This is the preprint version of an article published in *Pacific Philosophical Quarterly* 100 (3), 790-811. The final publication is available at Wiley Online: https://onlinelibrary.wiley.com/doi/10.1111/papq.12276. Please cite the published version only.

The Facts and Practices of Moral Responsibility

Benjamin De Mesel, Sybren Heyndels

Strawsonians about moral responsibility often claim that our practices of holding morally responsible fix the facts of moral responsibility, rather than the other way round. Todd (2016) refers to such 'reversal' claims in Watson (2004 and 2014), Kane (2005), Brink and Nelkin (2013), Tognazzini (2013), and Coates and Tognazzini (2013). Many have argued that reversal claims have an unwelcome consequence: if our practices of holding morally responsible fix the facts of moral responsibility, does this not imply, absurdly, that if we *held* the severely mentally ill responsible, they would *be* responsible? (Todd 2016: 211-212; Fischer and Ravizza 1993: 18-19; Fischer 1994: 212-213; Ekstrom 2000: 148; Nelkin 2011: 28) According to Todd (2016: 238), Strawsonians have not been clear about their answer to this question, to which we will henceforth refer as the absurdity challenge. We will provide an answer to the absurdity challenge, showing how Strawsonians could maintain that our practices fix the facts of moral responsibility without implying that if we were to blame the mentally ill, they would be blameworthy. We will proceed as follows. First, we will distinguish between two ways of reading the Strawsonian claim that our practices of holding morally responsible fix the facts of moral responsibility. Second, we will argue that the most plausible reading of the Strawsonian claim does not lead to an unwelcome conclusion about the mentally ill. Third, we will clarify our answer to the absurdity challenge in relation to Beglin's (2018) discussion of it. Fourth, we will reply to possible objections. Fifth, we will explore the relation between the reversal claim and (in)compatibilism.

1. A Criterion Thesis or an Obtainment Thesis?

This is the preprint version of an article published in *Pacific Philosophical Quarterly* 100 (3), 790-811. The final publication is available at Wiley Online: https://onlinelibrary.wiley.com/doi/10.1111/papq.12276. Please cite the published version only.

How is it possible that our practices fix the facts of moral responsibility? Does this not imply, absurdly, that if we *held* the mentally ill morally responsible, they would *be* morally responsible? Todd compares this problem to a standard objection to divine command theory: if something is right because God commands it, does this not imply, absurdly, that if God commanded murder, then murder would be right? According to Todd, what is needed are some examples with which the Strawsonian proposal might profitably be compared, examples in which something fixes something else in the way in which our practices of holding responsible fix the facts about moral responsibility. Todd suggests the following example. It seems as if it could not turn out that nothing we have ever laughed at was really funny. This impossibility indicates that our practices somehow fix the facts about funniness (Todd 2016: 236). But what if we were to laugh at genocide? Would genocide then be funny? The responsibility example, the divine command example and the funniness example seem to run parallel.

Consider, by way of a first attempt to clarify some of the issues involved, the following question: Could it turn out that nothing we have ever treated as a table really was a table? If not (and this answer seems plausible), Todd would agree that our practices somehow fix the facts about tables. But what if our practices were different? Suppose, for example, that we were to treat chairs as tables. Would chairs then be tables? In a sense, they clearly would not: treating something as a table does not make it a table. If we were to call chairs 'tables', for example, we would be mistaken. Our treating chairs as tables (including, among other things, our calling chairs 'tables'), would not change the fact that *x* is a table or that *y* is a chair. Similarly, an agent does not become responsible because we hold her responsible, and laughing at something does not make it funny. But that is only part of the story. If all of us (or a significant majority) were to start treating chairs as tables and calling chairs 'tables', and if we were to do this during a significant period of time, the *meaning* of 'table', our idea of what a table *is*, would gradually

This is the preprint version of an article published in *Pacific Philosophical Quarterly* 100 (3), 790-811. The final publication is available at Wiley Online: https://onlinelibrary.wiley.com/doi/10.1111/papq.12276. Please cite the published version only.

change. 'Table' would no longer mean *table* (it would no longer mean what it means now for us), but *table**, and the class of *tables** includes tables and chairs. Thus, the question, 'If we were to call chairs 'tables', would chairs then be tables?' can be answered in two ways: chairs would not be tables, but they would be tables*. What would change is not the fact that *x* is a table or that *y* is not a table, but the fact that 'table' means *table,* our idea of what a table is. A distinction could be made, in this respect, between 'facts of the world' (*x* is a table) and 'facts of meaning' ('table' means *table,* a 'table' is a *table*). In this case, only the latter are fixed by our practices.

The table example points at a distinction between two types of theses that is crucial for our response to the absurdity challenge, at two ways of reading the Strawsonian claim that our practices of holding responsible fix the facts of moral responsibility. The first type of thesis is the following: our practices of holding responsible fix the facts of meaning, they fix what it means to be morally responsible. They determine what the *criteria* for moral responsibility, the *responsibility-making and responsibility-defeating facts*, are.[1] Our practices fix that an agent is

[1] It may be held that there is a difference between criteria for moral responsibility on the one hand and responsibility-making and -defeating facts on the other. We do not think that the difference, if there is one, is important for our argument. As long as our response works with at least one of the two, there is no problem. Our use of the term 'criterion' has been influenced by Hacker (1993: 243-266) and Glock (1996: 93-97). They explain the term in the context of an account of meaning inspired by the later Wittgenstein. The point that our account of meaning and criteria is largely Wittgensteinian strengthens rather than weakens our claim to explain the Strawsonian proposal: despite certain disagreements that are of no consequence for our argument, Strawson seems to have supported a broadly Wittgensteinian account of meaning

morally responsible if and only if such and such (responsibility-making) facts obtain and such and such (responsibility-defeating) facts do not obtain. The thesis does not say *that* these facts obtain in a specific case, or that our practices make them obtain, it does not even say whether the responsibility-making or responsibility–defeating facts obtain at all. Let us call this type of thesis a criterion thesis.

Being mentally ill is a responsibility-defeating fact.[2] If our practices were such that we were to hold the mentally ill morally responsible, then the criteria (responsibility-making and responsibility-defeating facts) fixed by these practices would have to be different from the criteria fixed by our current practices: because the practices fix the criteria, being mentally ill could no longer be a responsibility-defeating fact. This is not to say that every change in our practices entails a change in criteria. Suppose, for example, that we hold an agent responsible because we believe that she has the mental capacities required for moral responsibility. We become convinced that she is mentally ill and, as a result, we no longer hold her responsible. Such a change in our practices keeps the criteria for moral responsibility constant: being mentally ill remains a responsibility-defeating fact. What changes is our belief about whether these facts obtain (about whether the agent is in fact mentally ill). The change in our practices described by those who confront Strawsonians with the absurdity challenge is of a different kind: if we were to hold the mentally ill (all of them, including those we *know* to be mentally ill) morally responsible in a systematic way, then being mentally ill could no longer be a

(see, for example, Strawson 2008b: chapter 4; Strawson 1992: chapter 8; Strawson 2008a: chapter 7).

