CRITICAL NOTICE

Excuses, Exemptions, and the Challenges to Social Naturalism

A Critical Notice by Sybren Heyndels of:

> *Freedom, Resentment, and the Metaphysics of Morals*, by Pamela Hieronymi, Princeton, NJ, Princeton University Press, 2020, $29.95/£25.00 (hardback), ISBN 9780691194035

Pamela Hieronymi has authored a very insightful book that focuses on one of the most influential articles in 20th century philosophy: P.F. Strawson's 'Freedom and Resentment' (1962). Hieronymi's principal objective in *Freedom, Resentment, and the Metaphysics of Morals* is to reconstruct and evaluate the central argumentative strategy in Strawson's essay. The author's aim is 'to show that it can withstand the objections that are both the most obvious and the most serious, leaving it a worthy contender' (3). In the present commentary, I summarize the main results of Hieronymi's analysis. I engage with the book's main themes, noting in due course certain unclarities and some shortcomings, while emphasizing the many valuable insights it offers.[1]

[1] References to Hieronymi's book, including its reprint of Strawson's 'Freedom and Resentment', will simply be by page number.

1. The Reactive Attitudes

Hieronymi offers a definition of Strawson's famous notion of the 'reactive attitudes':

> In general, then, a reactive attitude is *x*'s reaction to *x*'s perception of or beliefs about the quality of *y's* will toward *z*. (8)

Hieronymi's definition has the great merit that it captures succinctly the three different subtypes of reactive attitudes. These three subtypes are (a) the personal reactive attitudes, (b) the impersonal reactive attitudes, and (c) the self-directed reactive attitudes. When it concerns the *personal* reactive attitudes, the *x* and *z* (but not *y*) in Hieronymi's definition denote the same person. For example, gratitude and resentment are reactions to one's own perception or belief about someone else's quality of will toward oneself. Thus, gratitude is a reaction of *mine* towards my belief or perception of *your* goodwill towards *me*; resentment is a reaction of *mine* towards my belief or perception of *your* ill will towards *me*. When it concerns the *impersonal* reactive attitudes, *x*, *y* and *z* refer to three different persons. For example, indignation and moral admiration are a person's reactions to her perception or belief about someone else's quality of will towards a third person. When it concerns the self-directed reactive attitudes, *x* and *y* (but not *z*) are the same person. For example, guilt is a person's reaction to her perception or belief about her own quality of (ill) will toward another party.

I offer three remarks concerning Hieronymi's proposed definition. The first is that she does not adequately justify the definition of the reactive attitudes as reactions to someone's *perception* or *belief* about someone's quality of will, rather than as a reaction to someone's quality of will (tout court). One could object that the proposed characterization implicitly treats *all* reactive attitudes (not just the self-directed) as reactions to one's own states, given that these attitudes are defined as reactions towards one's own perceptual or cognitive states. While Hieronymi does not address this issue in any detail, I conjecture that her reason for invoking reactions to someone's 'perception or belief' (rather than to the quality of will without further ado) is motivated by the fact that we can be *mistaken* about someone's quality of will. I do think that the quality-of-will error possibility offers some justification for the incorporation of the qualification into her definition, but it has to be tacitly inferred from context, since it is not explicitly presented.

The second remark is that the selection of paradigmatic reactive attitude examples (such as gratitude, guilt and resentment) is problematic. Gratitude is regarded by both Strawson and Hieronymi as a paradigmatic example of a *personal* reactive attitude. But there are certain varieties of gratitude which are responses *not* to the way someone has treated *me* (as is the case of *personal* reactive attitudes) but rather responses to way a party has treated a *third person* I hold dear. For example, a mother can be grateful to someone who has gone to great lengths to help out her son. The third party directed gratitude amounts to an impersonal rather than personal attitude if one goes by Hieronymi's definition, given that in such a case, *x*, *y,* and *z* refer to distinct individuals. While Strawson and Hieronymi could be right that the standard cases of gratitude (or other examples of reactive attitudes) are to be captured by the proposed schematic analysis, it is possible for certain instances of a given subtype to exemplify another subtype as well (leading to an overlap in the categorization of the reactive attitudes). Such an observation reflects the actual complexity of the attitudes the definition aims to capture.

