**P.F. Strawson was neither an externalist nor an internalist about moral responsibility**

*Abstract*

Internalism about moral responsibility is the view that moral responsibility is determined primarily by an agent’s mental states; externalism is the view that moral responsibility is determined primarily by an agent’s overt behaviour and by circumstances external to the agent. In a series of papers, Michelle Ciurria (2014a; 2014b; 2015) has argued that most if not all current accounts of moral responsibility, including Strawsonian ones, are internalist. Ciurria defends externalism against these accounts, and she argues that, in contrast to his contemporary followers, P.F. Strawson himself was an externalist. I believe that Ciurria’s reading of Strawson is problematic. The aim of this paper is to elucidate Strawson’s position with regard to the internalism-externalism issue against the background of Ciurria’s reading of him. I conclude that Strawson was neither an internalist nor an externalist about moral responsibility. I draw extensively upon the whole body of Strawson’s work, much of which is sadly neglected in discussions of ‘Freedom and Resentment’, although it illuminates many of the issues discussed there.

In a series of papers, Michelle Ciurria (2014a; 2014b; 2015) distinguishes between internalism and externalism about moral responsibility. Internalism is the view that moral responsibility is determined primarily by ‘internal, psychological mechanisms’, such as an agent’s mental states, events or processes (Ciurria, 2014a, p. 546; Ciurria, 2015, pp. 601-602). Externalism is the view that moral responsibility is determined primarily by what is not ‘in the head’ (see the title of Ciurria, 2015), such as an agent’s overt behaviour and circumstances external to the agent. Internalism and externalism lie on a spectrum. Ciurria claims that pure internalism, according to which moral responsibility is determined solely by internal mechanisms, is implausible, and so is pure externalism. The correct account of moral responsibility is to be found somewhere along the spectrum (Ciurria, 2015, pp. 601-602).

According to Ciurria, most if not all current accounts of moral responsibility, including ‘Strawsonian’ ones, are internalist: they regard internal factors as more important for moral responsibility than external ones. Ciurria defends externalism, and she argues that, in contrast to his contemporary followers, P.F. Strawson himself was an externalist about moral responsibility. I believe that her reading of Strawson is problematic.[[1]](#endnote-1) The aim of this paper is to elucidate Strawson’s position with regard to the internalism-externalism issue against the background of Ciurria’s reading of him.

What is, according to Strawson, the role of internal and external factors in ascriptions of moral responsibility? First, I will discuss Strawson’s references to behaviour and circumstances in ‘Freedom and Resentment’ (Strawson, 2008a [1962]) which, Ciurria thinks, can only be taken seriously by an externalist interpretation of Strawson. Secondly, I will consider an argument invoked by Ciurria against the idea that Strawson was an internalist. The argument goes roughly as follows: explanations in terms of internal psychological mechanisms are causal explanations, and causal notions have their home in the objective perspective. Because, according to Strawson, moral responsibility is at home in the participant perspective and not in the objective perspective, Strawson was not an internalist about moral responsibility.

I will conclude that Strawson was neither an internalist nor an externalist about moral responsibility. I will extensively draw upon the whole body of Strawson’s work, much of which is sadly neglected in discussions of ‘Freedom and Resentment’, although it illuminates many of the issues discussed there. I hope that this paper provides a convincing example of how it does so.[[2]](#endnote-2)

1. **Internal and external considerations in ‘Freedom and Resentment’**

*1.1 Excuses and exemptions*

In ‘Freedom and Resentment’, Strawson emphasizes that it matters very much to us whether the actions and attitudes of other people reflect attitudes towards us of good or ill will. It matters very much, for example, whether someone steps on my hand on purpose, which reflects an attitude of ill will, or accidentally, which reflects neither good nor ill will. Roughly, a reactive attitude is a reaction to the ‘good or ill will or indifference or lack of concern’, that is, to the *quality of will* displayed in people’s attitudes or actions (Strawson, 2008a, p. 15).