[2] Note that, in the absurdity challenge, 'mental illness' refers to 'whatever sort of mental illness undermines moral responsibility' (Todd 2016: 212; footnote 7).

responsibility-defeating fact (because what these facts are is fixed by our practices). In that case, the change in our practices entails a change in criteria. To someone who knowingly and systematically holds the mentally ill morally responsible and calls them 'morally responsible', we might well say: 'You don't know what 'morally responsible' means!'

If the Strawsonian claim is a criterion thesis, do our practices of holding morally responsible then fix whether a given agent is morally responsible? The right answer here is, we believe, 'partly'. Our practices fix what the responsibility-making and –defeating facts are, but whether these facts obtain depends on how the world is. Our practices fix that having certain mental capacities is a responsibility-making fact, and if it turns out that an agent has these capacities (and other responsibility-making facts obtain while no defeating facts obtain), then she is morally responsible.

> [Criterion thesis] A is morally responsible if and only if, and because, the responsibility-making facts obtain (if and only if, for example, and among other things, an agent A has certain mental capacities) and no responsibility-defeating facts obtain.
>
> [World] The responsibility-making facts obtain (among other things, A actually has the required mental capacities) and no responsibility-defeating facts obtain.
>
> [Conclusion] A is morally responsible.

Thus, it may be said that our practices partly fix whether a given agent is morally responsible: they do not fix whether A is morally responsible 'by themselves', but they fix it in combination with how the world is.

The second type of thesis is not just a criterion thesis (though it includes a criterion thesis), but an obtainment thesis: our practices of holding responsible do not only fix what it means to be morally responsible, they also fix whether a given agent is morally responsible. They do not just fix the facts of meaning, they also fix the facts of the world. An obtainment

thesis is not just a thesis about what the responsibility-making and –defeating facts are, but it includes the further idea that our practices or dispositions of holding morally responsible *make it the case* that these responsibility-making facts *obtain* and that no responsibility-defeating facts obtain: an agent A is morally responsible if and only if, and because, we hold the agent morally responsible or are disposed to hold the agent morally responsible. The 'facts of the world' here are nothing more than facts about us. Understood as an obtainment thesis, the claim that our practices of holding morally responsible fix the facts of moral responsibility is the claim that in *holding* a given agent A morally responsible (or in being disposed to hold an agent morally responsible), we make it the case that the responsibility-making facts obtain and that no responsibility-defeating facts obtain, and, therefore, that A *is* morally responsible. If the Strawsonian thesis is an obtainment thesis, our practices of holding morally responsible fix whether a given agent is morally responsible. They do not fix it 'partly', as they would do if the Strawsonian thesis were a criterion thesis, but fully and straightforwardly.

2. No Absurdity Problem for a Criterion Thesis

The result of our discussion so far is that the so-called 'reversal claim', which states that our practices of holding morally responsible fix the facts of moral responsibility (rather than the other way round), is ambiguous. Either it is read as an obtainment thesis, or it is interpreted as a criterion thesis. If it is read as an obtainment thesis, the reversal claim does not allow for the possibility of *mistakenly* holding someone responsible. As it seems clear that we must allow for cases where we may hold (or be disposed to hold) an agent morally responsible when the agent is not in fact morally responsible, the reversal claim (understood as an obtainment thesis) is defective exactly because it states that holding an agent morally responsible (or being disposed to hold an agent morally responsible) makes it the case that the responsibility-making facts

This is the preprint version of an article published in *Pacific Philosophical Quarterly* 100 (3), 790-811. The final publication is available at Wiley Online: https://onlinelibrary.wiley.com/doi/10.1111/papq.12276. Please cite the published version only.

obtain, i.e. that the agent *is* morally responsible. Critics are right when they point out that the reversal claim read as an obtainment thesis is implausible.

An obtainment thesis is implausible when it comes to *our* practices, but this is not necessarily so when it comes to *God's* practices, and this difference indicates why Todd's divine command analogy breaks down. Our practices (at least according to the criterion thesis) fix only what the responsibility-making and –defeating facts are, and whether these facts actually obtain depends on how the world is. How the world is, is not fully under our control, but one may think that it is under God's control. It could be argued that God, by commanding murder, could *make it the case* that murder is right, or at least that he could make it the case that a given agent is morally responsible. God has control over the facts of the world and can change them at will, while we cannot, by changing our practices, *make it the case* that a given agent is morally responsible. At most, a change in our practices could make it the case that the *meaning* of 'morally responsible' would change in such a way that a non-responsible agent would rightly be called 'morally responsible', where 'morally responsible' would not mean *morally responsible* (what 'morally responsible' means for us now), but *morally responsible**. Only facts of meaning, but no facts about the world would thereby be changed.

With respect to our practices, a criterion thesis is much more plausible than an obtainment thesis. Moreover, Strawsonians have good reason to endorse such a reading, because the reversal claim read as a criterion thesis does not lead to an unwelcome result about the mentally ill. If Strawson's thesis is a criterion thesis, the mentally ill would not be morally responsible if we were to hold them morally responsible. A Strawsonian reply to the absurdity challenge can be outlined as follows:

(1) [Criterion thesis] Our practices determine what the criteria for moral responsibility (or the responsibility-making and responsibility-defeating facts) are.

(2) A criterion for moral responsibility (or responsibility-making fact), determined by our current practices, is that a morally responsible agent should have certain mental capacities that the mentally ill do not have (or, alternatively, that a morally responsible agent should be a possible fit target of the reactive attitudes, or that a morally responsible agent should be able to act from a will with morally qualitative content).

(3) If our practices were such that we knowingly and systematically held agents without those mental capacities (or agents who are not possible fit targets of the reactive attitudes, or agents without the ability to act from a will with morally qualitative content), such as the mentally ill, morally responsible, and if our practices determine the criteria for moral responsibility [criterion thesis], then having those mental capacities (or being a possible fit target of the reactive attitudes, or being able to act from a will with morally qualitative content) could not be a criterion for moral responsibility.

