then, to also parallel Kant in certain ways. 

Of course, Strawson’s interpretation is widely regarded as an extremely idiosyncratic and motivated reading of Kant’s resolution to the Third Antinomy, one stemming from Strawson’s rejection of Kant’s transcendental idealism. Henry Allison, for example, writes that “it is curious to find Strawson claiming that the critical solution to the antinomies is to deny the thesis and affirm the antithesis. …this fl[ies] in the face of Kant’s explicit claim that … neither the thesis nor the antithesis is true [Allison 1990, 23]. For more on this point, see Henry Allison, Kant’s Theory of Freedom, Cambridge University Press [1990]. 

On top of Allison’s criticism, there is a further reason to think Strawson’s interpretation isn’t quite right. For if Strawson’s interpretation of Kant on this point is correct, then it points to a deep tension in Kant himself. After all, you might think, in light of Kant’s claim in *Groundwork* III that the moral law is grounded in the *idea* of freedom (where an “idea” is cashed out in terms of a concept that no intuition can possibly correspond to), that Kant himself is guilty of “over-intellectualizing” our moral practices in
Of course, merely diagnosing the mistakes of optimism and pessimism is not Strawson’s only goal in “Freedom and Resentment.” As I’ve already suggested, he also means to offers us a way of reconciling the optimist’s commitment to the “facts as we know them” with the pessimist’s commitment to our status as morally responsible agents as not being derivable from the utility of our responsibility practices. In other words, he means to offer a positive justification for our moral responsibility practices. Accordingly, I turn to this below.

3. Freedom and Resentment, Kantian Style

In a crucial passage, Strawson offers more fully his own positive account for the justification of our moral responsibility practices. Specifically, he claims that:

I am strongly inclined to think that it is, for us as we are, practically inconceivable [that the truth of determinism could lead us to abandon the reactive attitudes]. 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 [Strawson 1974, 11, emphasis added].

And with that, we can see, in its broadest outline, the subtle but powerful justification Strawson advances on behalf of our responsibility practices: namely, that it would be “practically inconceivable” that we could abandon the responsibility-entailing relations.

But why, you might naturally ask, should the fact that it is “practically inconceivable” that we abandon ordinary interpersonal relationships serve as a justification for our moral responsibility practices? Indeed, what does it even mean to say that it is “practically inconceivable” that we abandon our commitment to ordinary human relationships? These questions become

precisely the way that Strawson accuses the pessimist of over-intellectualizing. On this final point, I am greatly indebted to Samantha Matherne [personal correspondence]
even more pressing when we follow the standard interpretative assumptions and adopt a naturalistic reading of Strawson’s claim that our “thoroughgoing and deeply rooted” commitment to “participation in ordinary inter-personal relationships” is practically necessary. To assess the justificatory prospects of this interpretation, then, I turn to the gold standard for naturalistic readings of Strawson below (i.e., Paul Russell’s interpretation of Strawson). And as we’ll see, two distinct naturalistic interpretations of Strawson’s proposed justification fail. However, rather than showing Strawson’s argument to be problematic, these failed interpretations point us to an alternative, Kantian interpretation.

***

First, in claiming that our inescapable commitment to responsibility-entailing relationships (and to the reactive attitudes themselves) justifies our responsibility practices, Strawson might be endorsing what Paul Russell has called “type naturalism.” According to Russell, type naturalism is the thesis that our proneness to the reactive attitudes “is natural to humans and requires no general justification of any sort,” [Russell 1992, 294]. On this interpretation of Strawson’s claim, the thoroughgoingness and depth of our commitment to the reactive attitudes is taken to show that worries about insufficient justification for our practices are misguided. After all, if it’s simply a fact about human nature that we are ineluctably prone to resentment, gratitude, et al., then like other facts about human nature, e.g., that we have eyebrows, it is not in need of justification.