According to Strawson, to adopt a reactive attitude is to regard an action or attitude as reflecting a certain quality of will, and it is for actions and attitudes reflecting a certain quality of will that we hold people responsible. Some actions and attitudes do notreveal anything about a person’s quality of will: if you step on my hand accidentally, it does not reveal anything about your good or ill will or indifference towards me. If an action does not reflect any quality of will, reactive attitudes are inappropriate: resentment may be fitting when you step on my hand on purpose, but not when you do so accidentally. Because adopting a reactive attitude and holding responsible are closely connected, Strawson thinks that we can learn a lot about the conditions under which ascriptions of responsibility are inappropriate by looking at the conditions under which reactive attitudes are inappropriate.

Under which conditions is it inappropriate to adopt reactive attitudes? Strawson distinguishes two groups:

To the first group belong all those which might give occasion for the employment of such expressions as ‘He didn’t mean to’, ‘He hadn’t realized’, ‘He didn’t know’; and also all those which might give occasion for the use of the phrase ‘He couldn’t help it’, when this is supported by the use of such phrases as ‘He was pushed’, ‘He had to do it’, ‘It was the only way’, ‘They left him no alternative’, etc. (Strawson, 2008a, pp. 7-8)

The pleas of the first group have in common that ‘… they do not invite us to see the *agent* as other than a fully responsible agent. They invite us to see the *injury* as one for which he was not fully, or at all, responsible’ (Strawson, 2008a, p. 8). This is what distinguishes the first from the second group of considerations, the *excuses* from the *exemptions* (as Watson, 1987 puts it). The latter *do* invite us to see the agent as other than a fully responsible agent. The second group can be divided into two subgroups, ‘of which the first is far less important than the second’ (Strawson, 2008a, p. 8).

In connection with the first subgroup we may think of such statements as ‘He wasn’t himself’, ‘He has been under very great strain recently’, ‘He was acting under post-hypnotic suggestion’; in connection with the second, we may think of ‘He’s only a child’, ‘His mind has been systematically perverted’, ‘That’s purely compulsive behaviour on his part’. (Strawson, 2008a, p. 8)

What distinguishes the first from the second subgroup of exemptions is that, while the first subgroup presents the agent as psychologically normal and morally developed, but the circumstances as abnormal, ‘the second and more important subgroup of cases allows that the circumstances were normal, but presents the agent as psychologically abnormal – or as morally undeveloped. The agent was himself; but he is warped, deranged, neurotic or just a child’ (Strawson, 2008a, p. 9). Thus, we get the following picture:

1. Excuses: the agent is a fully responsible agent, but not responsible for a particular injury.
2. Exemptions: the agent is not a fully responsible agent.

(2.1) The agent is psychologically normal and morally developed, but circumstances were abnormal.

(2.2) Circumstances were normal, but the agent is psychologically abnormal or morally undeveloped.

Suppose that we have to argue for or against the idea that Strawson was an externalist about moral responsibility. We then have to measure, somehow, the relative weight of internal and external considerations in (suspensions of) attributions of moral responsibility. Which considerations refer to mental states or inner psychological mechanisms, which ones refer to behaviour or external circumstances? Take, first, the phrases mentioned above. ‘He didn’t mean to’, ‘He hadn’t realized’ and ‘He didn’t know’ might seem internal, because they tell us something about mental states or at least about their absence. ‘He was pushed’, ‘He had to do it’, ‘They left him no alternative’ and ‘That’s purely compulsive behaviour on his part’ might seem external, because they refer to circumstances or behaviour. But what about ‘He couldn’t help it’, ‘He wasn’t himself’, ‘He has been under very great strain recently’, ‘He was acting under post-hypnotic suggestion’, ‘He’s only a child’, ‘He’s a hopeless schizophrenic’ and ‘His mind has been systematically perverted’? At least some of Strawson’s excuses and exemptions cannot easily be characterized as either internal or external.