(4) If the criteria for moral responsibility (or the responsibility-making and responsibility-defeating facts) determined by the practices in which we hold the mentally ill morally responsible are different from the criteria for moral responsibility determined by our current practices, then there are two different ideas of moral responsibility at play, because criteria at least partially determine the meanings of expressions for which they are criteria (Hacker 1993: 250). The idea of moral responsibility determined by the practices in which we hold the mentally ill morally responsible is not the idea of moral responsibility determined by our current

practices. It is not our idea of moral responsibility, not moral responsibility 'as we know it'.[3] It is moral responsibility*, not moral responsibility.

(5) If we knowingly and systematically held the mentally ill morally responsible, they would be morally responsible*. Because moral responsibility* and moral responsibility are different concepts governed by different criteria, there is no *prima facie* reason to worry about this result.[4]

A similar argument could be developed for the examples about tables and funniness. We will not undertake the task of specifying what the criteria for being a table or being funny, as determined by our current practices, are, and how the change in practices described in the examples would affect them, but we believe that in both cases, the radical change in our practices would entail a change in criteria.

3. Beglin's Concern-Based Approach

[3] The formulation 'as we know it' is Strawson's. He wants to recover from the 'moral life as we know it', from 'the facts as we know them' (a formulation he uses twice in the same paragraph), 'a sense of what *we* mean […] when […] we speak of […] responsibility' (emphasis added) (Strawson 2008a: 168).

[4] There is, of course, reason to worry about a world in which we knowingly and systematically hold the mentally ill morally responsible (as there is reason to worry about a world in which we laugh at genocide). But the idea that there is reason to worry about a world in which we hold the mentally ill morally responsible is different from the idea that there is reason to worry about a world in which the mentally ill would be morally responsible*.

This is the preprint version of an article published in *Pacific Philosophical Quarterly* 100 (3), 790-811. The final publication is available at Wiley Online: https://onlinelibrary.wiley.com/doi/10.1111/papq.12276. Please cite the published version only.

As we have indicated in the introduction, versions of the absurdity challenge have been around in the literature on Strawson and moral responsibility for a while. However, we agree with Todd's (2016) finding that, when he wrote his article, no distinctively Strawsonian, plausible and explicit answers to the absurdity challenge had been developed, and it is our aim to fill this gap. Our answer is much more explicit than extant Strawsonian suggestions (see Todd 2016 for examples), but that does not imply, of course, that it is incompatible with them (see below).

Todd's challenge has not gone unnoticed. In a reply to Todd, Beglin (2018) has worked out the only detailed and explicit Strawsonian answer to the absurdity challenge that is currently on offer.[5] According to Beglin, Strawson's reversal is the view 'that what it means to be morally responsible is determined (in some way) by our practices of holding responsible' (Beglin 2018: 612). This is right: as we have explained, our practices indeed determine the facts of *meaning*, not the facts of the world. Beglin's 'concern-based construal' of the reversal emphasizes that our practices of holding responsible express basic social concerns grounded in our social

[5] Shoemaker (2017) has recently defended a fitting response-dependence theory of moral responsibility. He claims that his theory is not vulnerable to Todd's challenge (Shoemaker 2017: 483), but he does not explicitly address it in detail. Shoemaker's theory is complex and requires lengthier treatment than we can give it here. We believe, in short, that the theory has a certain instability, depending on whether one emphasizes the descriptive *response*-element or the normative *fittingness*-element in his fitting response-dependence account. The exact relation between these aspects is not entirely clear. The criterion thesis, according to which our practices (descriptive element) determine the criteria (normative element) of moral responsibility, is one way of making explicit how both elements hang together. It could thus be seen as a proposal to develop Shoemaker's theory in a way that avoids the absurdity challenge.

sentimental nature. An example of such a concern is our concern about the quality of will with which people act (Beglin 2018: 620), but Beglin is not committed to a specific account of the relevant basic concerns. The main point is that there are some basic, human social concerns, 'deep-seated features of human psychology and sociality' (Beglin 2018: 621). These concerns can be expressed in different ways: different cultures have different responsibility practices. But these variable responsibility practices express the same basic human concerns:

> There do, after all, seem to be certain general patterns across cultures' responsibility practices. Strawson's quality-of-will proposal is perhaps a good example of such a pattern. All human communities seem to recognize the difference between someone's offering help out of genuine concern and someone's offering help for self-serving purposes. (Beglin 2018: 622)

So how does Beglin answer Todd's absurdity challenge? What if we held the mentally ill morally responsible? Beglin's answer, as we understand it, is that holding the severely mentally ill morally responsible (not occasionally, but knowingly and systematically) would not be an expression of the basic concerns underlying our human responsibility practices. It would thus not be a recognizably human practice. Whether a severely mentally ill person is morally responsible does not seem to be among the things that can vary by community (Beglin 2018: 622, n19). Thus, roughly, Beglin's answer to Todd's 'What if we held the severely mentally ill morally responsible?' is 'We cannot, because it's not in our nature.'

This, we take it, is correct (see Beglin 2018 for defense). But, as Beglin recognizes, it is not enough. He makes an empirical point about humans and human responsibility practices ('this is an empirical matter', Beglin 2018: 622), but what if we were different? What if there was a community of rational agents with responsibility practices that express different basic concerns? What if these rational agents held the severely mentally ill morally responsible?

This is the preprint version of an article published in *Pacific Philosophical Quarterly* 100 (3), 790-811. The final publication is available at Wiley Online: https://onlinelibrary.wiley.com/doi/10.1111/papq.12276. Please cite the published version only.

Would the mentally ill then *be* responsible? These are the questions that, we imagine, those who put forward the absurdity challenge will continue to ask in the light of Beglin's concern-based construal. Beglin addresses them briefly (Beglin 2018: 623-624), and he remarks:

> And it isn't obvious that rational agents who relate to each other in fundamentally different ways, who have fundamentally different evaluative standpoints, which are constituted by fundamentally different basic concerns, would or should operate with the same conceptual equipment as we do … (Beglin 2018: 623)

Our response to the absurdity challenge could be seen as an elaboration of this suggestion. We have tried to show how exactly the reversal, read as a criterion thesis (our practices determine the meaning of moral responsibility), allows for the thought that Beglin only tentatively expresses: that a fundamental change in concerns and practices will result in a conceptual change, a change in the meaning of moral responsibility.

We believe that our account and Beglin's are mutually supportive. Beglin explains how human responsibility practices are rooted in basic concerns, thus emphasizing the Humean, naturalistic strand of 'Freedom and Resentment'.[6] He shows that holding the mentally ill morally responsible does not square with human nature, so that the absurdity challenge gets no grip on us. We explain how concepts are rooted in practices (which is perfectly compatible with Beglin's idea that practices are rooted in basic concerns), thus emphasizing the Wittgensteinian, conceptual strand in Strawson's account. We show that, even if (human or non-human) practices were radically different, even if the mentally ill would be held morally responsible,

[6] Shoemaker (2017: 482, 519) also emphasizes the Humean, naturalistic strand of Strawson's essay. He could thus be taken to suggest an answer to the absurdity challenge that is similar to Beglin's. In that case, what we say about Beglin would hold for Shoemaker as well.

this would still not result in their being responsible. Even if the empirical claims on which Beglin's argument rests are false, the absurdity challenge can be warded off.

