Strawson's Ethical Naturalism: a Defense

Pamela Hieronymi
hieronymi@ucla.edu
June 28, 2024

I first present what I take to be Peter Strawson's "Social Naturalism," as applied to ethics. I then briefly present the way in which this Naturalism allows Strawson to resist skepticism about moral responsibility, as argued in "Freedom and Resentment."[1] Strawson's way of resisting skepticism makes plain that the view is open to another challenge: it seems to entail an objectionable relativism. I have provided a response to this challenge, on Strawson's behalf, in the final chapter of *Freedom, Resentment, and the Metaphysics of Morals*.[2] Here I expand upon that response.

1. Strawson's Social Naturalism Applied to Social Morality

I begin with a presentation of Strawson's Social Naturalism, as applied to ethics. Although the most thorough statement of his Social Naturalism appears in his Woodbridge Lectures, *Skepticism and Naturalism*, delivered in 1983, it can be seen in his much earlier "Social Morality and Individual Ideal," from 1961.[3] There Strawson points out that the mere existence of a working human society itself ensures the existence of some system of rules which are typically followed. I take it these are very minimal rules against murder, deceit, theft, and the like. As he puts it,

> … it is a condition of the existence of any social organization, any human community, that certain expectations on the part of its members should be pretty regularly fulfilled; that some duties, one might say, should be performed, some obligations acknowledged, some rules observed. We might begin by locating the sphere of morality here. It is the sphere of observation of rules, such that the observance of some such set of rules is the condition of the existence of society. (5)

[1] (Strawson 1962) Citations here are to the reprint in (Hieronymi 2020).

[2] (Hieronymi 2020)

[3] (Strawson 1961, 1985) It is worth noting that, in its preface, H. L. A. Hart thanks Strawson for reading the manuscript of his 1961 book, *The Concept of Law*. Much of Strawson's picture of the rules that regulate society resonates strongly with Hart's. Of particular interest to our topic is Hart's discussion of moral criticism at (Hart 1961, 183). I am grateful to both T. M. Scanlon and Michael Thorne for drawing my attention to Strawson's debt to Hart.

Strawson here presents transcendental argument from the existence of a working society to the satisfaction of the necessary conditions for its existence, namely, a system of rules that are "pretty regularly" followed. Like all transcendental arguments, it provides us with grounds for confidence in the existence of such a system. But it does not explain its nature or origins.

A year later, in "Freedom and Resentment," Strawson talks about a "framework" of expectations and demands that is "given with the existence of society." (131) Those expectations and demands, he says, are identical with our proneness to what he calls "reactive attitudes"—attitudes such as resentment, indignation, and gratitude.

To better understand this, we first need to better understand what Strawson has in mind by "reactive attitudes." He identifies them by first noting what he calls a "commonplace," namely,

> the very great importance that we attach to the attitudes and intentions towards us of other human beings, and the great extent to which our personal feelings and reactions depend upon, or involve, our beliefs about those attitudes and intentions. (111)

To illustrate, he points out,

> If someone treads on my hand accidentally, while trying to help me, the pain may be no less acute than if he treads on it in contemptuous disregard for my existence, or with a malevolent wish to injure me. But I shall generally feel in the second case a kind and degree of resentment I shall not feel in the first. (112)

Such cases, he points out, highlight,

> … in how much of our behavior the benefit or injury [to others] lies mainly or entirely in the manifestation of the attitude [of contempt, indifference, or good will] itself. So it is with good manners, and much of what we call kindness, on the one hand; with deliberate rudeness, studied indifference, or insult, on the other. (112)

To restate Strawson's commonplace: we care, not only about how others affect us materially, so to speak—whether our hand is damaged—but also, and often more, about how others think about us. We care about how we figure into one another's worlds, the esteem or disesteem with which others hold us, whether they accord us a basic degree of respect and good will, and whether they guide their own behavior accordingly. I will add: we care, too, about how we treat others, whether we accord others a basic degree of respect and good will, and whether others recognized that we do.[4] We care, in short, about standing in relations in which mutual regard is mutually recognized.[5]

More, the commonplace, the fact that we expect a some degree of respect and good will, is manifest in our reactions when those expectations are violated, in the "kind and degree of resentment" that is felt when someone treads on your hand from contempt or disregard. Such resentment is an example of what Strawson calls a "participant reactive attitude."[6] It is an attitude that is a reaction to "the quality of others' will towards us, as manifested in their behavior." (121) Such attitudes come in different forms: the "personal" reactive attitudes, such as resentment, are a reaction to the quality of someone's will towards you. The "impersonal" reactive attitudes, such as indignation, are a reaction to the quality of someone's will towards some third party. And, the "self-directed" reactive attitudes, such as guilt or remorse, are a reaction to one's own quality of will towards another.[7]

The reactive attitudes contrast, usefully and illuminatingly, with what Strawson calls a "more objective" attitude, an attitude we might adopt toward objects and impersonal events. It can be useful to consider two contrasting classes of attitude, here (though Strawson does not do so). "Objective" attitudes would be responses such as frustration or relief, which we might have to

[4] That this should be included in the "commonplace" is suggested by the "self-directed" attitudes of the next paragraph. I am grateful to Michael Thorne's careful reading, which drew my attention to the need to explicitly add this concern.

[5] This concern is also at the heart of T. M. Scanlon's contractualism (Scanlon 1998), though, for Scanlon "regard" will require according equal status or standing. See, e.g., (Hieronymi 2011). Strawson assumes only some reciprocity.

[6] The word "reactive" is somewhat unfortunate. Clinical psychology sometimes uses "reactive" to pick out responses to others that are, so to speak, "knee-jerk" (essentially unreflective) and that do not respect what such clinicians might refer to as "boundaries." "Reactive" attitudes, so understood, are problematic in interpersonal relationships. I do not believe this is what Strawson has in mind. Rather, in using "reactive," I believe he means to be noting that these attitudes are not voluntary forms of treatment (and so cannot be well understood as forms of sanction or punishment). In any case, I believe their non-voluntariness does not entail reactivity in the problematic sense. Thanks to Hannah Pickard for conversation.

[7] A reactive attitude is *p*'s reaction to *p*'s beliefs about or perception of the quality of *q*'s will towards *r*. If *p*, *q*, and *r* are different people, the reactive attitude is impersonal. When *p* and *r* are the same person, the attitude is personal. When *p* and *q* are the same person, the attitude is self-directed.

events and states of affairs we believe were not willed by anyone. While you might be frustrated to find your tire is flat, you do not resent the tire for losing air (or, if you do, you recognize this as a mistake). Though you may be disappointed when the strap on your bag breaks, you do not feel betrayed. If an unsteady board bears your weight in a time of need, you feel relieved, not grateful. If, on the other hand, you believe that the tire was flattened, the strap broken, or the board supported by someone, on purpose, with you in mind, then you might resent, feel betrayed, or feel grateful. Resentment, feelings of betrayal, and gratitude are reactive attitudes. When contrasted with these, frustration, disappointment, and relief are objective attitudes.[8]

With that much understanding of the reactive attitudes, we can turn to examine the framework of expectations and demands given with the fact of human society. Strawson's thought is that the framework of expectations and demands—the set of norms or standards—that are required for and to a large extent constitute human society are (not norms that were in away way written down, handed down, or chosen, but rather are) constituted by, or of a piece with, our readiness to react with this set of attitudes when those norms are violated. The norms are established and constituted by our readiness to react to what we perceive to be the quality of others' wills towards ourself and others, and whether their will is of good or poor quality depends on whether they violate those norms. As Strawson says, in "Freedom and Resentment," "the making of the demand *is* the proneness to the attitude." (129)

The relevant contrast, here, is not with a society with different set of reactive attitudes, but rather with a society which utilized *only* some *other* way of establishing and securing widespread conformity with the minimal norms required for its existence—say, one that utilized only a system of surveillance and sanctions, , or only desire for reward and fear of punishment in the afterlife, or only concern about karma in this or subsequent lives, rather than interpersonal reactions grounded

[8] I make these same points with these same examples, etc., in (Hieronymi 2020, 7–9).

in the commonplace. Such a society, I believe Strawson would say, would not be recognizably human.[9]

We can notice, further, that appeal to the commonplace, as opposed to surveillance and sanction, opens the door to what we might call "moral," rather than merely prudential, reasoning: it allows us to guide our actions while thinking, not merely about how it will impact our own interests, narrowly understood, but also about how it will impact our relations to others.