This could be taken to mean that there is a grey zone of considerations that do not clearly belong to either the internalist or the externalist group. So Strawson mentions some internal considerations, some external ones, and some that cannot readily be classified as either internal or external. The picture gets more complicated, however, when we consider the condition of being peculiarly unfortunate in formative circumstances, on which Ciurria (2014a) puts a lot of weight. This is an important external condition, according to her, and internalist accounts of moral responsibility (such as Watson, 1987 and Wolf, 1987) cannot explain the non-derivative difference that this factor can make to an individual’s moral responsibility-related status (Ciurria, 2014a, p. 547). It is noteworthy (and I will come back to this) that Strawson mentions this condition in relation to the *second* subgroup of exemptions (2.2), where *circumstances are normal* but the agent is psychologically abnormal or morally undeveloped. Strawson often refers to this important subgroup in the rest of his essay, but he does so in different wordings. Within subgroup 2.2, he consistently distinguishes between [1] the ‘psychologically abnormal’ and [2] the ‘morally undeveloped’. Thus, he speaks of agents who are ‘[1] warped or deranged, neurotic or [2] just a child’; ‘[1] warped or deranged or compulsive in behaviour or [2] peculiarly unfortunate in his formative circumstances’; ‘[1] by abnormalities or [2] by immaturity’; ‘[1] the compulsive behaviour of the neurotic or [2] the tiresome behaviour of a very young child’; ‘[1] by deep-rooted psychological abnormality – [2] or simply by being a child’; ‘[1] an idiot, or [2] a moral idiot’ (Strawson, 2008a, pp. 9-13; numbers added).

Here also, we see that Strawson mentions internal and external considerations, at least according to Ciurria’s way of distinguishing between them. For Ciurria, psychological abnormality counts as internal, while compulsive behaviour is external. Being morally undeveloped, or being a moral idiot, is internal (‘the criterion of moral sanity is irreducibly internalistic’ (Ciurria, 2014a, p. 552; see also Ciurria, 2014b, p. 272), while being unfortunate in formative circumstances is external. What we see here is not just that some considerations are internal while others are external, but also, and more importantly, that *Strawson describes what he takes to be the same consideration sometimes in internal and sometimes in external terms*. Thus, psychological abnormality (internal) is sometimes described in terms of behaviour (external), and being morally undeveloped (internal) is described in terms of tiresome behaviour and unfortunate circumstances (external).

What does this show? It could be taken to show that Strawson has not been careful: what exactly is excusing, the fact that an agent did not mean to act as she did *or* the fact that she was pushed? What is exempting, formative circumstances *or* being morally undeveloped, psychological abnormality *or* compulsive behaviour? Charity demands, however, that we ask why Strawson presents things as he does. In order to answer that question, we have to turn to Strawson’s philosophy of psychology.

*1.2 Strawson’s Wittgensteinian philosophy of psychology*

It is uncontroversial that Strawson’s philosophy of psychology was heavily influenced by Wittgenstein’s (Reinhardt, 1969), and that Wittgenstein rejects the separation between ‘inner’ mental states and ‘outer’ behaviour. According to Wittgenstein, ‘the inner is tied up with the outer not only empirically, but also logically’ (Wittgenstein, 1992, p. 63). William Child explains:

… Wittgenstein takes it to be essential to pain, for instance, that in normal cases it has the kind of outer behavioural expression it does. He recognizes, of course, that there can be instances of pain without any behavioural manifestation. But such cases, he thinks, are necessarily the exception, not the rule. It is no more possible for there to be a world in which people experience pain but never manifest it in their behaviour, or a world in which people have thoughts but never express their thoughts, than it is for there to be a world in which people play games but no one makes anything but false moves in the game. (For this analogy, see Wittgenstein, 2009, §§344-345.) (Child, 2017, p. 474)

That the inner and the outer are logically tied up means that the outer is more than just a *sign* or *symptom* or *evidence* of the inner.[[3]](#endnote-3) In normal cases, the mental *infuses* our behaviour and is *expressed* or *manifested* in it (Wittgenstein, 2009, §579), and our behavioural expressions or manifestations are *criteria* for the ascription of mental states to us (Wittgenstein, 2009, §580). Mental states are neither inner nor outer, they are not in our heads or outside of our heads (to locate them in space is to make a category mistake), but they are manifested in ways that are publicly observable.