The combination of a naturalistic and a conceptual strand fits well with Strawson's overall philosophical practice. In methodological essays, Strawson distinguishes between the philosopher's *descriptive, explanatory,* and *imaginative* tasks (see Strawson 1963, 2011a, 2011b). The descriptive task is to describe our conceptual scheme and is compatible with, and supported by, the explanatory task of attempting 'to show the natural foundations of our logical, conceptual apparatus, in the way things happen in the world, and in our own natures' (Strawson 1963: 516; Strawson 2011b: 86). Strawson clearly takes these two strands to be compatible. They function as constraints on the *imaginative* task of the philosopher. Although the latter task has a positive function as well, it often leads to 'merely rudimentary mistakes' (Strawson 2011b: 87) that result from a misrecognition of our actual use of linguistic expressions, which 'remains his [the philosopher's] sole and essential point of contact with the reality which he wishes to understand, conceptual reality' (Strawson 2011b: 90). An imagined radical change in our responsibility practices would not only imply that we would have to be creatures with different natures (as Beglin correctly points out); we would have to be creatures with different concepts as well. Not only would we have to transgress the bounds of our humanity, we would also have to transgress the bounds of our actual conceptual scheme. In the terminology of this article, we would treat one another as morally responsible*, not morally responsible.

4. Questions and Answers

We have distinguished between two readings of the Strawsonian reversal claim. We have argued that the absurdity challenge can be answered, as outlined in section two, if the reversal claim is read as a criterion thesis. We have discussed the relation between our proposal and

Beglin's response to the absurdity challenge. In this section, we look at some possible questions about our proposal.

(1) We have argued that the reversal claim read as a criterion thesis is not vulnerable to the absurdity challenge. Our practices of holding responsible do not make it the case *that* the responsibility-making facts obtain and that the responsibility-defeating facts do not obtain (the facts of the world); instead, they fix *which* facts count as responsibility-making and –defeating facts in the first place (the facts of meaning). But *how* do our practices of holding responsible fix these facts of meaning? Is this claim plausible and compatible with contemporary approaches in the philosophy of language?

We believe that our proposal most naturally goes together with use-based, broadly Wittgensteinian accounts of meaning. Take, for example, Paul Horwich's use-based metasemantics. Horwich's explanatory strategy takes as basic facts the (1) law-like *regularities* of word use (that can be specified in non-semantic terms). These facts then (2) engender facts about the *implicit* rules we follow when using these words and these, in turn, (3) fix the facts about the meaning of the words and sentences we use (Horwich 2010: 113). Similarly, we take it that (1) our practices of holding people morally responsible (which can be characterized in non-semantic terms) are the basic facts that (2) engender facts about the criteria (or rules) we implicitly follow when using the word 'responsibility' and that these, in turn, (3) fix the 'facts of meaning' of moral responsibility. If our practices would radically change (for example, if we would hold the mentally ill morally responsible), the criteria we implicitly follow and, therefore,

This is the preprint version of an article published in *Pacific Philosophical Quarterly* 100 (3), 790-811. The final publication is available at Wiley Online: https://onlinelibrary.wiley.com/doi/10.1111/papq.12276. Please cite the published version only.

the 'facts of meaning' would change as well.[7] 'Moral responsibility' would no longer mean what it means now for us.

(2) Given that our aim is to provide a Strawsonian answer to the absurdity challenge, it is appropriate to ask whether Strawson himself should be read as defending a criterion thesis. Did Strawson have a criterion thesis in mind?

The answer, we believe, is 'yes' (see also the last paragraph of section 3). First, in 'Freedom and Resentment' Strawson explicitly aims to 'recover from the facts as we know them a sense of what we *mean,* i.e., of *all* we *mean,* when, speaking the language of morals, we speak of desert, responsibility, guilt, condemnation, and justice' (Strawson 2008a: 168, emphasis added). Given that a criterion thesis about moral responsibility states that our practices of holding responsible fix the *meaning* of moral responsibility, our reading aligns well with Strawson's claim that he wishes to understand what we mean when we speak of (moral) responsibility. Second, Strawson supported a largely Wittgensteinian account of meaning (see footnote 1), and our proposal most naturally goes together with that type of account (see our reply to the first question above). Third, Strawson's focus on the meaning of responsibility 'as we know it' exemplifies the general philosophical method that he explicitly sets out and develops in other works. Strawson calls his method, as it is for example defended and practiced in *Individuals* (1959), 'descriptive metaphysics'. The aim of descriptive metaphysics is, first

[7] On this point, see also Hacker (1993: 255-256) and Glock (1996: 95): '[…] a change in criteria is a conceptual change, a change in the meaning of words: that *q* is a criterion for being *F* is partly constitutive of the concept of being *F*.' Glock (1996: 95-97) discusses (and, in our view, successfully refutes) several objections to this view.

and foremost, to describe our conceptual scheme and, further, to explain how 'the nature of our thinking is rooted in the nature of the world and in our natures' (Strawson 2011a: 36).

In his insightful discussion of Strawson's method, Hacker suggests that the term 'metaphysics' may mislead one here. Strawson aims to lay bare conceptual connections, and although these connections are rooted in our natures (that is, they would probably be different if our natures were different), 'it would be absurd to argue from conceptual connections in thought to existential truths about the world, or, in Wittgensteinian idiom, from grammatical propositions to empirical ones' (Hacker 2001: 363). In our terminology, it would be mistaken to think that Strawson's aim was to put forward an (existential) obtainment thesis rather than a (conceptual) criterion thesis. The descriptive metaphysician describes what we have called the 'facts of meaning' (or, as Strawson put it, the 'bounds of sense'; see Strawson 1975), not the facts of the world.

(3) We did once have criteria for moral responsibility different from the ones that we have now, and some cultures may have criteria that are different from ours. Does this imply that 'moral responsibility' means something different for them from what it means for us?

This worry can be mitigated by emphasizing that sameness of meaning should not be thought of as an all-or-nothing affair. Our concept of moral responsibility has evolved, and other people may have a somewhat different but overlapping concept. Whether a change in practices entails a change in criteria and a change in meaning, whether a change in meaning generates a modification of the concept rather than an entirely different concept, depends, in general, on how radical the change in practices is. The change in practices required by the absurdity challenge is so radical that the concept of moral responsibility* fixed by these practices would be very different from our concept of moral responsibility.