We can now surmise that the *content* of the demands in question—what is demanded—will be given by the apt triggering conditions, so to speak, of this set of attitudes. And those triggering conditions concern the quality of another person's will. The "quality of will" to which your attitudes react just is another's willingness to heed the norms established by the reactive attitudes, themselves—including their concern for how you perceive their quality of will.

Put this way, the account might seem circular, but it is not. It is, instead, formal, awaiting a further story about what gives content to the norms, which is to say, what gives content to the expectations and demands. Given content (avoid deceit, offer assistance when it is easy to do so, do not cause gratuitous pain, etc), the formal account claims, first, that the social reality of those expectations and demands is identical to our readiness to respond with reactive attitudes when they are violated (or superseded), and, second, that those reactive attitudes react, not merely to behavior as such, but rather to the quality of will manifest in it. The account is formal, but not circular.

Importantly, Strawson recognizes that both the particular demands we make (say "please" and "thank you") and the particular attitudes we adopt (resentment, say, rather than some special form of disappointment) are culturally contingent. The existence of a working society does not entail *our*

[9] Strawson's picture of human nature contrasts strongly with that imagined by Glaucon early in Plato's *Republic*: Glaucon imagined that anyone with the Ring of Gyges would misbehave wildly, and so seemed to imagine that only the risk of discovery deterred such misbehavior. (Plato 1992) Strawson suggests, instead, that his central commonplace—the way in which we care about the kind of relationships in which we stand—will, itself, constitute the framework of norms required for the existence of society. That framework can, of course, be supplemented by a system of surveillance and sanction, or belief in karma, etc. (and it seems that it has always needed to be so supplemented).

system of attitudes nor *our* expectations and demands—the view is not so parochial. Rather, the existence of a working society only ensures that there is *some such* system in place, something playing the role. Strawson would, I think, be happy to allow a recognizably human, working society that is constituted by a framework of demands for respect or regard that is, in turn, constituted by a different set of reactive attitudes and that required a different set of thoughts and behaviors.[10]

Pulling these ideas together, we can see that Strawson offers us a picture of what we can call "the natural form of human sociability." The existence of a functioning human society ensures the existence of a minimal framework or system of norms, where those norms are constituted by our proneness to the reactive attitudes. Our reactions are triggered by the quality of others' wills regarding those norms—i.e., showing respect or disrespect. We can say that human society, so understood, is based on Strawson's commonplace. In fact, we can say that this commonplace grounds what we can call the human form of sociability: we live together and abide by a system of norms because we care about whether others respect us, about whether we respect others, and whether others think that we respect them. That we care and so live is a "natural" fact (analogous to using language, reasoning inductively). (And thus Strawson, as I am interpreting him, is indulging in some armchair anthropology and sociology. I will be drawn into this highly dubious practice, myself, momentarily.)

More, as a natural fact about us, our form of sociability, Strawson argues, is not, itself, open to skeptical doubts. Yet, this form of sociability sets up a system *within which* justifications and criticisms can be made. The particulars of the system itself, its *content*, can be criticized and changed.

[10] See, e.g., (Goldman 2014).

Only the fact that it takes this form is immune from doubt.[11] Here Strawson evokes Wittgenstein ——"it is difficult to begin at the beginning. And not try to go further back."[12]

2. Strawson's Defense of Social Morality against Skeptical Doubt

I turn, now, to "Freedom and Resentment," where Strawson's defends our form of sociability against one kind of skeptical doubt, a doubt raised by the apparent incompatibility of free will and physical determinism. The skeptic claims that, if determinism is true, then we should not be engaged in this form of interaction at all—if determinism is true, then we have reason to abandon the reactive attitudes, ceasing to respond to the quality of others' (and our own) wills in this way. Or, if we cannot abandon them, if we are stuck with these reactions as a natural fact about our psychologies, we should at least acknowledge that we are trapped in an unjustifiable set of responses.

Strawson's most basic reply is simply to assert that our form of sociability—the fact that we live in a society held together by norms constituted by such reactions—is a natural fact about us, and thus it is not the sort of thing that can be shown to be either justified or unjustified. Only the particulars within the framework can subject to such evaluation.

However, the fact that the particulars, within the system, can be shown to be unjustified seems to open a route for the skeptic, often called "the generalization strategy." The skeptic can argue that certain conditions that, within the system, show it unjustified to hold certain people responsible—that is, unjustified to respond to them with these attitudes—will, if determinism is true, be true of everyone all the time: those already-recognized conditions will generalize. For example, it seems inappropriate to hold someone responsible for their impulsive behavior, if that behavior is due to an inhibitory control disorder. The disorder, it seems, exempts them of responsibility. The skeptic

[11] Questions can be raised: How to make precise the metaphor of within and without? What is form and what is content?

[12] Quoted by Strawson in (Strawson 1985, 24)

then argues that, if determinism is true, then everyone is always, in the relevant way, like the person who is subject to the inhibitory control disorder.[13]

What is this "relevant way"? In what way are all of us like the person suffering from an impulse control disorder, if determinism is true? Answers vary from skeptic to skeptic: our behavior is explained by our physiology, is out of our control, is not up to us, or has a source outside of us. Whatever the "relevant way," the basic strategy is clear: by understanding the already-recognized case (e.g., of an inhibitory control disorder) in some way that will generalize, if determinism is true, the skeptic can argue that, if determinism is true, then the system is unjustified on its own terms, so to speak: it condemns itself from within.

Strawson does not agree. In "Freedom and Resentment" he argues against the generalization strategy in an interesting, though difficult to follow, way.[14] He grants that the system contains, within it, certain conditions under which we suspend the reactive attitudes. Such case have come to be called, in the literature, cases of "exemption," but I have come to think this a misleading, and sometimes dangerous, label. I will say that these are cases in which the quality of a person's will ceases to "matter," to us—though, crucially, I am here using "matter" as a technical term: the quality of will does not cease to matter, *full stop*, but simply in the sense that we cease reacting with reactive attitudes and move to a more objective attitude. (I will put single quotation marks around 'matter' to remind us this is a technical term.) Strawson lists some such cases: when someone is under great stress or is suffering from mental illness, or when we are interacting with a small child.

In fact, Strawson divides these cases in into three classes.[15] The first are cases of abnormal circumstances, such as having a bad day or being under great stress. He dismisses these quickly, saying, "We normally have to deal with [a person] under normal stresses; so we shall not feel towards

[13] It is trickier to find examples for the generalization strategy than is typically realized. See (Hieronymi 2020, 40–41).

[14] This defense, given in "Freedom and Resentment," is not repeated in *Skepticism and Naturalism.*

[15] Strawson first separates cases in which we were mistaken about the quality of will from those in which the quality of will ceases to 'matter,' and then he divides the latter into three classes. See (Hieronymi 2020, 9–11)

him, when he acts as he does under abnormal stresses, as we should have felt towards him had he acted as he did under normal stresses." (115). The second sort of case "allows that the circumstances were normal, but presents the agent as psychologically abnormal—or as morally undeveloped" (115). These are cases of mental illness, unfortunate formative circumstances, and immaturity. Then, having presented these first two classes, he notes what he calls "something curious to add to this," namely, that we sometimes step away from the reactive attitudes on purpose, so to speak, for a variety of reasons—he says:

> The objective attitude is not only something we naturally tend to fall into in cases [of] abnormalities or immaturity. It is also something which is available as a resource in other cases, too… we *can* sometimes look with something like the same [objective] eye on the behavior of the normal and mature. We *have* this resource and can sometimes use it—as a refuge, say, from the strains of involvement; or as an aid to policy; or simply out of intellectual curiosity. (116)

When we use our "resource" we do not merely "naturally fall into" the objective attitude, but we adopt it for various purposes, more-or-less at will.

Notice that, when Strawson characterizes the cases in which we "naturally fall into" the objective attitude, he does so by noting what is "normal," "abnormal," "immature," or "undeveloped." He later characterizes these as cases in which a person is "incapacitated in some or all respects for ordinary interpersonal relationships"(119). Being incapacitated for ordinary interpersonal relationships is, for Strawson, that which explains why we move to a more objective attitude. We do so, not because that behavior is caused by their physiology or is out of their control or not up to them or has a source outside of them, but rather because the disorder renders them incapable for ordinary interpersonal relationships (across that certain range of behavior).