The problem with this view, which Russell notes, is that even if it is true that human are naturally disposed to engage with others through the reactive attitudes, that fact alone won’t undercut those pesky questions of justification that pessimists are likely to raise. Indeed, even the pessimist can agree that natural facts about humans don’t require any general justification. But it’s a further thing to therefore conclude that these facts do not need justification in specific contexts. For example, if having eyebrows came to be widely associated with support for a genocidal dictator, then one could certainly ask for justification from those who do not shave off the hair above their eyes. And of course, pointing to most humans’ natural disposition to

---

grow eyebrows does no justificatory work in this context. But plausibly, the truth of causal determinism also provides a context in which we might lacking justification for engaging with others through blaming attitudes like resentment and indignation, natural though such engagement may be. As Russell puts it:

The Pessimist does not (or need not) claim that we are capable of suspending our disposition or liability to the reactive attitudes—much less that the thesis of determinism requires us to do so. … Rather, the Pessimist claims only that we can and must cease to entertain reactive attitudes insofar as we have reason to believe that everyone is incapacitated in the relevant ways. If the thesis of determinism is true, the Pessimist argues, then we are, indeed, all morally incapacitated [Russell, 269].

In other words, the pessimist can accept type naturalism since she can admit that, of course, we would be thoroughly and deeply disposed to engage with others via the reactive attitudes even if determinism were true. However, as we've seen, this does not thereby commit the pessimist to the further claim that any particular instance of resentment or indignation is fitting.

To further motivate this claim, consider an analogous case concerning the emotion of fear (rather than say, blaming attitudes like resentment or indignation). When I step out onto the Skywalk—the glass-bottomed walkway that extends out over the Grand Canyon—I am ineluctably disposed to fear. Even if the fear doesn’t manifest itself in any particular instance, I am nevertheless prone to tremble and shake, while staring down at the canyon floor a mile below. Of course, the Skywalk is perfectly safe, and so no instance of fear would be fitting, since my safety is never threatened, and fear is fitting only if my safety is, in fact, threatened. Despite this, we could say that it’s just a fact about me—and about humans more generally—that we are disposed to fear our deaths when we step out onto high ledges like the Skywalk. So even if I was a pessimist about fear, and I thought that the world was ordered in a way that was incompatible with me ever being genuinely threatened (e.g., perhaps I have a guardian angel that follows me around and ensures that I won’t ever be threatened), I could nevertheless be a type naturalist about fear.

Thus, if fear type naturalism doesn’t commit me to thinking that any instance of fear is warranted, then neither does type naturalism about
resentment or any other blaming attitude commit me to thinking that any instance of blame is warranted. And if this is correct, then Strawson cannot appeal to the truth of type naturalism to undermine the pessimist’s position, since the pessimist can happily accept type naturalism. Thus, if Strawson is correctly read as a type naturalist, he fails to provide a positive justification for our responsibility practices, since the truth of type naturalism wouldn’t show our responsibility practices to be justified if causal determinism obtains.

***

Consequently, Russell claims that if Strawson is to undercut the pessimist’s position, he must do so by embracing token naturalism. Token naturalism is the claim that our proneness to attitudes like resentment and indignation is so deeply rooted in human nature that “we will inescapably or inevitably continue to entertain or feel [these attitudes], whatever reason suggests to us,” [Russell 1992, 295]. So unlike type naturalism, token naturalism commits Strawson to an implausibly strong claim. After all, on token naturalism (but not type naturalism), the natural human proclivity to engage with others through the reactive attitudes is one that is wholly insensitive to reason.

In other words, even if we recognize there to be undeniably weighty reasons to refrain from regarding others from this perspective—reasons that will undercut the legitimacy of our emotional engagement with others via the reactive emotions—according to token naturalism, our commitment to blaming attitudes like resentment and indignation is so unshakeable that we ineluctably find ourselves resenting or getting indignant, whatever else reason might say. Because of this inevitability, if token naturalism is true, our engagement with others via blaming attitudes can’t be any more rationally objectionable than other given facts of human nature—e.g., that we have eyebrows or that we are bipedal. And if we can’t refrain from feeling resentment and indignation, then we can’t be obligated to do so, since, after all, ‘ought implies can’. Thus, on this interpretation of Strawson, our responsibility practices are justified in virtue of the fact that no alternative set of practices could possibly be required.

Unfortunately for Strawson however, while token naturalism can answer the pessimist’s challenge, it is incredibly implausible as a thesis about human nature. Russell makes this point forcefully:
it seems clear why the Pessimist finds Strawson’s [token] naturalistic reply both misguided and disturbing. What is particularly disturbing about Strawson’s naturalistic strategy, expressed in more general terms, is that it casts doubt on our ability or capacity to curb or control our emotional life according to the dictates of reason. More specifically, it seems clear that, despite disclaimers to the contrary, Strawson’s naturalistic strategy invites us to accept or reconcile ourselves to reactive attitudes (and their associated retributive practices) even in circumstances when we have reason to repudiate them. Given this, it seems evident that we have good reason to reject Strawson’s suggestion that we dismiss the Pessimist and refuse to take his arguments seriously. We have, on the contrary, every reason to take the Pessimist seriously [Russell 1992, 297-298].