The third remark amounts to a sympathetic amendment to Hieronymi's definition. Hieronymi characterizes the personal reactive attitudes as attitudes in which 'the same person stands in for *x* and *z*'; and the self-directed reactive attitudes as attitudes in which 'the same person stands in for *x* and *y*' (8). It ought to be explicitly qualified that in the case of the *personal* reactive attitudes, it is not *only* that *x* and *z* refer to the same person, but *y* refers to *another* person distinct from *x* and *z*. Similarly, in the case of the *self-directed* reactive attitudes, *z* should be taken to refer to *another* person distinct from *x* and *y.* While I have rendered explicit the supplementary clause in my summary of Hieronymi's definition, it is not the author's own explicit refinement. Although it is clear from her examples that she indeed thinks that these further specifications have to be presupposed, explicitly supplementing the definition is necessary, in order to differentiate these types of reactive attitudes from *another* class of reactive attitudes, unconsidered by either Hieronymi or Strawson (or any other commentator, to my knowledge). The latter subclass is the class of reactive attitudes in which *x*, *y* and *z* all refer to the same person. This class of reactive attitudes are reactions of *mine* towards my perception or belief of the quality of will I express towards myself. Certain varieties of self-hatred, self-loathing, or self-pride constitute prime examples of such reactive attitudes. Similar to the class of self-directed reactive attitudes, these attitudes concern my *own* quality of will, but unlike the class of self-directed reactive attitudes, they do not concern how I treat *others,* but rather how I treat *myself*. In other words, it is logically possible that *x*, *y* and *z* might all refer to the same person, which ought to be accounted for in any adequate analysis. While it could be debated whether or not expressing ill will or goodwill towards oneself is conceptually possible or not, the possibility should not be discarded at the outset.

2. Excuses and Exemptions

Having proposed the definition of the reactive attitudes, Hieronymi offers an overview of the considerations under which it is appropriate to modify our reactive attitudes. More specifically, the overview concerns the ways in which we, according to Strawson, *excuse* and *exempt* people. Hieronymi argues that when we *excuse* people, ‘we were mistaken about the quality of the will in question, and therefore our reactive attitude – our reaction to our *perception of* or our *beliefs about* the quality of that will – must change.’ (9) And when we *exempt* people, these are cases in which we take it to be that ‘the ill will does not matter in the usual way’ (10). In cases of exemption, we shift to an *objective* attitude.

Hieronymi distinguishes between three subtypes of *exemption* as proposed by Strawson. The first subtype includes cases in which we discount someone’s quality of will temporarily, owing to extreme or unusual circumstances. We exempt an agent, for example, by pointing out that she ‘was not herself’, or ‘was under a lot of pressure’. The second subtype concerns cases in which we discount someone’s quality of will on the basis of a ‘more enduring condition’ (10), such as disease or immaturity. A third subtype is more peculiar, occurring when we use a ‘resource’ to adopt an objective attitude, in order to consider people who do not find themselves in extreme or unusual cases (as in the first subtype), or who are not incapacitated in any way (as in the second subtype). In these cases, Strawson writes, we use the given ‘resource’ ‘as a refuge, say, from the strains of involvement; or as an aid to policy; or simply out of intellectual curiosity’ (116). As Hieronymi is well aware, Strawson ‘seems reluctant’ to ‘class this third subvariety with the other two’ (11). Nevertheless, given that in these three cases, ‘ill will does not matter to us in the usual way’ (11), she takes them to be different species of the same genus.

There is a problem with this characterization of excuses and exemptions in relation to another point made by Hieronymi. The latter has to do with our expectations and demands for regard. In the case of *excuses*, Hieronymi argues that ‘[t]he demands stay in place’ (12). In the case of *exemptions*, on the other hand, we ‘cease to make the associated demands’ (12). If these features of excuses and exemptions are added to supplement the characterization in the previous paragraph, it emerges that Hieronymi is committed to the view that, when we *excuse*,

(i) We were mistaken about which quality of will was present.
(ii) However, the demands stay in place.

Whereas when we *exempt*,

(iii) The quality of will does not matter in the usual way.
(iv) The demands do not stay in place.

However, Hieronymi also argues that cases in which we *excuse*, and therefore we were mistaken about which quality of will was present, include cases where there isn't an operative will at all (9, fn5).[2] For example, we could be mistaken in thinking that there was a human agent (or, more generally, an agent with a will) involved in a particular interaction. Hieronymi emphasizes that her use of the term 'excuse' includes cases in which we would 'excuse' puppets or other things without a will (13 fn6). While she admits that this use of 'excuse' amounts to a technical term, diverging from Strawson's original discussion, there nevertheless remains an inconsistency between the use of the technical term 'excuse' in the case of puppets and her endorsement of the fact that in the case of such 'excuses', our moral demands stay in place. After all, in cases where we come to realize there is no operative will, our demands will *not* stay in place.