By "ordinary," I have argued, Strawson here means something like "statistically ordinary"—and it is on this basis that he can argue against the generalization strategy.[16] He can argue against it because, as he puts it, "it cannot be a consequence of any thesis which is not itself self-contradictory

[16] This interpretation of "ordinary" is both the key to, and what is distinctive about, my interpretation in (Hieronymi 2020). It is defended there.

that abnormality is the universal condition." (118)[17] Given that cases in which we move to a more objective attitude are those in which the person is not capable of (statistically) ordinary interpersonal relating, we know, in advance, that nothing true of all of us will be a reason to move to a more objective attitude—not only because we already know that most of us *are* so capable, but also because it could not be the case that everyone is an outlier.

We can now provide, on Strawson's behalf, a diagnosis of what went wrong with the generalization strategy: when we move to a more objective attitude in the case of, for example, someone with an inhibitory control disorder, our reason for doing so is not, as the generalization strategist would have it, that the person's behavior is determined in some way that might be true of all of us. Rather, our reason is that (the person's behavior is determined *in such a way* as to make it the case that) the person is not capable of ordinary interpersonal relating (across that range of behavior). The generalization strategist misidentified the "relevant way."

Strawson here provides a fascinating way to argue against the generalization strategy, and one that requires some reflection. Let us consider the underlying picture.

Notice that Strawson's picture seems to imply that, as what is ordinary changes, so will not only the exempting conditions but also the demands and expectations themselves—what we can call the "standards of regard." In fact, I think this implication is plausible.

To illustrate, consider drunkenness. Drunkenness is, in our society, a condition under which we often shift to more objective attitudes.[18] But suppose everyone in our society was naturally equipped with only the degree of inhibitory control, attention, and memory that we now exhibit when fairly intoxicated. In such a society, our expectations of and interactions with one another would look, from the outside, much the same as our current expectations of and interactions with people who are fairly drunk: we would not respond to certain outbursts with indignation or resentment; we

[17] For the argument that this is Strawson's argument, see (Hieronymi 2020).

[18] I consider this analogy in more detail in (Hieronymi 2020, 31–31, especially note 14).

would not expect our requests to be reliably remembered; we would not expect certain indulgences to be resisted.[19] But notice, it would not be the case that we would be treating everyone as we now treat the drunken: we would not be suspending, all the time, the usual expectations and demands, taking up a more detached, objective attitude and so overlooking constant disrespect or disregard. Rather, in this society, certain behaviors that we now normally regard as disrespectful, but which we might overlook given intoxication, would then simply be an unremarkable part of life. They would not be considered disrespectful. That is to say, in this society, *the content of the expectations and demands* would have shifted. In such a society, those outbursts, absent-mindedness, and indulgence would not be regarded as instances of ill will or disregard—they are no longer even prima facie wrong. The standards within the system thus shift, to match the ordinary capacities.

Thus, it seems, if we were all gradually to lose some of our ability to self-regulate and so to become as we are now when fairly intoxicated, the expectations and demands to which we hold one another would adjust: certain things we now regard as rude or disrespectful would cease to be so. The standards would shift to accommodate our new limitations.

We can note that this picture allows for subcultures with different standards of regard—perhaps elementary school classrooms, homes for the memory impaired, frat houses, finishing schools, or the mafia. The standards of regard—that is to say, the proneness to the reactive attitudes—can shift with the population.

This shiftiness, of course, raises the specter of relativism. But before taking up that objection, let us revisit how the picture allows Strawson to foil the generalization strategy:

On Strawson's picture, though we sometimes suspend the reactive attitudes, we could not have reason to so towards everyone all the time, because, given our natural human commitment to

[19] Given our lack of impulse control, we might also, of course, respond more quickly to a perceived offense. Even so, doing so would not seem to be responding *unusually* quickly, given that the condition is widespread, and so the quick response would not, itself, be remarkable. Thus it seems that our interactions would settle into something similar to (though not identical to) the current interactions between the drunken and the sober.

engaged interpersonal relating, we have reason to do so only in unusual or outlier cases. We can support this with two thoughts. First, given our *commitment* to engaged relating, any condition true of everyone would be accommodated with a shift in our expectations. The condition would then not trigger the reactive attitudes; there would be no need to suspend. Contrariwise, if we *were* to suspend the reactive attitudes towards everyone always, that would amount to dropping all demands and expectations and so exiting the natural human form of sociability—which we could not have reason to do. Thus, no "exempting" conditions, no conditions on the aptness of our proneness to the reactive attitudes, will generalize. Our natural commitment to ordinary relating means we not only that we will not "exempt" everyone, but also that the framework given by human sociability will not give us reason to do so—it will not show that we do not 'matter.' The skeptical doubt is idle.

Of course, the idea that we will accommodate anything true of everyone with a shift in the standards opens the picture to the charge of relativism. It implies that behavior will cease to be disrespectful simply because it is widespread.

3. Relativism

I will now try to address the charge of relativism. We can begin by noting that not all relativism is objectionable: I believe the shifts to accommodate natural limitations would not be objectionable, and some sort of relativism is important to accommodate cultural variations in social standards.[20]

However, some forms of relativism are objectionable; an acceptable picture of the nature of morality must (at a minimum) allow moral claims to have what I will call "critical purchase." That is to say, a picture of the nature of morality must not merely describe actual social norms and their workings; it must also both allow critical evaluation of those norms and leave open the possibility that norms could be properly criticized even if widespread or dominant.

Mere descriptions of the "moral system" at work in a society or culture, such as those that might be offered by an anthropologist or sociologist, do not provide critical purchase. On the basis of

[20] I find the final chapter of (Scanlon 1998), titled "Relativism," particularly illuminating.

their observations, the social scientist might correctly say that some action (or attitude, or practice) is "disrespectful in culture *x* (or at time *t*)," and yet "respectful in culture *y* (or at time *t'*)." If we had only such descriptions, based only on observations of how people act and react in a given culture or at a given time, the resulting picture of the nature of morality would be objectionably relativistic. We would have no ability to criticize the norms at work in a culture. We would have no critical purchase.

Insofar as the standards of regard adjust to what is statistically ordinary, they might seem to be objectively relativistic in the way such descriptions are: the standards might now seem simply to reflect what is ordinary in that culture—how people typically act and react. Thus it might seem that Strawson's picture of the nature of morality leaves morality without critical purchase and therefore objectionably relativistic. However, I believe Strawson's picture can allow critical purchase.

4. Securing Critical Purchase

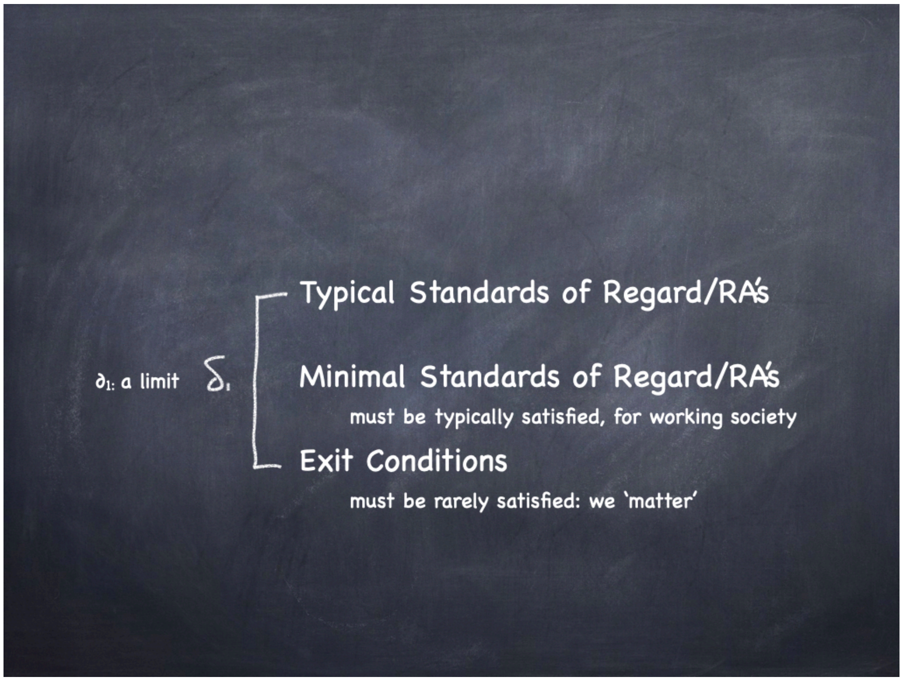
How to secure critical purchase? Strawson himself frequently mentions consistency. Appeals to consistency can indeed take us some distance, providing some critical purchase. It is a Kantian hope that they will take us the whole distance, providing adequate critical purchase. I do not share that hope.

I would instead suggest a simple and unapologetic appeal to *ideals*—that is, to the kind of thing that Strawson had in mind when writing "Social Morality and Individual Ideal." Ideals might include such things as equality, piety, the stewardship of nature, or avoiding pain in sentient creatures.