There is every reason to believe that Strawson accepted these Wittgensteinian points. He explicitly rejects the inner-outer picture in his review of Wittgenstein’s *Philosophical Investigations* (Strawson, 2008c [1958], p. 183)*.* He writes the following in *Individuals*:

What I have said is that one ascribes P-predicates [all predicates we apply to persons but not to material bodies; these include predicates ascribing mental states] to others on the strength of observation of their behaviour; and that the behaviour-criteria we go on are not just signs of the presence of what is meant by the P-predicate, but are criteria of a logically adequate kind for the ascription of the P-predicate. (Strawson, 1959, p. 106)

This point, as well as Strawson’s use of ‘criteria’ (see also Strawson, 2008d [1985], p. 63) and ‘of a logically adequate kind’, reveals Wittgenstein’s influence.[[4]](#endnote-4)

If Strawson thinks about the relation between mental states and behaviour in a Wittgensteinian way, his way of putting things in ‘Freedom and Resentment’ makes sense. He refers to mental states in some of his formulations of exempting conditions (2.2), but these should not be thought of as inner. His references to behaviour in other formulations of the same consideration do not mean that he has been careless; they should rather be understood as references to the behavioural manifestations or expressions of mental states. If the link between mental states and their behavioural manifestations is logical, there is no ‘logical gap’ or ‘logical wedge’ (Strawson, 1959, pp. 108-109) between them, which is why it is normally quite pointless to ask whether either the mental state or the behaviour excuses or exempts one from being morally responsible.[[5]](#endnote-5) From a Strawsonian perspective, the right answer is that behaviour normally excuses or exempts insofar as it manifests a mental state with which it is logically tied up, and mental states normally excuse or exempt insofar asthey are somehow manifested inbehaviour. Normally, as a rule, mental state and behaviour are tied up, and this helps to explain why Strawson uses behavioural descriptions and descriptions in terms of mental states interchangeably. In abnormal cases, mental state and behaviour can come apart, but I see no reason to think that Strawson was occupied with such cases in ‘Freedom and Resentment’.

If Strawson’s view is that mental states and behaviour are normally tied up, then Strawson is neither an internalist nor an externalist. If his excusing and exempting considerations can be divided into ‘internal’ and ‘external’ at all, the descriptions in terms of ‘internal’ mental states (see, for example, ‘warped or deranged, neurotic’ as a way of referring to psychological abnormality) and those in terms of behaviour (see, for example, ‘warped or deranged or compulsive in behaviour’ as another way to refer to psychological abnormality) refer to two sides of the same coin.

So far, I have discussed the relation between mental states and behaviour. But behaviour is not the only kind of ‘external’ factor that Ciurria mentions. Recall that she puts a lot of weight on external circumstances, especially on Strawson’s ‘peculiarly unfortunate formative circumstances’. Recall that Strawson mentions this consideration in relation to the *second* subgroup of exemptions (2.2), in which circumstances are normal but the *agent* is abnormal or undeveloped. How can we understand this? The answer draws upon the point outlined above about the logical relation between mental states and behaviour. Of course, the circumstances cannot be said to manifest or express mental states. But ‘peculiarly unfortunate formative circumstances’ can be *logically* instead of merely evidentially tied to being morally undeveloped or immature. *Peculiarly* unfortunate circumstances, then, are *essentially* circumstances that make it the case that someone is morally undeveloped or immature. So if we know that formative circumstances were unfortunate, but have no reason to say that an agent is morally undeveloped because of them, these formative circumstances will not count as *peculiarly* unfortunate in Strawson’s sense, and they will not exempt from moral responsibility.[[6]](#endnote-6)