In other words, because token naturalism commits us to an implausible conception of human nature—one in which we genuinely are slaves to destructive forces of our reactive emotions—we have good reason to doubt its truth. After all, we certainly seem to be capable of mollifying or withholding the reactive attitudes when the situation calls for it.

On this point, I heartily agree with Russell. Token naturalism is plausibly false since we are, in fact, able to refrain from resenting in particular instances if exempting or excusing conditions are present. But in addition to the fact that it implausibly characterizes human nature, the putative truth of token naturalism does not obviously square with general arc of Strawson’s argument. After all, Strawson himself is very sensitive to the role of exempting and excusing conditions in our ordinary responsibility practices. Indeed, in discussing these two families of pleas, Strawson admits that our practices are sensitive to considerations that undermine an agent’s capacity to fully participate in ordinary interpersonal relationships. That an agent wasn’t himself, that he was pushed, that he was coerced, that he was under significant duress, etc., are all facts that are relevant to apt application of the reactive attitudes. Concerning this, Strawson claims that:

Inside the general structure or web of human attitudes and feelings of which I have been speaking, there is endless room for modification, redirection, criticism, and justification. But questions of justification are internal to the structure or relate to modifications internal to it. The
existence of the general framework of attitudes itself is something we are given with the fact of human society” [Strawson 1974, 23].

In other words, Strawson accepts Russell’s point that the blaming attitudes are not wholly insensitive to reason. Thus, Strawson seems to think that there is room for rational revision of the sort Russell thinks would be impossible given the truth of token naturalism. And this suggests that we should not understand Strawson’s attempt to justify our practices as relying on the truth of token naturalism.

Moreover, since Strawson does think that it is possible that we can, at times, refrain from engaging with others via the blaming attitudes, his insistence that it is “practically inconceivable” that we do so upon learning that causal determinism obtains should not be interpreted naturalistically, as describing a kind of psychological, biological, or sociological necessity. Thus, if the inescapability of our commitment to responsibility-entailing relationships is to serve as a justification for our responsibility, it can’t be because such commitment is naturally necessary for humans.15

* * *

This, then, invites us to reconsider what it means for our commitment to ordinary interpersonal relationships to be practically inescapable. As I understand him, when Strawson claims that it would be “practically inconceivable” that we abandon our commitment to interpersonal relationships and to the class of human attitudes that shape such relationships (which for Strawson, includes resentment and indignation), he is not making a psychological (or biological or sociological) claim at all. To say that it is “practically inconceivable” is to say that it is inconceivable or inescapable from the practical point of view of agency that we abandon these ordinary relationships. In other words, Strawson is claiming that from the

---

15 In a footnote Strawson does compare the justification of our responsibility practices with the justification of inductive belief-formation, calling the latter “original, natural, non-rational (not irrational)” [Strawson 1974, 23]. He then approvingly cites Hume as giving making clear the facts about the external basis of induction. This suggests—more strongly than anything else—that Russell’s interpretation of Strawson is a bit closer to Strawson’s project than the interpretation I offer. Of course, I think there a host of reasons for thinking that this footnote is anomalous, but I would just say that even if Strawson is implicitly trying to ground moral responsibility through token-naturalist commitments, there is a position very close to the one that Strawson articulates—one that he hints at—and it is this position, one closer to Kant than to Hume, that I want to consider.
perspective we take on the world as agents, we cannot fail to regard ourselves as, in general, participants in ordinary human relationships. And because Strawson takes our ordinary relationships to be responsibility-entailing relationships, he concludes that we cannot abandon our responsibility practices from our practical perspective as agents. So the practical inescapability of our commitment to responsibility-entailing relationships serves as part of a transcendental argument for the legitimacy of our responsibility practices.