This is the preprint version of an article published in *Pacific Philosophical Quarterly* 100 (3), 790-811. The final publication is available at Wiley Online: https://onlinelibrary.wiley.com/doi/10.1111/papq.12276. Please cite the published version only.

(4) It may well be the case that our practices fix what 'table' means, and/or that they fix what 'funny' means, because it could not turn out that nothing we have ever treated as a table (or as funny) really was a table (or funny). But moral responsibility may be different. Maybe it is not to be compared to tables, but rather to witches, and in the case of witches we can say that nobody we have ever treated as a witch really was a witch. Does that mean that in some cases our practices do not fix the facts?

Not if this is read as a criterion thesis, that is, as the claim that our practices fix what the witch-making facts are. Our practices determine the criteria for witchhood, that is, how the facts would have to be in order for there to be witches: if there were women with evil magic powers flying on broomsticks, there would be witches. Our practices do not determine whether there are *in fact* any witches; that is up to the world. So even if moral responsibility is more like witches than like tables, the criterion thesis is not in trouble.

(5) We have distinguished between being morally responsible and being morally responsible*. But it seems like both senses of 'morally responsible' would entail its being appropriate to hold someone responsible. Is this not what proponents of the absurdity challenge are really worried about? The fact that we hold certain agents responsible does not make it *appropriate* to hold them responsible.

We would like to thank an anonymous reviewer for pressing this objection. We see two (distinct but related) points here, and we will address them in turn. First, suppose that we accept the conclusion of our argument in section two: if we knowingly and systematically held the mentally ill morally responsible, they would not be responsible, but they would be responsible*. The objection takes as its starting point that being responsible entails appropriately being held

responsible, and we accept that point. If the mentally ill were responsible, it would be appropriate to hold them responsible. Similarly, if the mentally ill were responsible*, it would be appropriate to hold them responsible*. This point, in combination with the conclusion of our argument in section two, thus leads to the following: if we knowingly and systematically held the mentally ill morally responsible, they would be responsible* (conclusion argument section two), and it would thus be appropriate to hold them morally responsible*. Should this worry us?

We do not think so (see, in this respect, also footnote 4). Recall that the concept of moral responsibility*, fixed by a radical change in our practices, would be very different from our concept of moral responsibility. Because the concept of responsibility* is very different from our concept of responsibility, it is to be expected that holding responsible* will be very different from holding responsible. Just as it is appropriate to hold people responsible only if they are responsible, it is appropriate to hold people responsible* only if they are responsible*. The responsible* include the severely mentally ill, so holding people responsible* (in contrast to holding people responsible) will only include attitudes appropriately adopted towards the severely mentally ill. It will not include reactive attitudes, but only 'objective' ones: no resentment or indignation, but disappointment or sadness, for example. This explains why we can accept the idea that holding certain agents responsible does not make it appropriate to hold those agents responsible: if we were to hold the mentally ill morally responsible, it would not follow that it would be appropriate to hold them morally responsible, only that it would be appropriate to hold them morally responsible*. But holding morally responsible* is very different from holding morally responsible.

The second worry that we discern in the objection is that our argument does not allow for the distinction between appropriately and inappropriately holding people responsible. If our

practices of holding responsible determine the criteria for moral responsibility, can we then make this distinction? It is important here to distinguish between two levels of appropriateness. First, suppose that an agent A holds another agent B morally responsible. On what basis can we say that A does so (in)appropriately? If our practices determine, for example, that having certain mental capacities is a criterion for moral responsibility, then (other things being equal) A will appropriately hold B responsible when B has those capacities (and the other criteria for moral responsibility are also satisfied), and A will inappropriately hold B responsible when B does not have them. Our criteria are norms against which individual ascriptions of responsibility can be measured (see also section one: not every change in our practices entails a change in criteria).

This, we take it, is rather unproblematic. But questions about appropriateness can also be raised at a second level: on what basis can we say that the criteria for moral responsibility, determined by *our* practices of holding responsible, are the appropriate ones? According to Strawson (2008a: 25), this is a question for 'external justification' of our practices and the criteria determined by them. Questions for internal justification (first-level appropriateness questions) can be answered by reference to our criteria, but what standard would we use to determine whether the general framework of our criteria and practices itself is the appropriate one? To ask for an external rational justification is, according to Strawson, to over-intellectualize the facts; it is to misunderstand that questions for justification only make sense *within* a framework of practices and criteria. Strawson does not answer questions for external justification, he tries to show that they are misguided.[8] A full Strawsonian account of moral

[8] This is not to say that Strawson dismisses the need for *explanation* of our practices and criteria. As we have seen, Beglin shows that our practices are grounded in certain basic concerns. This explains why we have the practices that we have, but it does not justify them.

responsibility should explain and defend this part of Strawson's view in detail. We cannot do that here, but we believe that convincing explanations and defenses of Strawson's distinction between internal and external justification have been provided (see De Mesel 2018). We will only offer an interesting analogy made by Strawson himself that helps to understand why questions for external justification are problematic.

In the first chapter of *Analysis and Metaphysics* (1992), Strawson compares the philosopher's task to the grammarian's. Based on how people actually use language, the grammarian formulates rules of grammar. Similarly, the philosopher looks at how concepts are used and, on that basis, formulates what the criteria for their use are. It is, of course, possible for people to make grammatical mistakes, that is, to go against the rules of grammar. And it is also possible to think mistakenly that something is a rule of grammar. But what about the second-level question: 'Is the general framework of our grammatical rules, determined by the way we speak, appropriate?' Here one could only answer: 'It is the framework that we have, the framework by reference to which particular uses of language can be (in)appropriate.' Analogously, Strawson writes about our practices of holding morally responsible that they are 'given with the fact of human society', and that 'questions of justification are internal to the structure or relate to modifications internal to it' (Strawson 2008a: 25). The question, 'Why do our practices, as opposed to something else, determine the criteria for moral responsibility?' is, according to Strawson, relevantly analogous to, 'Why do our practices of language use determine the grammatical rules of our language?' From a Strawsonian perspective, these questions are misguided rather than in need of a substantial answer.

5. Incompatibilism and the Criterion Thesis

This is the preprint version of an article published in *Pacific Philosophical Quarterly* 100 (3), 790-811. The final publication is available at Wiley Online: https://onlinelibrary.wiley.com/doi/10.1111/papq.12276. Please cite the published version only.