In order to preserve the idea that it is possible to 'excuse' things without a will, Hieronymi has several notable options. She could rescind her commitment to the claim that, in the case of 'excuses', our moral demands always stay in place. Or she could effect a distinction between two kinds of *excuses* in which we are either mistaken about which quality of will was present, while the demands stay in place; or mistaken about which quality of will was present, while the demands do not stay in place. My own inclination is to refrain from using the term 'excuses' when there is no will at all, for the following threefold reasons. In the first place, there seems to be no perspicuous need, nor clear advantage, in extending the use of the term to cover 'no operative will' cases. Moreover, attempting to extend the analysis leads to inconsistency in cases where the demands stay in place, even though there is no operative will at all; which entails that our moral demands stay in place when we discover the agent was in fact a puppet.

[2] Although Hieronymi makes this claim in a (lengthy) footnote, I think it is important to discuss it given that it has consequences (see section 3) for her overall evaluation of Strawson's argumentative strategy. In general, Hieronymi too often makes important philosophical claims in footnotes. They should have been discussed more elaborately in the main body of the text.

And finally, Strawson himself does not make this claim – a fact amply known to Hieronymi (13 fn6).[3]

3. Exemptions and Statistics

Strawson's account of responsibility, and Hieronymi's reconstruction of it, allow for a reformulation of the putative incompatibility of determinism and moral responsibility: is there any sense of 'determinism', or 'being determined', that would make it the case that either (a) we are always *excused*, or (b) we are always *exempted*? Strawson argues that the answer in each case is 'no'.

While Hieronymi's main focus throughout the book is on (a) (concerning *exemption*), several remarks arise about her abbreviated discussion of (b) (concerning *excuses*). Hieronymi correctly observes that according to Strawson, the truth of determinism cannot have the consequence that we are always *excused*, given that this would imply 'the reign of universal goodwill' (117). Indeed, if considerations amounting to an excuse occur, we learn thereby that we were mistaken in thinking an agent acted on the basis of ill will. Hieronymi does not enter into a full explication of these matters, simply concluding that 'the first sort of revision (in which we come to see that "the will was not ill") is not fit for general application' (16). Note that given Hieronymi's idiosyncratic understanding of the class of *excusing* considerations, she is in fact committed not merely to Strawson's proposition that the universal application of excusing considerations implies a reign of 'universal goodwill'; but universal application could also imply that no one has any will at all, or that people either have goodwill or no will at all. This follows from her characterization of *excuses* as appropriate, including in circumstances in which there is no will at all. Plausibly, Hieronymi may well think that there isn't any sense in which 'being determined' entails that no will is ever present, but an explicit elucidation would not have gone amiss.

Hieronymi focuses on Strawson's claim that the truth of determinism cannot make it the case that we are always and everywhere *exempted*. As we have seen, Hieronymi distinguishes

[3] Note as well that there is an ordinary sense of talking about 'excuse' which does not presuppose that one is mistaken about an agent's quality of will at all in excusing their behavior. One might excuse an agent without ever having made a mistake concerning that agent's will. Thanks to Jim O'Shea for suggesting this further point (O'Shea questions whether Strawson's text either states or implies that excusing agents for their actions normally follows upon having made any such initial 'mistake').

between three subtypes of exemption: (a) cases in which we discount someone's will, owing to extreme or unusual circumstances, (b) cases in which we discount someone's will, for reasons of disease, immaturity or incapacity, (c) cases in which we resort to a 'resource' in adopting an objective attitude towards a person, as a palliative against 'the strains of involvement'. Hieronymi discusses (a)-(b) separately from (c). Accordingly, I will first focus on (a)-(b).

(a)-(b) are similar because they both concern *outlier* cases. Strawson's argument is that since exemption is reserved for outlier cases, the truth of determinism cannot entail that we are always exempted, because outlier cases cannot possess universality, on pain of self-contradiction (clearly, they would not be *outliers* if this were the case). As Strawson writes, 'it cannot be a consequence of any thesis which is not itself self-contradictory that abnormality is the universal condition' (118). The third subtype of exemptions differs from outlier cases, because it concerns cases in which we elect to take up an objective attitude towards a person, without its necessarily being the case that she finds herself in extreme or unusual circumstances, or is otherwise incapacitated.