So understood, Strawson argues that our responsibility practices are justified because they are conditions for the possibility of the ordinary interpersonal relationships that we are committed to from our practical perspective as agents. But why should we accept this? Derk Pereboom and Tamler Sommers, to name two examples, have argued very forcefully that ordinary interpersonal relationships are not responsibility-entailing relationships. So it’s not enough to simply assert that these relationships are, in fact, responsibility entailing. In what follows, then, I will fill in this argument, showing that the undeniable legitimacy of our ordinary human relationships does depend on such relationships being responsibility-entailing relationships. I will then offer an account of how this claim and the claim that our commitment to ordinary human relationships is practically inescapable, serve as the basis for Strawson’s transcendental argument.

4. Reactive Attitudes and Interpersonal Relationships

Strawson’s argument crucially turns on what he takes to be a commonplace of human life: our participation in meaningful forms of reciprocal relations. This fact about us, Strawson thinks, isn’t merely important for our standing as responsible agents—it’s ineliminably vital (indeed, this is what the optimist misses). After all, it is through interpersonal relationships that we, “as sharers of a common interest; as members of the same family; as colleagues; as friends; as lovers; as chances parties to an enormous range of transactions and encounters” [Strawson 1974, 6], are meaningfully engaged with one another directly, and reciprocally, through “non-detached” attitudes and emotions. And Strawson (plausibly) thinks that it only because we engage in these kinds of relationships that we are morally responsible to one another. In other words, for Strawson, moral responsibility (or the property of being

---

16 Cf. Derk Pereboom [2001] and Tamler Sommers [2011].
morally responsible) is a property instantiated by human agents only because we are creatures who participate in these sorts of relationships.

It’s worth noting that this is one of Strawson’s most important insights regarding the nature of moral responsibility. And it’s one that was largely ignored prior to Strawson and only haltingly heeded since.

To see this, compare Strawson’s conception of moral responsibility with those of Nietzsche and of Galen Strawson. For Nietzsche, the *causa sui*, i.e., freedom of the will in the “metaphysical superlative sense,” and the kind of moral responsibility that such freedom underwrites, is grounded in “the desire to bear the whole and sole responsibility for one’s actions and to absolve God, the world, ancestors, chance, and society from responsibility for them,” [BGE 21].

This conception of freedom and responsibility is certainly a far cry from Strawson’s idea that morally responsible agency is the glue that holds together ordinary human relationships.

Similarly, Galen Strawson, writing two and half decades after his father, identifies the property of being morally responsible with the property of being an apt target of divine damnation (or beatification, as the case may be). This, too, takes us away from (Peter) Strawson’s more minimal, but no less robust, conception of moral responsibility as being grounded in participation in distinctively human forms of engagement. For even if Galen Strawson is

---


19 See Galen Strawson, *Freedom and Belief*, Oxford University Press [1986].
correct to think that it is impossible for us to be ultimately morally responsible in the sense that it is fitting for God to eternally damn us, why should it thereby follow that we are not morally responsible in the sense that it is fitting for others to engage with us via the reactive emotions? Moreover, other than “metaphysical megalomania,” why should we even care about a sense of moral responsibility that abstracts away from our ordinary interpersonal relationships? Of course, in light of the emphasis Strawson places on ordinary interpersonal relationships, it's natural to wonder about their import. Undoubtedly, I stand in such relationships: I am a partner, a child, a friend, a colleague, a sibling, etc. But so what? That is, what's so special about these relationships? And why should we think that they can serve as the justificatory basis of our responsibility practices?

Well, for Strawson, these relationships are special because they manifest our commitment to the values of good will and respect—values located at the heart of morality. In fact, Strawson identifies this class of relationships exclusively in terms of a general concern for good will (a concern he calls “the basic demand”), where this concern is reflected in the “non-detached” attitudes and emotions that play a role in structuring reciprocal interpersonal engagement. These emotions, which Strawson calls the reactive attitudes, include blaming attitudes like resentment, indignation, and guilt, but also their positive counterparts of gratitude, esteem, and pride. And the reason that these emotions are essential to reciprocal interpersonal relationships (i.e., relationships that require each party will regard the other with good will and respect) is that our proneness to the reactive attitudes is itself, according to Strawson, “the making of the demand [for good will and respect]” [Strawson 1974, 22]. In other words, if I'm not prone to respond to your actions with gratitude and resentment, esteem and indignation, then I am not really regarding you as being subject to the basic demand that we