We have argued that Strawsonians could avoid the absurdity challenge while maintaining that our practices of holding responsible fix the facts of moral responsibility, rather than the other way round. However, an important aspect of the reversal claim has not been mentioned: it is often thought to play a key role in a Strawsonian *compatibilist* strategy about free will and determinism. Todd (2016), for example, does not deny that Strawsonians could respond to the absurdity challenge (although he believes that extant responses oscillate between being implausible and being un-Strawsonian); rather, he denies that a plausible Strawsonian response to the absurdity challenge could entail compatibilism about free will and determinism, as many Strawsonians have suggested (see Todd 2016 for examples). Does the reversal, read as a criterion thesis, entail compatibilism?

We believe that the answer is 'no', and we agree with Todd's claim that no plausible response to the absurdity challenge could straightforwardly entail compatibilism. An incompatibilist could accept the idea that our practices fix the criteria for moral responsibility. Accepting it means forsaking a line of reasoning that is *not* consistent with the Strawsonian view. The incompatibilist could say: 'Why should I accept that what is relevant for the attribution of moral responsibility is determined by our practices? Our practices are one thing, moral responsibility is another.[9] The truth of determinism may not count as a responsibility-defeating fact in our practices, but that only shows that something is deeply wrong with our practices. If determinism is true, our practices of holding morally responsible are unjustified.' Why should the incompatibilist abandon this line of reasoning? Why accept the criterion thesis, that is, the idea that our practices, and not something else, determine the criteria for moral responsibility? This is the question that we discussed in our reply to objection (5) at the end of

[9] On this line of reasoning, see McKenna 2017: 77-78.

the previous section. From a Strawsonian perspective, the question is relevantly analogous to (and equally misleading as) the question why our practices of language use, and not something else, determine the grammatical rules of our language.

The fact that the criterion thesis can be accepted by incompatibilists does not preclude it from playing an ineliminable role in (at least one of) Strawson's compatibilist argument(s), which can be reconstructed as follows:

(1) [Criterion thesis] Our practices of holding morally responsible fix what the responsibility-making and responsibility–defeating facts are.

(2) Incompatibilists introduce a fact that does not count as responsibility-defeating in our practices of holding morally responsible (or at least a fact that, given our current practices, we do not take to have responsibility-defeating implications): the truth of determinism.

(3) It follows from (1) and (2) that the truth of determinism is not a responsibility-defeating fact. Moral responsibility is compatible with the truth of determinism.

While a criterion thesis has no immediate compatibilist implications, it is essential to a Strawsonian compatibilist argument. The fact that incompatibilists can accept it is not a problem for the Strawsonian, because the criterion thesis is not *meant* to entail compatibilism or to be unacceptable for incompatibilists. It can best be understood as nothing but a *reminder* of something that Strawson took to be quite uncontroversial, one of the 'commonplaces' he wanted to remind his readers of (Strawson 2008a: 152). He was looking for a possible way 'of reconciling these disputants [compatibilists and incompatibilists] to each other and the facts' (Strawson 2008a: 170). His point is not that incompatibilists have to *deny* the criterion thesis; rather, the point is that they (or at least those incompatibilists who are inclined to say that our practices are one thing and moral responsibility is another) have somehow tended to *overlook*

it, as well as some compatibilists who have been looking for a consequentialist justification of our practices of holding responsible and are equally guilty of 'over-intellectualizing' the facts (Strawson 2008a: 152). The reversal claim, rather than excluding incompatibilism, marks out the terrain on which, according to Strawson, the battle between compatibilists and incompatibilists should take place.

Only in combination with (2) does the criterion thesis yield a compatibilist result. But why accept (2)? Strawson not only attempts to remind us of the fact *that* our practices determine what the relevant criteria for moral responsibility are, he also indicates *what these criteria, implicit in our practices, are.* A criterion for moral responsibility suggested by Strawsonians is the ability to act from a will with morally qualitative content. The incompatibilist could agree, but has to add that this ability requires leeway freedom (the ability to do otherwise) or source freedom (or something else), and, therefore, the falsity of determinism. Similarly, the incompatibilist who accepts the criterion thesis could endorse the Strawsonian idea that being excused or exempted is a responsibility-defeating fact, but has to add that we already recognize within our practices that someone who could not have done otherwise, or someone whose actions were not up to him/her, is excused or exempted, that the truth of determinism would make it the case that nobody has leeway freedom or source freedom, and that, therefore, the truth of determinism would make it unfair to hold people responsible. Strawsonians deny these incompatibilist claims. But how could they?

Let us focus on leeway freedom first. Strawsonians could deny that the ability to do otherwise plays the role in our practices that incompatibilists need it to play. Wallace, for example, has argued that the standards of fairness implicit in our practices of holding responsible do not imply the principle that it is only fair to hold someone responsible if she could have done otherwise (Wallace 1994: chapters 5 and 6). Strawsonians could also accept

This is the preprint version of an article published in *Pacific Philosophical Quarterly* 100 (3), 790-811. The final publication is available at Wiley Online: https://onlinelibrary.wiley.com/doi/10.1111/papq.12276. Please cite the published version only.

the incompatibilist's claim that the ability to do otherwise plays a crucial role in our practices of holding responsible. Wiggins (1987) is a perfect example of an incompatibilist who clearly accepts the criterion thesis, but maintains that the ability to do otherwise, with the meaning determined by our practices ('we are characterizing an existing everyday notion', Wiggins 1987: 285), is incompatible with determinism. As Wiggins (1987: 299) suggests (and we agree here), Strawson thought it too obvious in 'Freedom and Resentment' that the incompatibilist's acceptance of the criterion thesis would lead to a compatibilist position (not because he thought that, in itself, the criterion thesis entails compatibilism, but because step (2) of the compatibilist argument outlined above is almost taken for granted).

Thus, Wiggins accepts to fight the compatibilist on the terrain demarcated by Strawson's criterion thesis. The dispute between compatibilist and incompatibilist does not concern the truth of the criterion thesis, but deals with the further question of *which* criteria are fixed by our practices of holding responsible. Strawsonians could accept that the ability to do otherwise is central to our practices, but if they want compatibilism, they have to make clear that, given what 'ability to do otherwise' *means* within our practices, the truth of determinism would not threaten the ability to do otherwise *in that sense* (it could, at most, threaten an ability to do otherwise*). And this seems to be Strawson's own position:

> It is certainly true that often, in the context of a moral judgment (especially if disapprobative) one may utter the words, "He could have acted otherwise," or other words to the same effect. But are such words, as then uttered, really equivalent to "There was no sufficient natural impediment or bar, *of any kind whatsoever, however complex,* to his acting otherwise"? I find it difficult, as others have found it difficult, to accept this equivalence. The common judgement of this form amounts rather to the denial of any sufficient natural impediment *of certain specific kinds or ranges of kinds.* For example,

This is the preprint version of an article published in *Pacific Philosophical Quarterly* 100 (3), 790-811. The final publication is available at Wiley Online: https://onlinelibrary.wiley.com/doi/10.1111/papq.12276. Please cite the published version only.