Ciurria rejects ‘internalist’ Strawsonian views, because they have difficulties to account for the fact that external factors sometimes *directly* or *non-derivatively* exempt. Gary Watson writes, for example, that historical factors can be ‘at most, evidence that some other plea is satisfied’ (Watson, 1987, p. 274). According to Ciurria, Wolf’s (1987) view suggests that ‘moral insanity is the inherently relevant factor, while external circumstances are merely evidential’ (Ciurria, 2014a, p. 550; see also Ciurria, 2014b, p. 265, p. 272). It may well be the case that some Strawsonians do not adequately account for the exempting force of external factors. The relevant point is that, even if they are internalist in the sense of thinking that moral responsibility is determined primarily by mental states, they *could* maintain (and should, if they take Strawson seriously) that external factors directly or non-derivatively (in Ciurria’s sense of ‘not merely evidentially’) exempt: the relation between internal and external factors is not merely evidential (which would warrant the idea of indirect or derivative exemption), but logical. This speaks against Ciurria’s (2014a, p. 545; 2014b, p. 265) claim that her interpretation is the *only* interpretation capable of according intrinsic relevance to behaviour and external circumstances in attributions of moral responsibility. Ciurria writes that, ‘without some story for how [Wolf’s view], as an internalistic theory, can be reconciled with inherently externalistic excuses – a story that I myself cannot envision, and which may not be possible – [Wolf’s view] must be seen as insufficient’ (Ciurria, 2014b, p. 267). I believe that Strawson’s Wittgensteinian philosophy of psychology provides that story: it shows that talk of mental states as internal and behaviour as external is misleading, and that mental states, behaviour and circumstances can be logically related to each other.

*1.3 Quality of will*

The above reading of Strawson on the relation between mental states, behaviour and circumstances, is confirmed by several passages in ‘Freedom and Resentment’ and elsewhere in which Strawson discusses quality of will. As we have seen, Strawson argues that whether an agent is morally responsible depends on whether their attitudes reflect a certain quality of will (for development and defense of Strawson’s claim, see, for example, McKenna, 2012). Ciurria mentions this, but she rejects the idea for being too internalist:

It has been suggested to me that an internalist account which emphasizes the quality of an agent’s will is preferable to externalism because quality of will seems essential to Strawson’s account, insofar as Strawson thinks that the reactive attitudes must be deployed toward an agent’s good or ill will as opposed to his overt behaviour as such. (Ciurria, 2014a, p. 554)

Compare this to Strawson:

What I have called the participant reactive attitudes are essentially natural human reactions to the good or ill will or indifference of others towards us, as displayed in *their* attitudes and actions. (Strawson, 2008a, pp. 10-11)

The reactive attitudes I have so far discussed are essentially reactions to the quality of others’ wills towards us, as manifested in their behaviour … (Strawson, 2008a, p. 15)

First, it is an understatement to say, as Ciurria does, that quality of will *seems* essential to Strawson’s account; it *is* essential (and is quite explicitly said to be so, at least if one takes reactive attitudes to be essential, but that seems uncontroversial). Secondly, ‘internal’ good or ill will are not *opposed* to ‘external’ overt behaviour. Rather, the logical linkbetween quality of will and behaviour is spelled out in exactly the terms one would expect (quality of will is ‘displayed in … attitudes and actions’, ‘manifested in behaviour’) in the light of my remarks on Strawson’s philosophy of psychology. See also this passage: ‘There [in ‘Freedom and Resentment’] I spoke of our natural responses … to certain *states of mind we discern as manifested by agents in their actions* – notably their good or ill will or indifference towards ourselves or others’ (Strawson, 1998c, p. 261, my emphasis).