---

20 This description of what the incompatibilist wants is due to John Martin Fischer, “Playing the Cards That Are Dealt You,” Journal of Ethics 10, [2006]: 235-44.
21 To follow Daniel Dennett, it seems that ultimate moral responsibility is not a “variety of [moral responsibility] worth wanting.” See Daniel Dennett, Elbow Room: The Varieties of Free Will Worth Wanting, MIT Press [1984]. Fischer [2006] also makes this point.
22 Strawson also includes some harder to characterize emotions such as love, forgiveness, and hurt feelings. Although these are of much interest, because they are not Strawson's focus, I will have little to say about them here (however, I do say more about love in particular in D. Justin Coates, “In Defense of Love Internalism,” unpublished manuscript, [2012]).
show others good will and respect. And if I fail to regard you as subject to this demand, then I can’t consistently engage with you in relationships that require you to be subject to this demand, relationships such as friendships, mature loves, etc.). Therefore, if I fail to hold you to this expectation, I am not justified in treating you as a friend.

* * *

More baldly, Strawson’s argument for thinking that engagement with others through the reactive attitudes is a condition for the possibility of genuinely reciprocal interpersonal relationships is as follows. First, to engage in reciprocal relations of the sort Strawson discusses (e.g., friendship, mature love, etc.), we must, *inter alia*, expect of our friend (lover, neighbor, co-worker, etc.) that she will show us good will and respect. In other words, we must hold our friend to the basic demand of good will and respect. Undoubtedly, this is an aspirational ideal of sorts—quite often we know of ourselves and our friends that we will sometimes fail to live up to the basic demand. 23 But what’s important for Strawson is that, however haltingly, we actually do hold ourselves and our friends to this demand. This suggests the first plank in Strawson’s argument:

1. When we interact with another agent S as a friend, we expect of S that she will show us good will and respect.

But as Strawson says, to demand or expect of another that she treats you with good will and respect *just is* the proneness to the reactive attitudes. Building on this point, R. Jay Wallace argues that expectations of this sort—i.e., *normative* expectations—are constitutively tied to the reactive attitudes. 24 And though I can’t offer a full defense of this claim here, it is initially plausible. After all, if I’m, for example, playing Monopoly with a friend, I’m not really expecting her to follow the rules of the game if I knowingly allow her to collect $300 every time she passes GO. To genuinely *hold* her to the rules of the game, I must be prepared to address her unscrupulous playing. Similarly, in more straightforwardly moral domains, I’m not really holding someone to the expectation that they show persons good will and respect if I’m not disposed to address their wrongdoing. And because blaming

---

23 I am grateful to Martha Nussbaum for emphasizing this point to me.

24 See R. Jay Wallace [1994] for a full defense of this claim.
attitudes like resentment and indignation are plausibly thought to be the most basic forms of moral address,\textsuperscript{25} our proneness to these attitudes is inextricably tied to the activity of holding persons to moral norms. Hence,

\begin{enumerate}
\item Expectations for good will and respect are constituted by our proneness to the reactive attitudes.
\end{enumerate}

And from (1) and (2), it seems to follow:

\begin{enumerate}
\item Therefore, when we interact with another agent $S$ as a friend, we are prone to engage with $S$ via the reactive attitudes.
\end{enumerate}

The upshot of (3) is, of course, that:

\begin{enumerate}
\item Therefore, \textit{justified} engagement with another agent $S$ as a friend presupposes that $S$ is an \textit{apt} targets of the reactive attitudes.\textsuperscript{26}
\end{enumerate}

In other words, it would also follow from these claims that if we are not disposed to engage another agent via the reactive attitudes, then we cannot consistently stand in a friendship with that individual.\textsuperscript{27}

\textsuperscript{25} Watson [1987] calls them "incipient" forms of moral address. Darwall [2012], takes it further, argument that the attitudes of resentment and indignation constitute forms of moral address.

\textsuperscript{26} I develop this argument in more detail in Coates[2012].