> "He could (easily) have helped them (instead of withholding help)" may amount to the denial of any lack on his part of adequate muscular power or financial means. Will the response, "It simply wasn't in his nature to do so" lead to a withdrawal of moral judgment in such a case? I hardly think so; rather to its reinforcement. (Strawson 1992: 136-137)

This passage does not occur in 'Freedom and Resentment'. Strawson may have thought it quite obvious that phrases such as 'He had to do it', 'It was the only way' and 'They left him no alternative', *as we mean them within our practices*, would not be true of everyone in all circumstances if determinism were true. Strawson could (and should) have tried to show in some detail that the criteria for the use of 'He could have acted otherwise', 'It was the only way', etc. within our practices are different from the criteria for the incompatibilist's use of 'He could have acted otherwise', 'It was the only way', etc. What would thereby be shown is that, while the compatibilist means by these expressions what these expressions mean in our practices, the incompatibilist means something different. If the incompatibilist accepts the criterion thesis however, she is committed to the idea that our practices fix the meaning of these expressions. If she uses 'ability to do otherwise' in a way that is out of line with how it is used in our practices, if her criteria for the application of the concept are radically different, she will be talking about the ability to do otherwise*. Even if that ability is incompatible with the truth of determinism, the Strawsonian compatibilist could maintain that the ability to do otherwise, for which the criteria are fixed by *our* practices, is compatible with the truth of determinism. It is clear that such a strategy is only hinted at in Strawson's work and would need further elaboration.

Not all threats to free will arise due to a supposed lack of leeway freedom. Instead, source incompatibilists emphasize that free actions must be in some sense up to us, that actions

can only be free if we are the source or origin of these actions. Strawson does not deny these perfectly familiar points (see Strawson 1992: 134-138), but notes, referring to P. H. Nowell-Smith (1954), that the incompatibilist's language 'is apt to alternate between the very familiar and the very unfamiliar' (Strawson 2008a: 3). 'The very unfamiliar' is talk of 'contra-causal freedom' (Strawson 2008a: 25) or the idea that free actions must issue from 'uncaused acts of will' (Strawson 1980: 260). These expressions are highlighted by Nowell-Smith (1954: 319-322), whose article is mainly a discussion of C. A. Campbell's *Scepticism and Construction* (1931). It is clear that Strawson's characterization of the pessimist draws heavily on Nowell-Smith's critique of Campbell, and it is no coincidence that, in Jonathan Bennett's (1980) discussion of 'Freedom and Resentment', Campbell figures as the prototypical incompatibilist.

Strawson suggests, following Nowell-Smith, that contra-causal freedom or uncaused acts of will play no role in our ordinary concepts of freedom and moral responsibility. Bennett casts some doubt on this claim. He refers to Christian thought and the influence of the Kantian idea that 'the source of moral thought and action must be located outside the empirically conditioned self' (Bennett 1980: 26). Bennett's point about source freedom is structurally similar to Wiggins' point about leeway freedom: is it not the case that this kind of freedom plays the role in our practices that incompatibilists who accept the criterion thesis need it to play?

Again, we concede that Strawson did not say nearly enough about this in 'Freedom and Resentment'. And again, a passage from other work by Strawson contains the germ of a Strawsonian reply to the incompatibilist who accepts the criterion thesis:

> Bennett himself is prepared, or half-prepared, to allow this [Christian-Kantian] thought some place, though not a dominant place, in 'our ordinary concept of accountability'. But there is a quite general ambiguity in the notion of 'our ordinary concept' of whatever

> it may be. Should the lineaments of such a concept be drawn exclusively from its use, from our ordinary *practice,* or should we add the reflective accretions, however confused, which, naturally or historically, gather around it? The distinction is hardly clear-cut; but where it can be made, I prefer the first alternative. (Strawson 1980: 265)

Strawson confirms what we have been emphasizing: our ordinary practice determines the criteria for moral responsibility ('lineaments of such a concept'). He adds, however, that our practice does not include 'the reflective accretions … which … gather around it', and understands the Christian-Kantian conception of freedom and responsibility as such a reflective accretion. An analogy used earlier helps to explain the distinction that Strawson has in mind here. Strawson compares the philosopher's task to the grammarian's. The grammarian's primary aim is to formulate rules of language based on how people actually use language (the ordinary practice), not on how people have formulated or understood the rules of language (the reflective accretions gathered around it). Similarly, the philosopher primarily takes into account the ordinary rather than the philosophical use of a concept (see, for this contrast, also Nowell-Smith 1954: 319-322).

That does not mean, of course, that reflective accretions cannot influence ordinary concepts, or that the Christian-Kantian thought is, as Strawson thinks, a reflective accretion rather than a part of the ordinary concept. Just as it is, according to Strawson, the compatibilist's task to show that the Christian-Kantian thought is only a reflective accretion, it is the incompatibilist's task to show that her criteria for moral responsibility are the criteria determined by our practices. And just as Wiggins is an example of someone who has taken up the Strawsonian challenge in trying to show that, if determinism were true, we could not have the ability to do otherwise in the sense of that expression determined by our practices, it could be argued that incompatibilists have tried to show that, if determinism were true, we could not

This is the preprint version of an article published in *Pacific Philosophical Quarterly* 100 (3), 790-811. The final publication is available at Wiley Online: https://onlinelibrary.wiley.com/doi/10.1111/papq.12276. Please cite the published version only.

be the source or origin of our actions in the sense of these expressions determined by our practices (one could think here of Derk Pereboom's (2014) source incompatibilism). It is beyond the scope of this paper to evaluate the extent to which these proposals are successful. Suffice it to say that, according to our understanding of Strawson, popular incompatibilist appeals to some 'ultimate' kind of freedom or sourcehood, or to 'basic' or 'genuine' or 'real' desert, often signal that an alternation from a very familiar concept to a much less familiar one has taken place (Strawson 1980: 264; Strawson 2008a: 3; see also Nowell-Smith 1954: 320).