The analysis of Strawson's argument concerning *outlier* cases lies at the heart of Hieronymi's overarching thesis. Drawing not only from 'Freedom and Resentment', but also from Strawson's important, yet often neglected, 'Social Morality and Individual Ideal' (published a year earlier in 1961), she constructs a nuanced interpretation of Strawson's position. Strawson argued that the existence of human society presupposes a minimal set of rules concerning our demands and expectations towards one another; moreover, the minimal set of rules requires that the demands of such a system are 'pretty regularly fulfilled' (Strawson 1961: 5). Hieronymi emphasizes correctly that Strawson's characterization of rule minimality allows for a comprehensive variety in the *kinds* of demands and expectations on which society is founded. Despite the existence of variation, the point adduced by Strawson is that any recognizable instance of human society must be governed on the basis of some manifestation of a cohesive set of moral demands and expectations.

Hieronymi (87) expands Strawson's account, arguing that a system of expectations and demands is partly determined by what is usual or ordinary, relative to our *purely natural* capacities and *socially developed* capacities. In effect, she argues that the nature of the mutual expectations and demands in any given society depends on (potentially culturally diverse) social practices of holding responsible, and on human emotional constitution. To clarify how a given society's system of expectations and demands depends on natural constitution, Hieronymi offers the illuminating example of a society where:

> [W]e all naturally possessed only the degree of inhibitory control, attention, and memory that we now possess when fairly intoxicated. The system of demands and expectations that would form, in our society, would be sensitive to those limitations. Certain expectations and demands would be unreasonable and unsustainable (31-32).

In the permanently intoxicated society sketched by Hieronymi, reasonable expectations concerning memory, recall of events, emotional outbursts, intimate interactions, etc., would be quite different, given that an altered mental state would be the norm. Our standards of regard would adjust downwards. Similarly, if memory and attention were improved (through "human enhancement"), the resulting expectations and demands would have to be set at a much higher standard, compared with our actual society (81). Hieronymi's generalization is that '[i]f we had different capacities, we would live under a different system of demands', producing a difference 'in what *counts* as showing ill-will or disregard' (32). To reiterate, our system of demands and expectations would 'adjust to what is typical or tolerably ordinary' (33).

Hieronymi's analysis of Strawson's argument to the effect that determinism cannot have the consequence that 'abnormality is the universal condition' (118) is illuminating in itself. Overall, Hieronymi offers a plausible interpretation of Strawson's view, given the wealth of insight it derives from Strawson's 'Social Morality and Individual Ideal' paper. Commentators on 'Freedom and Resentment' have unjustly neglected this valuable resource; Hieronymi's discussion has the additional merit of showing how our understanding Strawson's 'Freedom and Resentment' can be deepened in juxtaposition with this essay.

I return to Hieronymi's account of Strawson's 'metaphysics of morals' in section five. To anticipate that discussion, I make two observations here. First, Hieronymi argues that Strawson offers 'the ingredients for a transcendental argument moving from the existence of society to the satisfaction of the conditions required for it – the typical observance of a minimal set of rules.' (28) In a recent article, Coates (2017) argues that Strawson develops a 'modest transcendental argument' in 'Freedom and Resentment', which moves from the fact that we are involved in interpersonal relationships to the satisfaction of the conditions required for such relationships, i.e., that we are sometimes responsible for our actions. Coates refers to Strawson's work on Kant to render plausible his transcendental reading. While Hieronymi notes in passing that the ingredients for a 'transcendental argument' are present, she does not return to elucidate her envisaged form of transcendental argument, or explain how her reconstruction differs from Coates' interpretation.

The above observation relates to a more general second point. At the beginning of the Introduction, Hieronymi claims that ‘the central argument of’ Strawson’s article ‘has received relatively little attention’ (1). But this is simply incorrect. The historical and recent literature on Strawson’s main argumentative strategy is voluminous. While Hieronymi admits in a footnote that other philosophers have discussed these issues prior to the publication of her book (she refers to Shoemaker & Tognazzini 2015), the book contains no discussions of how her interpretation relates to many of the most prominent discussions of Strawson’s main arguments (and only a few references to these important critical antecedents). Importantly, works published in recent years *are* quite relevant to Hieronymi’s own analysis. In particular, in recent years much has been published on the so-called ‘reversal move’ in Strawson, which concerns the relation between ‘being responsible’ and ‘holding responsible’. The ‘reversal move’ denotes the idea, often attributed to Strawson, that ‘holding responsible’ is, in some sense to be specified, prior to ‘being responsible’; or that moral responsibility is a kind of ‘response-dependent’ notion. Todd (2016) convincingly argued that many commentators failed to clarify adequately the ‘reversal move’. Recently published works strive to rectify this omission (Shoemaker 2017, Beglin 2018, McGeer 2019, De Mesel & Heyndels 2019, De Mesel 2021).