\textsuperscript{27} Of course, it doesn’t require an explicit argument to appreciate this latter point. We need simply imagine what it would be like trying to be friends with someone who isn’t even \textit{disposed} to respond to your good will (or ill will) with gratitude (or resentment). The former failing would undoubtedly lead you to regard the person as too entitled, as never seeing our contributions to the friendship as mattering. And the latter failing would undoubtedly lead you to regard the person as a pushover, someone to take advantage of or (if you’re cut from a nicer stock), someone to engage with paternalistically, since they never stand up for themselves. This further suggests that if someone is not even disposed to be grateful when you show her good will or if your ill will cannot possibly exercise her resentment, it’s hard to see how you could be genuine friends. Indeed, if she isn’t even disposed to those sorts of emotional responses, she is failing to regard your actions as having much significance for her. But of course, when you’re genuinely friends with someone, their actions have a great deal of significance for you; at least, their actions aren’t simply things that you stoically endure. Thus, a form of emotional engagement that involves susceptibility to the reactive emotions seems necessary for mature, intimate friendships of the sort we take ourselves to enjoy, since our susceptibility to the reactive emotions grounds our commitment to the norms internal to such relationships.
But recall that for Strawson, the property of being morally responsible is understood in terms of the reactive attitudes. Specifically, the Strawsonian theory of moral responsibility holds that:

5. To be an apt target of the reactive attitudes just is to be a morally responsible agent.

So from (4) and (5), we can conclude:

6. Justified engagement with another agent $S$ as a friend presupposes that $S$ is a morally responsible agent.

And this point generalizes, since it’s not just friendship that seems to depend on seeing others’ actions as manifesting good or ill will. Rather, this seems to be a fact of all reciprocal interpersonal relationships, since like friendship, all reciprocal relations seemingly require mutual expectations for good will and respect. Strawson then, shows that ordinary reciprocal interpersonal relationships essentially involve emotional engagement through the reactive attitudes. Strawson thus concludes that:

7. Our ordinary reciprocal interpersonal relationships are responsibility-entailing relationships.

In other words, that we instantiate the property of being morally responsible is a condition for the possibility of our ordinary interpersonal relationships.

* * *

Even if the above argument is sound, however, it does not yet show our responsibility practices to be justified. Of course, Strawson doesn’t just leave the issue here. He uses (7) as a crucial premise in a further argument, the conclusion of which purports to show our responsibility practices to be justified given “the facts as we know them.” And because the falsity of causal determinism is not among these facts, it is irrelevant to the justification of our practices.

And as we saw earlier, this further argument turns on Strawson’s claim that:
8. Our commitment to ordinary interpersonal relationships is inescapable from the practical perspective we take on the world as agents.

Together with (7), he takes this to entail:

9. Our responsibility practices are justified.

Of course, it’s natural to wonder how (7) and (8), even when it is not mistakenly understood as a naturalistic claim, can entail (9). To begin to see why one might think that (9) follows from (7) and (8), consider the following argument, which is familiar from debates in metaethics.²⁸

10. The constitutive norm of activity is such that an action is good qua action only if it conforms to the Categorical Imperative.

11. Activity is practically inescapable for us.

12. Thus, we are bound by the constitutive norm of activity, viz., the Categorical Imperative.²⁹

Now, I don’t mean to be defending this argument as sound. And I certainly don’t think Strawson needs to accept it. But this style of argument, specific to Kantian constitutivism, is roughly parallel to the style of argument that I take Strawson to be advancing.

In the above argument, the Kantian constitutivist tries to show that because some activity is practically inescapable for us (in this case, the most general activity—activity itself), the presuppositions of that activity—viz., constitutive norms of that activity—have normative authority for us.


²⁹ This argument is most reminiscent of Christine M. Korsgaard’s Kantian-inspired account of the source of normativity. See Korsgaard [1996, 2009].
Crucially, then, the constitutivist bootstraps the normativity of morality from the presuppositions of an activity in which we are all inescapably engaged in. But this is precisely what Strawson is attempting to do when he appeals to the practical inescapability of our commitment to interpersonal relationships. He is, in effect, attempting to bootstrap the justification for our responsibility practices from the fact that we are inescapably committed to interpersonal relationships. A consequence of this move, if it is indeed successful, is that “the facts as we know them”—facts about what it is to have a practical standpoint—are sufficient to justify our responsibility practices.