Questions arise, at this point, about the relation between quality of will, mental states, behaviour and external circumstances. The answers will be similar to earlier proposals: behaviour (for example, ‘deranged behaviour’) normally exempts insofar as it manifests a certain mental state (for example, ‘psychological abnormality’), and mental states normally exempt insofar as they preclude someone’s actions from manifesting either good or ill will. Quality of will, mental states, behaviour and external circumstances are inextricably interwoven. Because there is no logical gap between them, quality of will-approaches do not have to be described as internalist in Ciurria’s sense, and they need not be rejected even if one puts a lot of weight on the non-derivatively exempting force of external circumstances. Proponents of a quality of will-view need not be either internalists or externalists. Strawson’s statements about quality of will confirm that he himself was neither.

* 1. *The epistemological problem*

Strawson’s view, as I have explained it, avoids a problem that Ciurria attributes to ‘internalist’ Strawsonian accounts. According to internalists, moral responsibility is determined primarily by mental states. But others’ mental states are hidden from us, they are ‘extremely obscure’ (Ciurria, 2015, p. 613), we can never really know them, we can only ‘conjecture’ (Ciurria, 2015, p. 612). Mental states are only privately accessible and subjective, while behaviour is interpersonally or publicly accessible and objective (Ciurria, 2014a, p. 555). If that is the case, internalists have an epistemological problem that externalists do not have: if we can never really know others’ mental states, and moral responsibility is determined primarily by mental states, it seems as if, on an internalist view, we can never really know whether people are morally responsible.

Strawson’s Wittgensteinian view does not have this problem. If mental states are manifested in behaviour, we can often see what mental states other people are in: ‘Well, one might say this: if one sees the behaviour of a living being, one sees its mind’ (Wittgenstein, 2009, §371); ‘But of course it isn’t true that we are never certain about the mental processes in someone else. In countless cases we are’ (Wittgenstein, 1992, p. 94). Glock summarizes Wittgenstein’s position as follows:

… others can, and often do, know that I am in pain. It is also misleading to claim that such knowledge is ‘indirect’: the sufferer does not know directly or indirectly of his being in pain, and for us there is no more direct way of knowing than by seeing him moan and writhe. In such cases we do not *infer* – draw the conclusion – that he is in pain, we *see* that he is suffering. Nevertheless, one might hold, I cannot see the pain itself, only the behaviour which expresses it. But this is like saying that I cannot see sounds or hear colours. It indicates only a categorial distinction between mental and behavioural terms …

It is tempting to protest that the mind is hidden in that there is always the possibility that others are lying or pretending. This shows that our third-person judgments are fallible. It does not establish the skeptical conclusion that, in a particular case, we are or could always be mistaken. … Nor is pretence possible in all cases, for example when someone falls into a fire and screams with pain. (Glock, 1996, pp. 176-177)

Again, it is fairly obvious that Strawson, in many ways an anti-skeptic (see Strawson, 2008d), accepts these Wittgensteinian points. In his review of Wittgenstein’s *Philosophical Investigations,* he confirms that it is ‘empty and pointless … to raise generalized doubts about other people’s experience of pain, or about one’s knowledge of them’ (Strawson, 2008c, p. 171). In ‘Classical Empiricism. The Inner and the Outer’, a chapter in *Analysis and Metaphysics,* Strawson distinguishes between what he calls ‘internalism’ and ‘externalism’, and his characterization of these views comes close to what Ciurria understands by them.