What, then, is Strawson's contribution to the debate between compatibilists and incompatibilists? The reversal claim rules out forms of compatibilism and incompatibilism which rest on a denial of the criterion thesis, a denial according to which our practices are one thing and moral responsibility is another. Compatibilists targeted by Strawson's criticisms include Schlick (1939) and Nowell-Smith (1948) (for more references, see Bennett 1980: 19, footnote 5), the main incompatibilist target was Campbell (1931).[10] Although Strawson embraces a form of compatibilism in 'Freedom and Resentment', he does not offer a clear path from the reversal to compatibilism (that is, a defense of premise (2) in the argument outlined above), mainly because, as Wiggins remarks, he thought it too obvious that the reversal claim would lead to compatibilism. However, in other works, often neglected in discussions of Strawson's view, Strawson sketches compatibilist replies to both leeway and source

[10] It may be argued that there are not many contemporary incompatibilists who favor this line of reasoning (see Wallace (1994: 97)). Strawsonians could agree with this point: it could be taken to show that the reminder has been successful and has been accepted by compatibilists and incompatibilists alike. It is beyond the scope of this paper to evaluate whether or to what extent contemporary (in)compatibilists accept the criterion thesis.

incompatibilists who accept the criterion thesis. We believe that a thorough exploration of these replies, which we cannot offer here, will help those who seek a path from the reversal claim to a plausible form of Strawsonian compatibilism. We hope to have shown that the reversal claim itself is best understood, both on exegetical and substantive philosophical grounds, as a criterion thesis.

References

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

Bennett, Jonathan. (1980) Accountability. In Zak van Straaten (ed.), *Philosophical Subjects. Essays Presented to P. F. Strawson.* Oxford, Clarendon, 14-47.

Brink, David and Dana Nelkin. (2013) Fairness and the Architecture of Responsibility. In David Shoemaker (ed.), *Oxford Studies in Agency and Responsibility. Volume 1.* Oxford, Oxford University Press, 284-313.

Campbell, C. A. (1931) *Scepticism and Construction.* London, George Allen & Unwin Ltd.

Coates, Justin and Neal Tognazzini. (2013) *Blame. Its Nature and Norms.* Oxford, Oxford University Press.

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, 603-614.

Ekstrom, Laura. (2000) *Free Will. A Philosophical Study.* Boulder (CO), Westview Press.

Fischer, John Martin and Mark Ravizza. (1993) Introduction. In John Martin Fischer and Mark Ravizza (eds.), *Perspectives on Moral Responsibility.* New York, Cornell University Press.

Fischer, John Martin. (1994) *The Metaphysics of Free Will.* Oxford, Blackwell.

This is the preprint version of an article published in *Pacific Philosophical Quarterly* 100 (3), 790-811. The final publication is available at Wiley Online: https://onlinelibrary.wiley.com/doi/10.1111/papq.12276. Please cite the published version only.

Glock, Hans-Johann. (1996) *A Wittgenstein Dictionary.* Oxford, Wiley-Blackwell.

Hacker, P.M.S. (1993) *Wittgenstein. Meaning and Mind. Volume 3 of An Analytical Commentary on the Philosophical Investigations. Part I. Essays.* Oxford, Wiley-Blackwell.

Hacker, P.M.S. (2001) On Strawson's Rehabilitation of Metaphysics. In his *Wittgenstein. Connections and Controversies.* Oxford, Oxford University Press, 345-370.

Horwich, P. (2010) *Truth – Meaning – Reality.* Oxford, Oxford University Press.

Kane, Robert. (2005) *A Contemporary Introduction to Free Will.* Oxford, Oxford University Press.

McKenna, Michael. (2017) Theories of Moral Responsibility and the Responsibility Barter Game. In Zachary Goldberg (ed.), *Reflections on Ethics and Responsibility. Essays in Honour of Peter A. French.* Cham, Springer, 71-84.

Nelkin, Dana. (2011) *Making Sense of Freedom and Responsibility.* Oxford, Oxford University Press.

Nowell-Smith, P.H. (1948) Freewill and Moral Responsibility. *Mind* 57, 45-61.

Nowell-Smith, P.H. (1954) Determinists and Libertarians. *Mind* 63, 317-337.

Pereboom, Derk. (2014) *Free Will, Agency, and Meaning in Life.* Oxford, Oxford University Press.

Schlick, Moritz. (1939) *Problems of Ethics.* New York, Prentice Hall.

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

Strawson, P.F. (1959) *Individuals. An Essay in Descriptive Metaphysics.* London, Routledge.

Strawson, P.F. (1963) Carnap's Views on Constructed Systems versus Natural Languages in Analytic Philosophy. In P.A. Schilpp (ed.), *The Philosophy of Rudolf Carnap*. La Salle, Open Court, 503-518.

Strawson, P.F. (1975 [1966]) *The Bounds of Sense. An Essay on Kant's* Critique of Pure Reason. London, Routledge.

Strawson, P.F. (1980) P.F. Strawson Replies. In Zak van Straaten (ed.), *Philosophical Subjects. Essays Presented to P.F. Strawson.* Oxford, Clarendon, 260-296.

Strawson, P.F. (1992) *Analysis and Metaphysics. An Introduction to Philosophy.* Oxford, Oxford University Press.

Strawson, P.F. (2008a [1962]) 'Freedom and Resentment'. In P.F. Strawson, *Freedom and Resentment and Other Essays.* London, Routledge, 1-28.

Strawson, P.F. (2008b [1985]) *Scepticism and Naturalism. Some Varieties.* London, Routledge.

Strawson, P.F. (2011a [1956]) Construction and Analysis. In his *Philosophical Writings*. Oxford, Oxford University Press, 30-38.

Strawson, P.F. (2011b [1967]) Analysis, Science, and Metaphysics. In his *Philosophical Writings*. Oxford, Oxford University Press, 78-90.

Todd, Patrick. (2016) Strawson, Moral Responsibility, and the 'Order of Explanation'. An Intervention. *Ethics* 127, 208-240.

Tognazzini, Neal. (2013) Blameworthiness and the Affective Account of Blame. *Philosophia* 41, 1299-1312.

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

Watson, Gary. (2004 [1987]) Responsibility and the Limits of Evil. In his *Agency and Answerability. Selected Essays.* Oxford, Oxford University Press, 219-259.,

Watson, Gary. (2014) Peter Strawson on Responsibility and Sociality. In David Shoemaker and Neal Tognazzini (eds.), *Oxford Studies in Agency and Responsibility. Volume 2.* Oxford, Oxford University Press, 15-32.

This is the preprint version of an article published in *Pacific Philosophical Quarterly* 100 (3), 790-811. The final publication is available at Wiley Online: https://onlinelibrary.wiley.com/doi/10.1111/papq.12276. Please cite the published version only.

Wiggins, David. (1987 [1973]) Towards a Reasonable Libertarianism. In his *Needs, Values, Truth.* Oxford, Basil Blackwell, 269-300.