To see this response, let us first review Strawson's picture. It includes both the standards of regard—that is, the norms that are identical to the proneness to the reactive attitudes—and the "exempting" or, as I will now refer to them, "exit" conditions. These are the conditions under which participant relating becomes unworkable, which are also the conditions under which the quality of a will no longer 'matters'—under which we cease to react. Individuals "exit" when they lack the capacity for tolerably ordinary interpersonal relations. As we have seen, these cases must be rare. This is for two reasons. First, if the exiting conditions were universal, we would, by exiting everyone, abandon human society altogether—something we both would not do and could not have reason to do. Second, if a so-called "demand" or "expectation" is rarely reacted to, it thereby ceases to be a demand or expectation—the standards thereby shift.

We can now complicate the picture, as in **Figure 1**. We can distinguish between the **minimal standards of regard,** which, we have said, must be typically satisfied for the existence of any society, and what we can now call the **typical standards of regard** at work across a given society (typically above the minimum). We could further distinguish the typical standards of regard at work in *portions* of a society.

Figure 1



∂_1: a limit δ_1

Typical Standards of Regard/RA's

Minimal Standards of Regard/RA's

must be typically satisfied, for working society

Exit Conditions

must be rarely satisfied: we 'matter'

Now we can reconsider the threat of relativism. In Strawson's picture, because the **exit conditions** must be rarely satisfied, there is a kind of limit in the possible distance or difference between the typical standards and the exit conditions, marked in Figure 1 as "**∂1.**" This limit ensures that the typical standards of regard will shift, in response to changes in the typical capacities of the population.[21]

Making matters worse, it seems a naturalist should think that even what we might call "moral" capacities could undergo widespread change for the worse: the naturalist should think that these capacities can be naturally or contingently limited just as capacities for memory or inhibitory control can be naturally or contingently limited. Moral development takes time and can go badly, and some people, I am afraid, arrive at adulthood too frail of ego, or too insensitive, or too bigoted to satisfy the typical standards of regard. And so there is no reason to think an entire society might not slide downwards, so that, eventually, what was once treated as disrespect is no longer seen as disrespectful—no longer something reacted to with resentment or indignation.

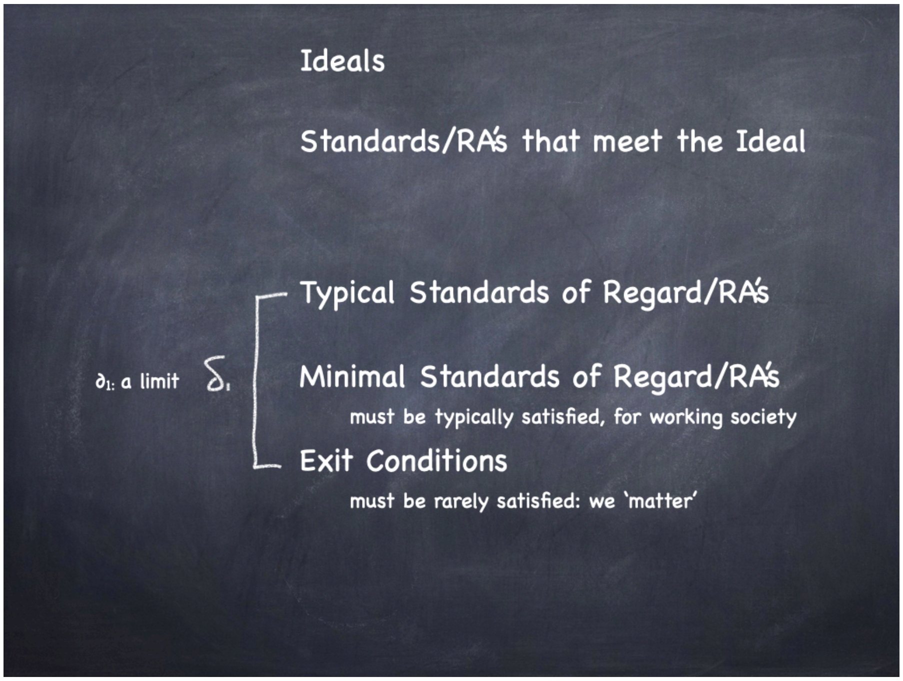
Thus $\partial 1$, together with the changeability of our capacities, generates the threat of relativism: a widespread change in capacities could, it seems, render what we now regard as disrespect no longer disrespectful.

To address this threat, I suggest a simple and unapologetic appeal to **ideals** (moving to **Figure 2**). Again, ideals might include such things as equality, ecological preservation, or avoiding pain. Such ideals will be essentially contestable and will be contested—to call something an ideal is not yet to call it a *good* ideal. But to call it an ideal is to say that it is the sort of thing that is *meant* to be a good ideal. With the ideals will come another set of possible standards: the standards of regard that would **meet**—adequately embody, or do right by—**the ideal**.

The presence of ideals will allow for critical purchase—it will allow us to criticize the typical standards of regard. An ideal of preserving nature would allow us to criticize the typical standards as destructive of it. An ideal of equality would allow us to criticize the typical standards as failing to recognize as disrespectful what, given the ideal of equality, is disrespectful. In fact, I believe that

[21] The idea of "distance" or "difference," here, is difficult, because it is between two different types of things: capacities and standards. The underlying thought is that the standards must be such that they are tolerably workable, given typical capacities. I try to flesh out "tolerably workable" below, in providing details on the dynamics that create $\partial 1$.

Figure 2



Ideals

Standards/RA's that meet the Ideal

Typical Standards of Regard/RA's

∂_1: a limit δ_1

Minimal Standards of Regard/RA's

must be typically satisfied, for working society

Exit Conditions

must be rarely satisfied: we 'matter'

allowing this kind of double-talk about respectfulness—allowing us to talk about what is disrespectful, given the typical standards at work in a society, and what is disrespectful, given the ideal of equality—is a strength of the view.[22]

5. Objections

Two objections can be raised to this naked appeal to ideals. First, one might object that ideals, as I have introduced them, are not yet sufficient for critical purchase—for that we would need the *correct* ideals.[23]

This objection misunderstands what (I believe) is required for critical purchase: we do not need to know which ideals are correct. We only need the ideals to be something on the basis of which we could aptly (if, ultimately, incorrectly) evaluate and criticize an entire society, and for that it is sufficient if the ideals can be intelligibly used to evaluate the norms at work in the society and are contestable. It is sufficient that they are the kind of thing that *might* be correct.

[22] The ideal of equality, I suspect, has a special place in this picture of the natural form of human sociability.

[23] Special thanks are due to Jeremy Fix for raising a version of this objection in commentary.

This reply is unlikely to satisfy the objector, who may now claim that such "critical purchase" is insufficient to meet the threat of objectionable relativism. But I suspect that such an objector is now confusing the threat of objectionable relativism with the threat of skeptical doubt or our own fallibility, neither of which I think an acceptable account of the nature of morality must answer.[24] The appeal to fallible ideals is, indeed, insufficient either to eliminate our fallibility, to answer a skeptic, or even to eliminate the doubt about our own ideals that might arise once we see that others disagree. But the threat of objectionable relativism is neither the threat of error, of disagreement, nor of doubt. So long as we retain our confidence in our own ideals, we can avoid objectionable relativism.

The second objection claims that these ideals, which are now divorced from the typical standards of regard at work and so divorced from the typical reactive attitudes, are too removed from the actual workings of what has been identified as social morality to show that morality, itself, allows critical purchase—we are now simply pointing to some abstracted, intellectual, pie-in-the-sky thing that may or may not have anything to do with recognizably moral attitudes and interpersonal expectations. I

[24] The objector may disagree about what is required. Sometimes ethical theory seems to be in pursuit of some argument or set of facts that would secure something like Cartesian certainty about the correct ideals and thereby eliminate the fallibility, or at least show the skeptic guilty of some error of rationality or prudence. (See, for example, (Korsgaard 1996), which seems to make roughly the same trip from ordinary conviction into doubt, to contradiction, to certainty, and then back to ordinary conviction as Descartes' *Meditations*.) Seeking such certainty seems, to me, a mistake, and, in fact, the same mistake made by certain forms of religious thought: appeal to metaphysics, reason, or self-interest as a substitute for admittedly fallible reflection about what is of value, hoping either to allay one's own doubts, answer skeptics, or secure agreement. In contrast, it seems to me that if we believe ourselves to be in possession of the one and final demonstrable truth about value, we would be both in a kind of bad faith—denying our inevitable responsibility for our own convictions—and in a compromised relation to others. We can have confidence in our own ideals even while admitting that we are not in possession of the one final truth about value.