According to a common construal of the ‘reversal move’, the act of holding responsible makes it true that someone is responsible for a particular action. The shortcomings of this idea are evident. The proposed view does not allow for the possibility of *mistakes* when attributing moral responsibility for a particular action. There are cases in which holding responsible and actually being responsible come apart. While many authors have criticized the ‘reversal move’ for this very reason, they have often taken this objection to amount to a criticism of Strawson’s actual views. However, in our co-authored article, ‘The Facts and Practices of Moral Responsibility’ (2019), Benjamin De Mesel and I argued that Strawson’s reversal move should not be understood along those lines in the first place. Rather than fixing whether someone *is* responsible for a particular action, we argue that our practices of holding responsible fix the *criteria* for someone to *count* as responsible for a certain action (or the criteria for *counting* an action as expressing disregard or ill will in the first place). Furthermore, we add that these criteria for *being* responsible are fixed not just by our *social* practices but also by certain *natural* facts about us. If certain natural facts about us were to change, it would equally affect the *criteria* for what counts as being responsible for a particular action, or the criteria for what *counts* as disregard.

While Hieronymi does not make this connection, her account fits well with our and other accounts of the reversal move. What is novel about Hieronymi's proposal is that it shows how an understanding of the 'reversal move' as fixing the *criteria* for moral responsibility illuminates Strawson's argument that the truth of determinism could not lead to the fact that we are universally exempted. Given that we exempt in *outlier* cases only, it cannot be the consequence of any claim that outlier cases apply all the time. Statistics matter, in the sense that what *counts* as being morally responsible depends on our actual practices of demanding and expecting things from each other, which is subject to facts about *social* practices, as well as facts about our *natural* capacities (87). If these social practices or capacities change, then the criteria for what counts as being morally responsible for a particular action would likewise undergo a shift. As Hieronymi writes: 'On Strawson's socially naturalistic picture, moral standards […] are constituted, at least in part, by actual moral practice' (75). This is compatible with the idea that *any* kind of recognizable human society must have *some* set of rules, underlying our practices of moral responsibility, which govern our system of demands and expectations.

4. The 'Resource' and the Justificatory Demand

So far, I have commented on Hieronymi's discussion of the first two subtypes of *exemptions*: exemptions in outlier cases, in which we disregard someone's will on the basis of extreme or unusual circumstances, and cases in which we discount someone's will because of disease, immaturity or incapacity. Strawson thinks that the truth of determinism cannot universally exempt us, because it cannot be true that outlier cases apply universally. But Hieronymi finds a third subtype of exemption in 'Freedom and Resentment': exemptions in normal (non-outlier) cases, in which we nevertheless choose to adopt an 'objective attitude', because we seek to avoid the strains of involvement. Correspondingly, Strawson considers whether the truth of determinism could *exempt* us in this third way. Given that Strawson admits that we might take up an 'objective attitude' in *normal* cases, an incompatibilist might ask whether we should *always* adopt such an attitude (e.g., if determinism is true)? In the present context, it is no use to point out that the truth of determinism cannot have as a consequence that *outlier* cases apply universally, since the situations under discussion are those in which we adopt an objective attitude in cases that are not outliers.

Hieronymi thus further analyses Strawson's discussion of the possibility that the truth of determinism implies that we should resort to this resource at all times. Hieronymi points out that initially, Strawson seems to answer the variant question as to whether we *could* use this resource all the time. Strawson writes that a 'sustained objectivity of the interpersonal attitude … [is not] something of which human beings would be capable' (118) The 'crucial objection' against Strawson's response is that stressing the strength of our commitment to interpersonal relationships (as a matter of empirical fact) does not show that we are *justified* in being engaged in interpersonal relationships. As Hieronymi writes, Strawson 'seems to leave untouched whether we *should* use our resource at all times. I call this the "crucial objection"' (51). An incompatibilist might agree that we *could* not give up such relationships, but that we nevertheless *should* give up such relationships, if determinism is true. The possibility arises that we are naturally committed to a practice that is in itself unjustified.

The argument instantiates an often reiterated point in the literature on Strawson's 'Freedom and Resentment'. For example, Gary Watson refers to Strawson's response (which he regards as inadequate) as the 'psychological inescapability argument':

> What puzzles me, however, is the prior question of how this claim is dialectically relevant. So what if it's true? What prompts "pessimism" or scepticism is the worry or conviction that there is or might well be a general "theoretical ground" for abandoning our sense of responsibility […] To add that even if that conviction is correct, we couldn't adjust our lives accordingly (because we could not accept or "absorb" its truth) leaves these interlocutors' basic position completely intact. What work, then, is the psychological inescapability argument supposed to be doing? (Watson 2014, 25-26)

Similarly, the 'psychological impossibility' claim is what Paul Russell (1992) takes to lie at the heart of Strawson's naturalistic (as opposed to his rationalistic) strategy. And it has been characterized by András Szigeti as Strawson's 'inescapability arguments' because 'they move *from* the diagnosis of inescapability of the practice *to* the conclusion that the practice is justified' (Szigeti 2012, 92).