Of course, this invites the important question: does this sort of bootstrapping get us the grand conclusion that Strawson wants—viz., that our responsibility practices are justified given the facts as we know them? I’m (at least) sympathetic to the idea that it does. To see this, consider that for some activity to be practically inescapable for an agent, it must be the case that that agent cannot, on pain of practical self-contradiction, fail to engage in that activity. So our commitment to participation in interpersonal relationships is practically inescapable only if, were we to wholly abandon such relationships, we’d necessarily be engaged in self-contradiction. And plausibly, wholesale rejection of interpersonal relationships does lead to self-contradiction. Since consistency demands that we engage in ordinary interpersonal relationships, it seems, then, that these relationships are justified. And because these ordinary human relationships are responsibility-entailing relationships, it follows that the responsibility practices which are presupposed by these relationships are also justified.

This argument for the legitimacy of our responsibility practices is, I think, the heart of Strawson’s “Freedom and Resentment.” Of course, as is the case with other transcendental arguments, we might plausibly worry about whether it really is the case that failing to regard ourselves participants in interpersonal relationships leads to self-contradiction. If it is not, then our commitment to ordinary human relationships, thoroughgoing and deep as it might be, is not practically inescapable, and so it cannot serve as a premise in an argument of the sort I outlined above. And Strawson has given us no real reason, beyond his insistent assurances, that this is the case. In what follows, then, I want to offer a further argument for this claim. And though it is not found in Strawson, it is, I think, a friendly amendment.

***
To begin, consider what it means to have a practical point of view. Plausibly, it entails that:

13. From the practical point of view, we necessarily see ourselves as agents.

But in seeing ourselves as agents, we don’t see ourselves as mere causes. That is, we don’t take ourselves to merely be effects-makers. Rather, we see ourselves as creatures that make causal contributions to the world that are expressive of our own agency. Thus,

14. From the practical point of view, we regard our actions as attributable to us in some deep sense.

But an action isn’t attributable to me (in this deep sense) solely in virtue of its causal etiology—that, for example, I played a causal role in its history. Rather, an action is attributable to me in virtue of normative properties of that action—properties like whether I should feel prideful or ashamed for that action, whether it licenses you to feel grateful towards me or resentful, and so on. Hence,

15. An action is attributable to an agent if and only if, in virtue of performing that action, that agent is an apt target of the reactive attitudes.

So, putting it all together:

16. So, from the practical point of view, in seeing ourselves as agents, we are committed to seeing ourselves as apt targets of the reactive attitudes.

But if it’s true that, from the practical perspective we are rationally committed to seeing ourselves as apt targets of the reactive attitudes, then, on pain of self-contradiction, it seems that we are committed to being at least of being candidates for justifiable interpersonal relationships. And we cannot consistently regard ourselves and others as apt candidates for interpersonal relations while simultaneous abandoning these relationships and thereby deeming no one to be suitable for interpersonal relations. It seems plausible,
then, that our commitment to ordinary interpersonal relationships is inescapable from the practical point of view because part of what it is to have a practical point of view in the first place is to regard oneself as a potential candidate in ordinary human relationships. In other words, it does seem that this strategy of argumentation gives us some reason to think that Strawson is correct when he claims that it is practically inconceivable that we abandon our commitment to ordinary interpersonal relationships.

§5. Conclusion
As we saw, (1) – (6) gave us a reason to accept (7), i.e., the claim that ordinary interpersonal relationships are responsibility-entailing relationships. And likewise, (13) – (16) give some teeth to (8), i.e., Strawson’s claim that our commitment to interpersonal relationships is practically inescapable. From (7) and (8), Strawson concludes (9), i.e., that our responsibility practices are justified.

Because Strawson’s conclusion rests only on what we know concerning ordinary interpersonal relationships, it, like other forms of optimism respects the facts as we know them. In other words, Strawson thinks that, the facts as we know them provide a sufficient basis for our responsibility practices. But unlike other optimists, Strawson grounds our practices’ justification not in the good that such practices might do in establishing a healthy social order, but in our human commitment to participation in interpersonal relationships. Thus, while respecting “the facts as we know them,” Strawson shows our practices to be justified. But unlike the pessimists, he does so without appealing to extra-relational, metaphysical considerations. The falsity of determinism, then, is not a requirement on the legitimacy of our practices. Accordingly, it seems to me that Strawson has provided us with a very subtle defense of compatibilism. And because this defense relies so heavily both on a Kantian critique of the alternatives and a distinctively Kantian transcendental argument for the legitimacy of our responsibility practices, I hope that I have shown that we will properly understand Strawson’s argument only if we interpret the infamously difficult “Freedom and Resentment” à la Kant.