In any case, it seems to me that the best way to defend an ideal is not to show it grounded in *something else* (metaphysics, human nature, rational nature), but rather to put it on vivid display—a task perhaps better accomplished by art than argument. In fact, grounding support for an ideal in some other kind of fact or argument risks changing the subject, and very often simply amounts to an appeal to some other ideal (consistency, say, or the avoidance of pain). (Kant noticed the problem and thought he could solve it by appeal to a "formal" constraint.) Returning to the quotation from Wittgenstein, above, it is difficult to begin at the beginning and not try to go further back.

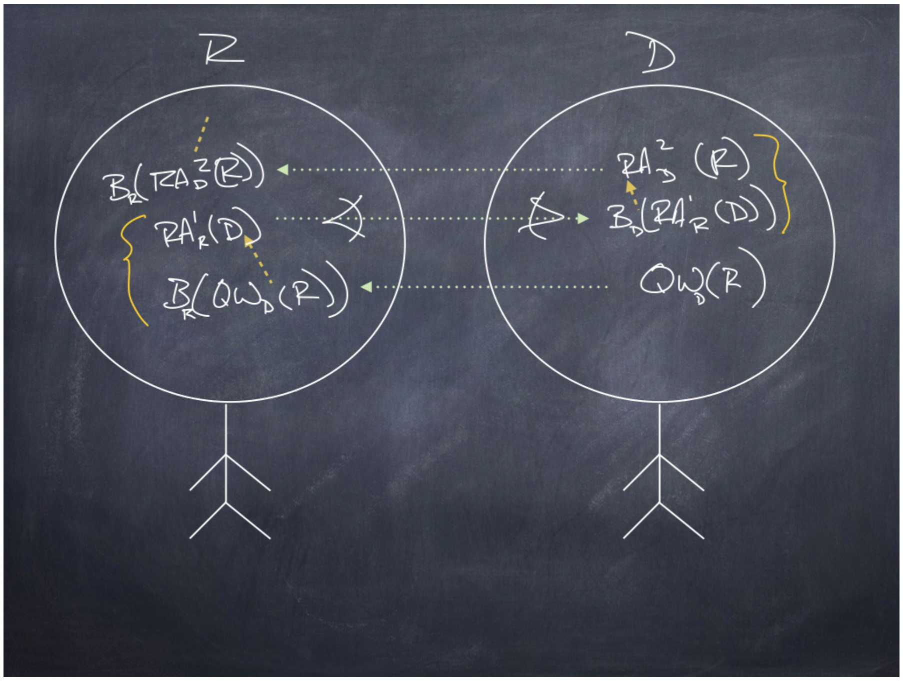
(Still, if you find such grounding important, please provide it. I do suspect that fairness or equality, as ideals, will have a special place in the picture presented—though I am not yet prepared to say how.)

now turn to address this objection by showing how, through ordinary, recognizable social dynamics, ideals will not only allow us to criticize the typical standards of regard but can also be incorporated into and thereby change those standards, becoming part of morality as understood on this picture.

(At this point I am following Strawson into indulgence in armchair social science. I am aware of leaving my lane, and I welcome input from those more informed. What I say below seems to me both plausible and illuminating; I cannot speak to its accuracy.)

To start, we can examine more closely why there might be a limit like ∂1, by focusing on the interpersonal dynamic highlighted by Strawson. For this, turn to **Figure 3**, which depicts stick people with very big heads, looking at one another. Here, person **D** (Doer) has some quality of will (**$QW_D(R)$**) towards **R** (Reactor), and R reacts to the quality of D's will towards R—or, more accurately, to R's *belief* about the quality of D's will towards R (**$B_R(QW_D(R)$**)—with a reactive attitude towards D (**$RA^1_R(D)$**): R resents D.

Figure 3



But the story does not stop there. D, in turn, reacts to R's reactive attitude, $\mathrm{RA}^1_\mathrm{R}(\mathrm{D})$—or, more precisely, to D's belief about R's reactive attitude ($\mathbf{B_D(RA^1_R(D))}$)—with another reactive attitude, now toward R ($\mathbf{RA^2_D(R)}$)—maybe D is indignant that R should resent him. And of course, R can now react to D's reaction to R's reaction—and it might go on to exhaustion.

Notice, here, that the dotted, horizontal interactions are matters of perception and belief, which can be evaluated for accuracy. Meanwhile, the dashed, vertical developments—the reactions had on the basis of those perceptions or beliefs—both embody and manifest the individual's operative standards of regard.[25] Demands of rational or intrapersonal consistency generate the yellow brackets (if you do not react as your own standards of regard would make fitting, you will feel yourself in some way inconsistent).

Notice, too, in this case, with D's *indignation* at R's resentment (rather than, say, guilt or remorse), presuming no inaccuracy (that is, no mistakes or misunderstandings), these individual's operative standards of regard are out of alignment.

The situation between D and R is uncomfortable. This discomfort is a manifestation of Strawson's commonplace: we care both about the quality of our own and others' will and about others' perception of our quality of will. We do not want them to violate standards of regard, nor do we want to do so, nor do we want them to think that we have done so. This commonplace generates at least three kinds of pressure:

First, and most obviously, it generates pressure to meet the standards of regard, or, failing that, to present yourself as having done so. Second, it generates pressure for uniform standards of regard, because disagreement about the standards will generate the conflict depicted. Third, Strawson's commonplace, our natural form of sociability, leaves us vulnerable to what Strawson calls "the strains of involvement," we can tolerate only so much disregard, malice, indifference,

[25] I am using "operative" standards of regard to refer to those that are revealed by the person's (patterns of) reactive attitudes. The label expands on T. M. Scanlon's notion of "operative reason" in (Scanlon: 1998, chapter 1).

indignation, outrage, failure to understand, or failure of empathy before we feel pressure to use our "resource" and adopt a more objective attitude. Often enough, we also generate a story, theory, ideology, or diagnosis about the other that allows us to more easily use our resource—to more easily interact with them in a more objective way. This third kind of pressure, the strains of involvement, pushes us away from participant engagement and the reactive attitudes.[26]

But, as we have seen, on Strawson's picture, a widespread lack of reaction simply *amounts to* a change in the standards of regard. Because our tolerance of interpersonal conflict is limited, and because a widespread failure to react to so-called violations results in a shift of the standards, we encounter ∂_1—the typical standards must be such that our relationships are typically tolerably workable, given the typical capacities. If the typical capacities drop, the standards must drop, or, contrariwise, if standards rise, the typical capacities must also rise. (The result is something close to ought implies can, but applied with a very broad brush, at the level of the society on the whole.)

I turn, now, to consider how the ideals can be incorporated into, and change, the standards. We can begin by distinguishing between what I will call "naturally given" limitations, such as our natural capacities for memory, attention, or inhibition control, and what we might call "socially malleable" limitations, such as our capacity to lose gracefully, avoid xenophobia, delay gratification, or respect women.[27] Let us focus first on the naturally given limitations and examine, in a bit more detail, how a natural limitation in our capacity for, say, delay of gratification, will shift the typical standards downwards.

[26] Some people seem to thrive in certain situations of conflict, apparently finding them satisfying. Such people are not then experiencing Strawson's "strains of involvement."