As Strawson's opponent emphasizes, the claim that it is 'practically inconceivable' to abandon the reactive attitudes at the heart of our responsibility practices (and the system of demands and expectations that go along with these) does not entail that these practices are *justified*. Hieronymi does well to treat by way of a rejoinder Strawson's *own* actual response to this line of thought. Strawson did not ignore the possibility of just such an objection, as he writes that '[i]t might be said that all this leaves the real question unanswered [...] For the real question is not a question about what we actually do, or why we do it. It is not even a question about what we would *in fact* do if a certain theoretical conviction gained general acceptance. It is a question about what it would be *rational* to do if determinism were true, a question about the rational justification of ordinary interpersonal attitudes in general' (120). Strawson then goes on to argue that such a line of thought can only occur to someone 'who had utterly failed to grasp the purport of' his answer because '[t]his commitment [to ordinary interpersonal attitudes] is part of the general framework of human life, not something that can come up for review as particular cases can come up for review within this general framework' (120). Hieronymi justly connects these ideas to Strawson's claim that 'the general framework of attitudes [...] neither calls for nor permits an external "rational" justification' but only allows for 'questions of justification [that] are internal to the structure' (131). Therefore, in order to respond to Strawson's opponent, Hieronymi must make sense of Strawson's *rejection* of the *justificatory demand*. This is an important point, which has not been sufficiently appreciated in the literature on 'Freedom and Resentment' (but see De Mesel 2018).

The internal/external distinction, Hieronymi argues, lies at the heart of Strawson's naturalism. While some authors have connected Strawson's naturalistic strategy in 'Freedom and Resentment' with his later work, *Scepticism and Naturalism* (1985), many have overlooked the ninth chapter of Strawson's first book, *Introduction to Logical Theory* (1952). This chapter, praised by Quine (1953, 433) in his review as 'an excellent little philosophical essay', concerns the topic of justifying our basic canons of induction. Hieronymi, however, notices that Strawson's rejection of the justificatory demand in 'Freedom and Resentment' resembles his rejection of the justificatory demand for induction in *Introduction to Logical Theory* (a relatively less familiar source to many). In the latter, Strawson argues that it makes sense to *justify* particular pieces of inductive reasoning to support *particular* beliefs, but not to seek to *justify* induction itself. As Hieronymi argues, Strawson's rejection of the justificatory demand resembles Wittgenstein's anti-skeptical strategy in *On Certainty*, and she gives clear examples to substantiate the observation that Strawson explicitly refers to *On Certainty* as an influence on his social naturalist account in 'Freedom and Resentment'.

While I endorse it as the correct exegesis of Strawson, I don't think that Hieronymi quite succeeds in convincing incompatibilists that their question concerning the 'external' justification of our practices as a whole is meaningless. What is necessary yet absent in

Hieronymi's discussion is a far more detailed account of the Wittgensteinian foundation of Strawson's views, rendered plausible by a far richer arsenal of examples of similar cases, in which justificatory demands are shown to be meaningless. A promising strategy, in my view, is to connect the proposed Wittgensteinian line of thought with Carnap's distinction between internal and external questions, which Strawson himself may well have envisaged when writing 'Freedom and Resentment'. A good case can be made that Strawson's general method in 'Freedom and Resentment' is to *describe* our actually adopted criteria for holding someone responsible (*internal* to our practices), in order to *explain* how these criteria are rooted in our natural capacities; and to argue that any notion of moral responsibility that allows for the possibility that our moral responsibility attributions are *always and universally mistaken* would be radically different from our actual notion of moral responsibility (a notion *external* to our practices) (see Heyndels 2019).