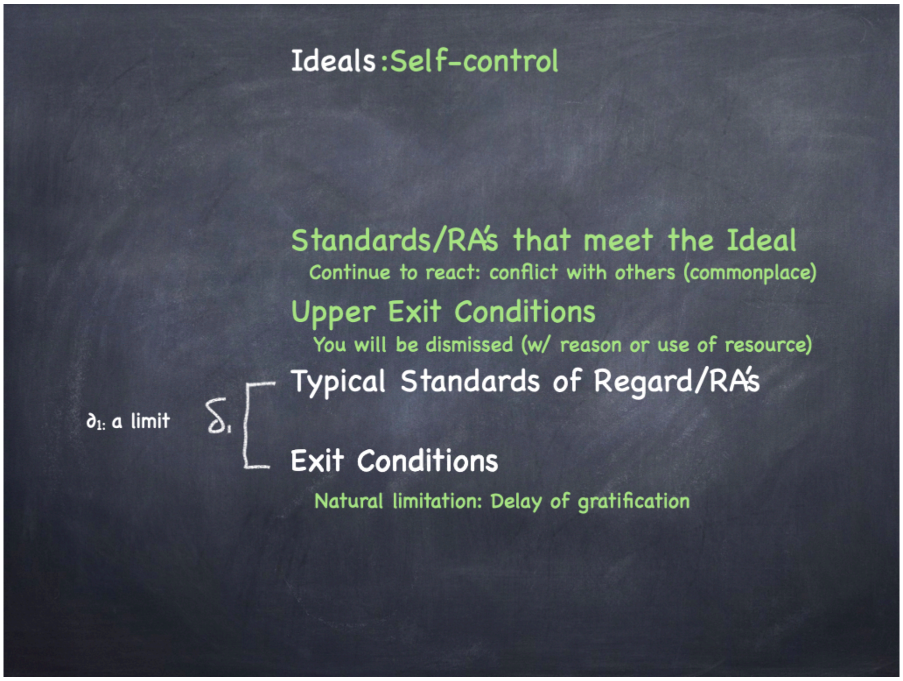
Strawson also notes the emotional and interpersonal difficulty of sustaining an objective attitude. He says, "Being human, we cannot, in the normal case, do this for long, or altogether." (117) This difficulty generates a very important fourth pressure, opposing the third. (Strawson does not, in "Freedom and Resentment," consider the ideologies that dehumanize or "other," and so, apparently, make this easier.)

[27] This distinction is not sharp, and perhaps all limitations are ultimately socially malleable. This will not be a problem for the picture, but it is easier to illustrate by starting with the idea that there are some natural limits to our capacities.

Suppose that, as our capacity for delay of gratification slips downwards, you, as a would-be reformer, try to hold out against the shift: you hold onto an ideal of self-control. That ideal will dictate a set of standards above the ones tied by ∂1 to the natural limits. You now have a few options.

First, you might continuing to engage with others, reacting with resentment, indignation, and the like, as would fit those higher standards. This will be difficult, given the commonplace: It will be *personally* emotionally costly, as those negative reactive attitudes are personally exhausting. It will be *interpersonally* costly, as it will generate interpersonal conflict. And, eventually, it will lead others to use their resource to avoid the strains of involvement with you. You will be dismissed by others. This dismissal of the would-be reformer reveals what we might call **upper exit conditions**, as in **Figure 4.**

Figure 4



Ideals:Self-control

Standards/RA's that meet the Ideal

Continue to react: conflict with others (commonplace)

Upper Exit Conditions

You will be dismissed (w/ reason or use of resource)

Typical Standards of Regard/RA's

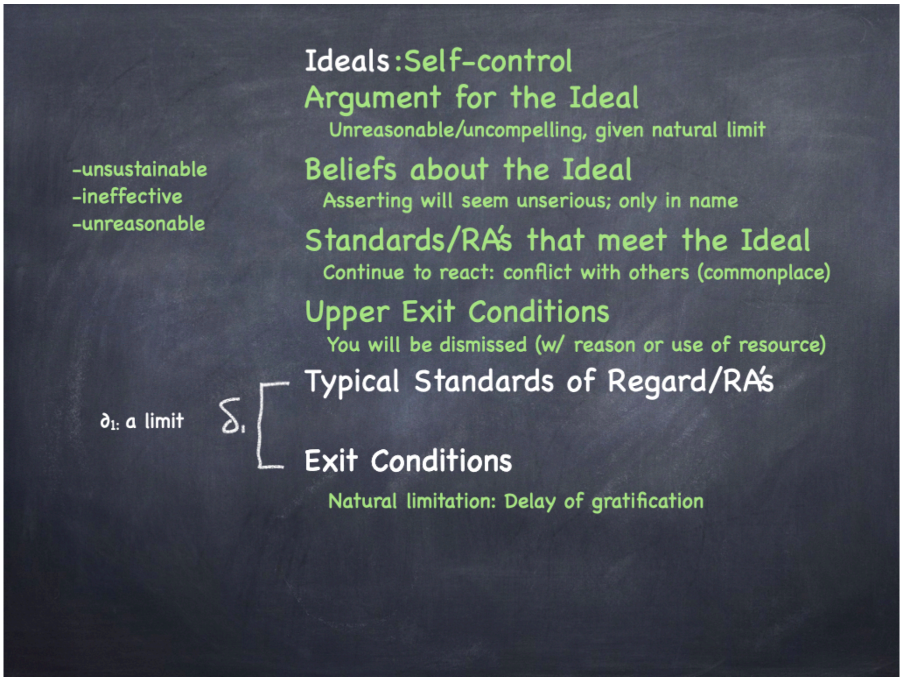
Exit Conditions

Natural limitation: Delay of gratification

How or why will you be dismissed? If you can, yourself, satisfy the higher standard of self-control, you will be regard as gifted but ungenerous, and therefore not to be taken seriously. If you cannot, yourself, satisfy the standards, you will be regarded as hypocritical, and so not to be taken seriously. Either way, others will end up using their resource and/or telling stories, and you will end up, at best, a kind of irritating curiosity with a weird hang up.

Moving now to **Figure 5**. If you understand the ineffectiveness of your emotional engagement, you might choose instead to use *your* resource, yourself, while maintaining your **belief** in correctness of the higher standards. But now we encounter the objection: this bloodless reaction will make it seem as though you are not really committed to these ideals as important. You are not serious about them. You hold them only in name. (We can think, here, of the position I believe many people find themselves in with respect to the treatment of animals.)

Figure 5



Ideals :Self-control

Argument for the Ideal

Unreasonable/uncompelling, given natural limit

-unsustainable
-ineffective
-unreasonable

Beliefs about the Ideal

Asserting will seem unserious; only in name

Standards/RA's that meet the Ideal

Continue to react: conflict with others (commonplace)

Upper Exit Conditions

You will be dismissed (w/ reason or use of resource)

$\partial_{1:}$ a limit δ_1

Typical Standards of Regard/RA's

Exit Conditions

Natural limitation: Delay of gratification

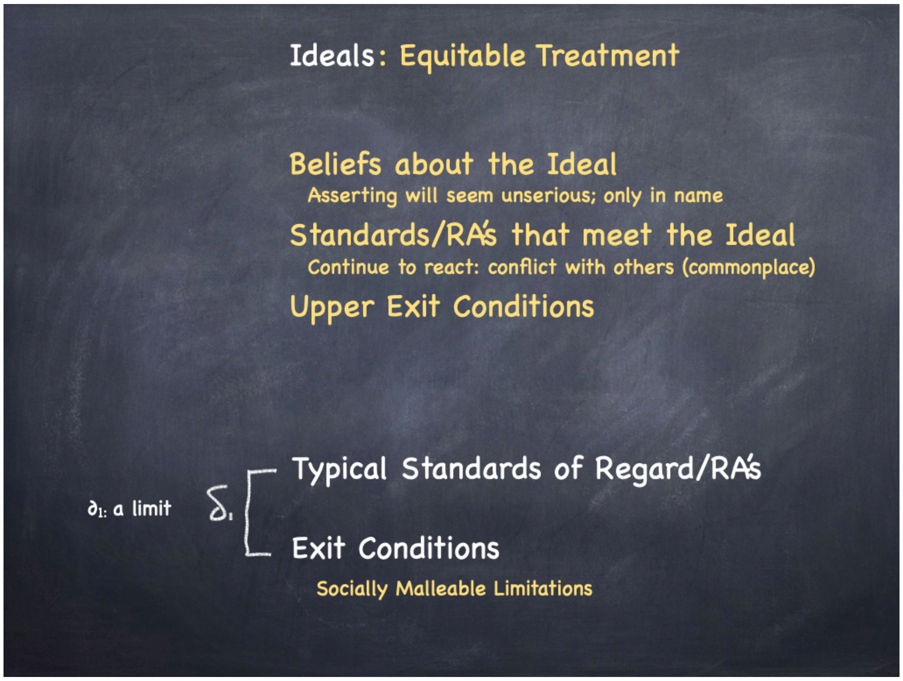
In either case (whether you engage and continue to react or instead use your resource while expressing beliefs), you may try to make **arguments**.[28] However, given that the limitation in self-control is (by stipulation) a natural one, your arguments will seem unreasonable.

[28] These arguments, as I imagine them, would appeal to the ideal. They might also attempt to establish it. See note 24.

Thus, it seems, holding out for expectations that exceed natural limitations will be neither sustainable, effective, nor reasonable. The standards will move downwards to meet the limitations.

Let us now turn to what I have called the socially malleable limitations, and so replace the capacity to delay gratification with the capacity to respect those in a subjugated group, looking at **Figure 6**.

Figure 6



Ideals: Equitable Treatment

Beliefs about the Ideal

Asserting will seem unserious; only in name

Standards/RA's that meet the Ideal

Continue to react: conflict with others (commonplace)

Upper Exit Conditions

Typical Standards of Regard/RA's

∂_1: a limit δ_1

Exit Conditions

Socially Malleable Limitations

Suppose again that, as a would-be reformer, you try to hold out against the downward shift. In this case, let us say that you are holding out for an ideal of **equality**. That ideal will, again, dictate a set of standards, above the typical ones. As before, you might continue to engage, reacting to violations of the standard with resentment and indignation. Again, this will be difficult, given the commonplace. It will be personally emotionally costly, as those negative reactive attitudes are exhausting. It will be interpersonally costly, as it will generate conflict. And, eventually, it will again lead to you being dismissed by others, or worse—you will again encounter the upper exit conditions. Others will end up using their resource or telling stories, and you will end up at best an irritating curiosity, or, more likely, a target of exile or violence.

Alternatively, you might try to use your resource while asserting your beliefs. Again, this will seem like believing without caring: being committed only in name.

Yet, in this case, there may also be significant counter-pressure that makes the difficulty of holding out more sustainable. The counter-pressure might come from one's commitment to the ideal itself, or from one's own self-interest, or from the demands of self-respect. It has been noted that resentment, or some such reaction, may be required for self-respect and may function, socially, as a kind of protest.[29]

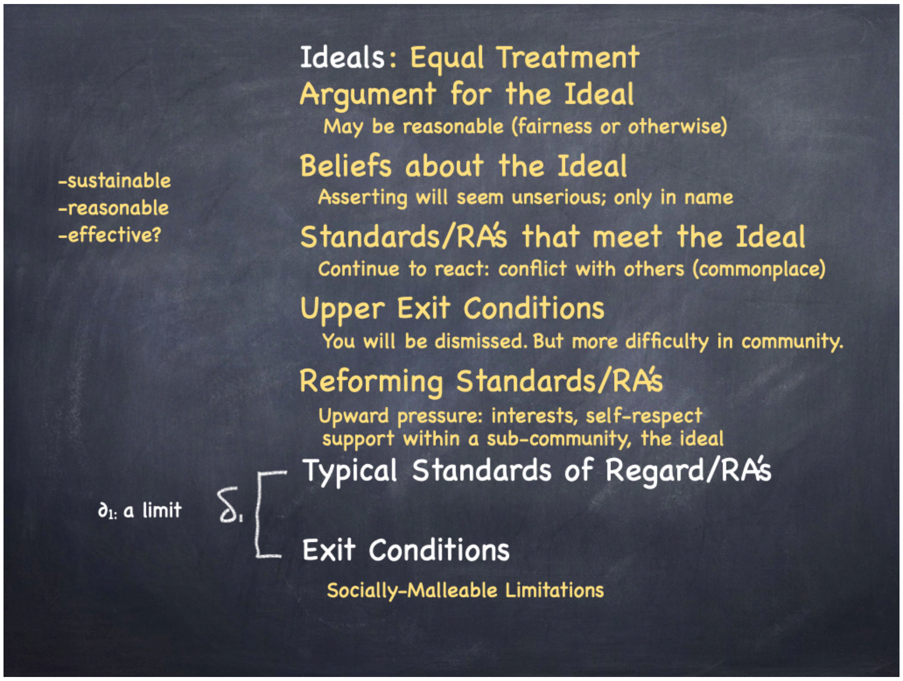
A different sort of counter-pressure will be available if you are not alone—if there is a sub-community of reformers. Such a sub-community will not only offer support to its members, but will also make it (somewhat) more costly and difficult for the larger society to "exit" (dismiss) the entire sub-community, given the pressures of the commonplace. Doing so will amount to, and so require, "dehumanizing" an entire class of people.

Meanwhile, within the sub-community, the standards will change. And there may appear, in the larger society, what we can call the **reforming standards**, now above the typical standards but *below* the upper exit conditions, as in **Figure 7**: these are standard of regard that challenge the status quo but can be tolerated, albeit with difficulty and conflict.

The reformers can, again, make arguments—and, in this case, the arguments might have *some* purchase with *some* people. First, because the limitations are socially malleable, it may be that the cost to those hurt by the status quo is, arguably, greater than the cost to others to make the change, which would render the reforming standard "reasonable," in the sense of "fair." If we presume a commitment to live together while treating each person and their interests symmetrically, the change would be required by that commitment to fairness. It will then be difficult to argue against the change without either again dehumanizing or else relying on metaphysics or ideology to generate (spurious) distinctions between, say, women and men, or races, or what have you.

[29] See, for example, (Hieronymi 2001; Smith 2013).

Figure 7



Ideals: Equal Treatment

Argument for the Ideal

May be reasonable (fairness or otherwise)

Beliefs about the Ideal

Asserting will seem unserious; only in name

-sustainable
-reasonable
-effective?

Standards/RA's that meet the Ideal

Continue to react: conflict with others (commonplace)

Upper Exit Conditions

You will be dismissed. But more difficulty in community.

Reforming Standards/RA's

Upward pressure: interests, self-respect
support within a sub-community, the ideal

Typical Standards of Regard/RA's

∂_1: a limit δ_1

Exit Conditions

Socially-Malleable Limitations

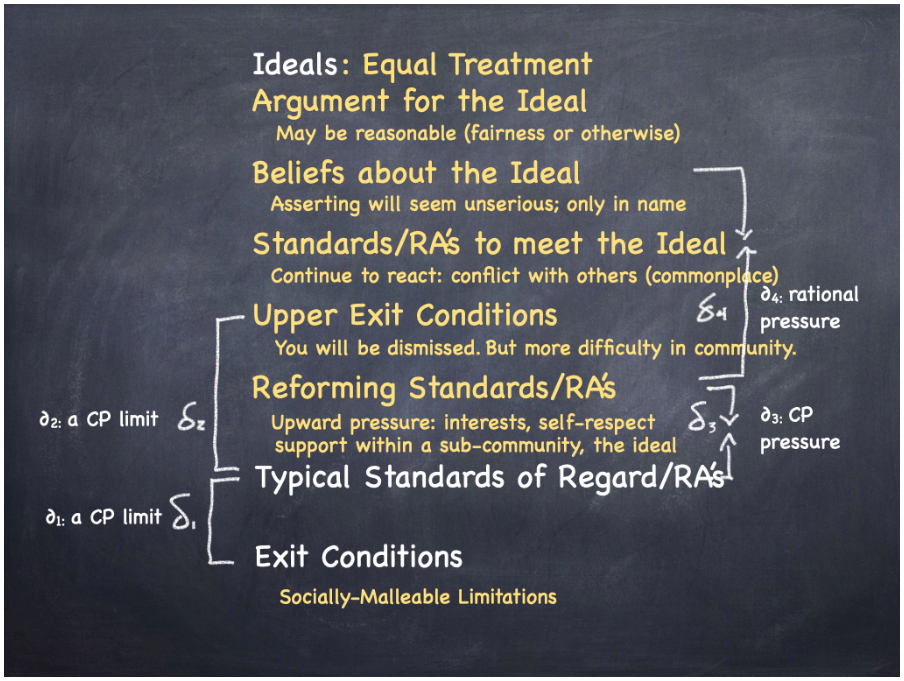
Alternatively, it may be that the ideal itself (e.g., beauty, ecological integrity, avoiding torture of sentient creatures) may be compelling enough reason to make the social change, rendering the expectation "reasonable" in a sense that appeals to the ideal, rather than to fairness.

In either case, it may be that these upwards pressures can render the reforming standards reasonable and sustainable. The reforming standards may, then, effectively become the typical standards—though whether they do will, of course, depend on how the non-reformers respond, whether they are moved by claims of consistency, empathy, or the ideal, or perhaps are simply, as individuals, replaced over time. This sometimes happens. The socially malleable limitations then shift upward, to (better) satisfy the standards.

Let us now zoom out and consider the pressures at work in more detail, in **Figure 8**. We have seen that there is another kind of limit, similar to $\partial 1$, between the typical standards and the *upper* exit condition. I will call this limit **$\partial 2$**. And, as we have seen, the commonplace generates pressure towards *uniformity* in the standards. I will call this pressure **$\partial 3$**. All three of these interpersonal pressures are generated by features noted by Strawson, in calling attention to his commonplace (**CP**)—including our limited capacity to tolerate what he called the strains of involvement.