I think such an approach has the benefit of relating to contemporary discussions in neighboring fields. A relevant case in point is the revival of neo-Carnapian approaches in (meta)ontology, where the 'internal/external' distinction is invoked to show that certain eliminativist proposals, according to which ordinary objects such as tables and chairs don't exist, should actually be interpreted as (tacit) novel uses of words like 'table' and 'chair' that substantially depart from our actual usage. Given the divergence from actual usage, the eliminativist's answer is without impact on any actual ontological questions that might arise in our practices. Another case in point concerns approaches which aim to show that the incompatibilist minimally succeeds in arguing that only an excessively demanding notion of the ability to do otherwise is trivially not fulfilled if determinism is true (a notion *external* to our practices), but that there is a robust and ordinary sense of the ability to do otherwise which is not threatened by determinism (a notion *internal* to our practices). I take certain *contextualist* accounts (which are influenced by similar contextualist strategies against skepticism about knowledge) to be promising in this relation. I also highlight Christian List's (2019) recent argument that determinism is incompatible with the *physical* possibility of doing otherwise, yet compatible with the *psychological* possibility of doing otherwise. (The present remarks are merely suggestive; a full discussion of the viability of the account lies beyond the scope of this critical notice.)

5. Social Naturalism

The most original contribution of the book by far is Hieronymi's defense, and extension of, Strawson's social naturalism. 'Social naturalism' was a term coined by Strawson in his *Skepticism and Naturalism,* and thus does not appear in his earlier work, 'Freedom and Resentment'. Nevertheless, Hieronymi uses the term to refer to Strawson's 'metaphysics of morals' (as she calls it) in 'Freedom and Resentment'. Hieronymi's position is outlined in response to a possible objection against the claim that our system of demands and expectations would 'adjust to what is typical or tolerably ordinary' (33). This view seems to imply that what is *moral* is to be equated with what is *ordinary*. But this seems profoundly wrong. There are numerous examples in the history of our species where morally questionable or wrong behavior is normalized. In order to make Strawson's position credible, Hieronymi needs to explain how Strawson's account can incorporate the idea that *criticism* of our ordinary standards is possible. While Strawson hints at the idea that there is room within our practices for 'endless modification, redirection, criticism' (131), he does not elaborate his thought. Hieronymi's objective is to make good the omission.

Hieronymi develops the outlines of (what I would call) a *dynamic* social naturalist account of how our system of demands and expectations is constituted. There are two central ideas. The first is that there is *pressure* within our practices to adjust our system of demands and expectations to the limitations of our actual natural constitution and the social habits imposed by traditions. This, as Hieronymi emphasizes, reflects an actual (and sometimes tragic) fact about the human condition. But while there 'will be emotional and interpersonal pressure to conform to the majority or the dominant' (90), there is also *counter-pressure* which allows for the criticism of these limitations. Criticism may take the form of defending standards of regard that constitute certain *ideals* to which one is committed and can be upheld, even if the ideal standards of regard are most often violated. Defending such ideals in a world where the standards of regard are almost universally violated is far from easy. While criticism of actual practices can lead to actual change, it can also lead to one's becoming *exempted* from ordinary interpersonal relationships. As Hieronymi writes, '[a]s time goes on, it is likely they will either turn against you (the list of martyred moral reformers is long) or else begin to use their resource to respond to you more objectively: you will become a problem, an issue, or perhaps a kind of curiosity or museum piece. You will then be left outside the scope of ordinary interpersonal relationships' (86).

Hieronymi's account sketches a kind of 'metanormative' theory of how a system of moral norms emerge, evolve and change. Hieronymi's characterization of the 'pressure and counter-pressure' of systems seems to capture a fundamental tension at the heart of our moral practices, in which there are both traditional forces that aim at keeping the status quo and more revolutionary forces that aim at transcending it. Rather than seeing this duality as a contradiction, in a proto-Hegelian manner Hieronymi argues that the 'contradiction' is in fact an essential aspect of our community as moral agents. Her views undoubtedly amount to an exciting proposal, deserving a far more detailed elaboration than has been given here.

The best insights in the book emerge when Hieronymi sets aside the question about the (in)compatibility of moral responsibility and determinism in order to characterize and expound Strawson's social naturalist construal of systems of moral demands and expectations. This positive account is interesting in its own right, independently of the incompatibilist/compatibilist debate. Personally, I think that the value of Strawson's account lies not in its potential to 'defeat' the incompatibilist, but rather in its ability to *explain* the nature of the debate between incompatibilism and compatibilism. In fact, I think that Hieronymi's *dynamic* picture of how our moral demands and expectations emerge and evolve contains the basic ingredients for characterizing the substance of the debate between compatibilism and incompatibilism as the dynamic tension at the heart of our moral practices itself. Our moral practices themselves contain both the drive towards the status quo of our *actual* standards as well as a longing to *transcend* and *change* these standards. On such an account, the incompatibilist's position is rooted in our actual moral practices, in the sense that the demand for transcending our actual criteria for moral responsibility is part of this very practice, just as the actual criteria themselves are part of this very practice (which the compatibilist will emphasize time and again). In light of the perennially ongoing debate between compatibilism and incompatibilism, the free will problem does not seem to require solving, but rather *explaining* how the tension at the heart of the problem itself is rooted in our practices.