We can add, to this, an additional kind of *intra*personal pressure: standards of rationality or what is sometimes called "fit" generate pressure to align one's own operative standards—one's own reactions—with one's beliefs about one's ideal. Distance here is the sort of thing experienced by those of us who find extreme poverty or factory farming unspeakable and yet fail to react, emotionally, personally, or interpersonally, in a way that would match our beliefs. I will call this **∂4**.

Figure 8



Ideals: Equal Treatment

Argument for the Ideal

May be reasonable (fairness or otherwise)

Beliefs about the Ideal

Asserting will seem unserious; only in name

Standards/RA's to meet the Ideal

Continue to react: conflict with others (commonplace)

Upper Exit Conditions

You will be dismissed. But more difficulty in community.

Reforming Standards/RA's

Upward pressure: interests, self-respect
support within a sub-community, the ideal

Typical Standards of Regard/RA's

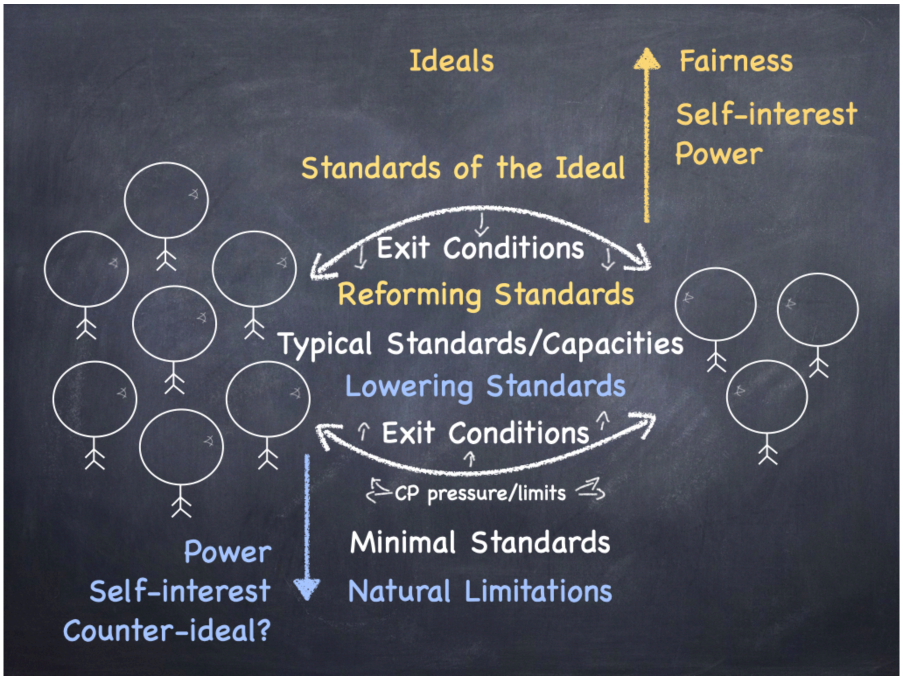
Exit Conditions

Socially-Malleable Limitations

δ_4
∂_4: rational pressure
δ_3
∂_3: CP pressure
∂_2: a CP limit δ_2
∂_1: a CP limit δ_1

Finally, we can reconsider the same dynamic in a slightly different format, in **Figure 9**. The items in the center, in white, are aspects of Strawson's picture. The facts Strawson noted in his commonplace, the fact that we care about our relations with others, generate pressure to keep the standards uniform and typically satisfied, and it eventually generates the limits we have called ∂1 and ∂2. Because these pressures and limits are generated by commonplace, i.e., by our caring about our relations with others, I will call them "horizontal"—though the lines are curved, here, because of the vertical pressures to keep this envelope narrow. It is the fact that we live in this envelope,

Figure 9



created by the commonplace, that Strawson takes to be a natural fact. But the envelope, itself, might move up and down vertically.

I have therefore added to Strawson's picture the possibility of "vertical" pressures, coming not only from natural limitations but also from ideals—and, we can now add, such pressure could also come from self-interest and from differences of power that will be exploited to advance self-interests or ideals. I suspect equality and fairness have a special place.[30]

Though this is all admittedly, embarrassingly armchair social science, it seems to me to fairly accurately capture certain dynamics I take to be actual. If something like it is possible, then it is possible for ideals to be incorporated into the standards of regard in a way that allows the critical purchase afforded by the ideal to be a part of the nature of morality. That is to say, what we can rightly call "morality" can have, within it, the resources to evaluate and criticize the social norms and

[30] I suspect it has special place because I suspect the reactive attitudes, which are attuned to matters of respect and disrespect, are especially attuned to matters of equality and fairness. This is what made for the doubleness in claims about equality that did not appear in claims about caring for the natural world. But this claim about equality and fairness is vulnerable to sociological and anthropological findings.

workings of an entire society, not just as an ideal pointed to only in name, but as part of morality itself. Morality thereby has the critical purchase it requires.

There will, of course, be objections to this dynamic picture. First, it might be objected that the dynamic of pressure-and-counter-pressure puts further, very significant, burdens on the already disadvantaged, and is therefore unjust. Alas, I have no answer to this. It seems to me this is so—a sorry consequence of our sorry condition.

Second, it might be objected that, on this picture, both morality and the reforming standards of regard are socially and historically contingent. That is true, and its truth might account for some of the urgency of the reformer's efforts—there is a real and present danger that the ideals might be lost to history, so to speak. That historical contingency is also, of course, one way in which the picture is naturalistic: it understands both the reactive attitudes and the social norms they constitute to be a product of our contingent, natural history. It could, I think, also understand the ideals in that way, while nonetheless insisting that the ideals remain both essentially contestable and possibly correct. Such a picture is not merely descriptive, insofar as it allows that the system described can be criticized by the ideals and can also incorporate them into itself. Strawson does think that his naturalism limits the possible ways of questioning our way of life: we could not have reason to exit our natural form of sociability.[31]

[31] I have not here considered a different challenge to Strawson's picture, namely, that the historically contingent development of our standards has left us in incoherence. I take up this objection in (Hieronymi 2020, 97–104).

Adopting Wiggins' phrase for what I suspect is a very similar position, I would call this a sensible naturalism.[32]

[32] (Wiggins 1987)

Thanks are due, first, to John Hyman, Yuuki Ohta, and Michael Thorne for their work in organizing the conference for which the paper was prepared, *Skepticism and Naturalism: Hume, Wittgenstein, and Strawson*, as part of their *Roots of Responsibility* project from the European Research Council, and for their further work in editing this volume. Michael provided an especially helpful set of editorial comments. Special thanks are also due to Jeremy Fix for his very helpful comments delivered at the conference, as well as to the other participants for their questions and conversation. This paper has further benefitted from conversation with the audiences at the University of Nevada, Reno; the Center for Human Values at Princeton University; the University of North Carolina, Chapel Hill; and the *Agency and Responsibility Research Group*.

BIBLIOGRAPHY

Goldman, David. 2014. "Modification of the Reactive Attitudes." *Pacific Philosophical Quarterly* 94 (4):1–22.

Hart, H. L. A. 1961. *The Concept of Law*. Third Edition ed. Oxford: Oxford University Press. Reprint, 2012.

Hieronymi, Pamela. 2001. "Articulating an Uncompromising Forgiveness." *Philosophy and Phenomenological Research* 62 (3):529–555.

Hieronymi, Pamela. 2011. "Of Metaethics and Motivation: The Appeal of Contractualism." In *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*, edited by Rahul Kumar, Samuel Scheffler and R. Jay Wallace, 101–128. Oxford: Oxford University Press.

Hieronymi, Pamela. 2020. *Freedom, Resentment, and the Metaphysics of Morals*. Princeton: Princeton University Press.

Plato. 1992. *Republic*. Translated by G. M. A. Grube. Edited by C. D. C. Reeve. Indianapolis: Hackett Publishing Company. Original edition, c. 380 B.C.

Scanlon, T. M. 1998. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.

Smith, Angela M. 2013. "Moral Blame and Moral Protest." In *Blame: Its Nature and Norms*, edited by Justin Coates and Neal Tognazzini, 27–48. New York: Oxford University Press.

Strawson, Peter F. 1961. "Social Morality and Individual Ideal." *Philosophy* 36 (136):1–17.

Strawson, Peter F. 1962. "Freedom and Resentment." *Proceedings of the British Academy* xlviii:1-25.

Strawson, Peter F. 1985. *Skepticism and Naturalism*: Columbia University Press.

Wiggins, David. 1987. "A Sensible Subjectivism?" In *Needs, Values, Truth*, 185–213. Oxford: Oxford University Press.