6. Conclusion

On the whole, Hieronymi's *Freedom, Resentment, and the Metaphysics of Morals* is a very insightful book. Not only does it yield a plausible interpretation of Strawson's influential 'Freedom and Resentment', but it succeeds in developing Strawson's account at great length in original and exciting ways. It is one of the best and most detailed discussions of Strawson's argument ever to have appeared. For these reasons, it constitutes necessary reading for anyone interested in Strawson's philosophy, or in the broader literature on moral responsibility to which it equally contributes. Apart from a number of minor criticisms, I have noted two general objections. The first is that Hieronymi does not sufficiently address the plethora of similar and dissimilar interpretations of Strawson in the extensive literature to date. While Hieronymi's account is sufficiently lucid and original to avoid becoming enmeshed in an excessively intricate analysis, a little more attention might have been welcome in these areas. The second remark is that the social naturalist picture developed in the course of the book deserves ampler treatment. For all that Hieronymi's account is fascinating, it nevertheless cries out for refinements and nuances. If the general objections should come to no more than the desire to hear more from the author on these topics, then it is safe to say that Hieronymi has succeeded in producing an excellent work of philosophy.[4]

References

Beglin, David. 2018. Responsibility, Libertarians, and the "Fact as We Know Them": A Concern-Based Construal of Strawson's Reversal. *Ethics* 128 (3): 612-625.

Coates, D. Justin. 2017. Strawson's modest transcendental argument. *British Journal for the History of Philosophy* 25 (4): 799-822.

De Mesel, Benjamin. 2018. Are our moral responsibility practices justified? Wittgenstein, Strawson and justification in 'Freedom and Resentment'. *British Journal for the History of Philosophy* 26 (3): 603-614.

De Mesel, Benjamin. 2021. Early View. Being and Holding responsible. Reconciling the Disputants Through a Meaning-Based Strawsonian Account. *Philosophical Studies*.

De Mesel, B. & Heyndels, S. 2019. The Facts and Practices of Moral Responsibility. *Pacific Philosophical Quarterly* 100 (3): 790-811.

[4] This research was funded by the Irish Research Council. I would like to thank Benjamin De Mesel and Jim O'Shea for helpful feedback.

Heyndels, S. 2019. Strawson's Method in 'Freedom and Resentment'. *Journal of Ethics* 23 (4): 407-423.

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

List, Christian. 2019. *Why Free Will is Real.* Cambridge: Harvard University Press.

McGeer, Victoria. 2019. Scaffolding agency: A proleptic account of the reactive attitudes. *European Journal of Philosophy* 27 (2): 301-323.

Quine, Willard van Orman, 1953. Mr. Strawson on Logical Theory. *Mind* 62: 433-451.

Russell, Paul. 1992. Strawson's Way of Naturalizing Responsibility. *Ethics* 102 (2): 287-302.

Shoemaker, David. 2017. Response-Dependent Responsibility: or, A Funny Thing Happened on the Way to Blame. *The Philosophical Review*, 126 (4): 481-527.

Shoemaker, D. & Tognazzini, N. 2014. *Oxford Studies in Agency and Responsibility, Volume 2: 'Freedom and Resentment'.* Oxford: Oxford University Press.

Strawson, P.F. 1952. *Introduction to Logical Theory*. London: Methuen.

Strawson, P.F. 1961. Social Morality and Individual Ideal. *Philosophy* 36 (136): 1-17.

Strawson, P.F. 1985. *Skepticism and Naturalism*. New York: Columbia University Press.

Strawson, P.F. (1962) 2020. Freedom and Resentment. In *Freedom, Resentment, and the Metaphysics of Morals* by P. Hieronymi*,* 107-133. Princeton: Princeton University Press.

Szigeti, Andras. 2012. Revisiting Strawsonian Arguments from Inescapability. *Philosophica* 85 (2): 91-121.

Todd, Patrick. 2016. Strawson, moral responsibility, and the "Order of Explanation": An intervention. *Ethics* 127 (1): 208-240.

Watson, Gary. 2014. Peter Strawson on Responsibility and Sociality. In *Oxford Studies in Agency and Responsibility, Volume 2*, 15-32. Oxford: Oxford University Press.

Sybren Heyndels
University College Dublin
sybren.heyndels@ucd.ie