Internalism treats the inner subjective life of thoughts, sensations, and inner experience in general as a series of unproblematic private entities – and regards the physical world as problematic. Externalism treats the physical public world of bodies moving and interacting in space as unproblematic and the subjective and inner life as problematic. …

I shall not attempt a full diagnosis of the condition in which externalism is felt to be attractive. (It is perhaps felt to be the hard-headed or scientific approach.) … One at least of the reasons for the attraction of externalism is certainly the thought that the characteristics, the relation, and the behaviour of bodies, including human bodies, in space are, or seem to be, satisfactorily definite and observable; whereas the mental or inner life seems to be characteristically elusive and indefinite, not available to public inspection or scientific verification. (Strawson, 1992, pp. 74-75)

Strawson rejects both internalism and externalism, because they make a similar error: both ignore that the ‘internal’ and the ‘external’ are two sides of the same coin, the one not more or less problematic than the other. We often know very well what mental state others are in, so that these states cannot be said to be in general or in principle inaccessible to us. In some cases, we can be mistaken or deceived, while in others we cannot, and this holds for others’ mental states as well as for their behaviour. Sometimes it is easier to identify a person’s state of mind than their behaviour, sometimes it is the other way round (see Kenny, 1978, pp. 11-21 for examples). The possibility of deception in some cases should not lead us to think that deception is always possible. There is no epistemological problem for Strawson; not because he was an externalist, but because mental states are normally manifested in behaviour.[[7]](#endnote-7)

**2. The humanistic and the scientific perspective**

*2.1 Causation and its relation to objective and participant perspectives*

I have tried to show that Strawson’s references to external circumstances and behaviour need not be taken to indicate that he was an externalist about moral responsibility. Another reason for thinking that Strawson was an externalist is provided by the following argument:

1. Internalism explains moral responsibility in terms of inner, psychological mechanisms (see the definition of internalism in the introduction).
2. Explanations in terms of inner, psychological mechanisms are a specific variety of causal explanation.
3. Causal notions have their natural home in the domain of determinist discourse, that is, in the objective perspective.
4. [From (1), (2) and (3)] Internalism explains moral responsibility in terms at home in the objective perspective.
5. According to Strawson, moral responsibility has to be explained in terms at home in the participant perspective, not in the objective perspective.
6. [From (4) and (5)] Strawson’s account of moral responsibility cannot be internalist.

A version of this argument was first made by Sneddon (2005, p. 245) and has been approvingly cited by Ciurria, who stresses that ‘causal explanations [are] *ipso facto* … external to the participant perspective’ (Ciurria, 2014a, p. 551; see also Ciurria, 2014b, p. 273).

It is precisely this point, that is, premise (3) of the argument, that Strawson would reject. The participant perspective is, roughly speaking, the perspective that we adopt when we adopt reactive attitudes, while the objective perspective is the perspective that we adopt when we suspend our reactive attitudes. In his characterization of the objective perspective in ‘Freedom and Resentment’ (Strawson, 2008a, pp. 9-10), Strawson makes no connection between the objective perspective and causal notions. Moreover, he seems to reject the association. In ‘Causation and Explanation’, a chapter in *Analysis and Metaphysics,* he explains that the general notion of a cause belongs to our basic stock of concepts. It is most naturally at home in our ordinary, everyday talk (Strawson, 1992, pp. 115-116). Of course, we also use the notion of a cause in scientific theories, but ‘there is a great gap between our ordinary causal explanations of particular events and circumstances and the notion of explicit appeal to strict law’ (Strawson, 1992, p. 126).[[8]](#endnote-8)

Three remarks are in order here. First, we have shifted from the distinction between a participant and an objective perspective to the distinction between an ordinary, everyday and a scientific perspective. While there may be subtle differences between the two contrasts, I do not think that they are relevant here. Sneddon and Ciurria clearly associate the participant perspective with an ordinary perspective and the objective perspective with a scientific perspective. And so does Strawson: at several places in his work, he explicitly distinguishes between an everyday personal or ‘humanistic’ and a specialized or ‘scientific’ perspective (Strawson, 1998a, p. 87; Strawson, 2008d, p. 46), and I see no reason to think that he saw this distinction as importantly different from his distinction between a participant and an objective perspective.

Secondly, it does not seem controversial to hold that we use causal explanations in our ordinary, everyday talk. Donald Davidson fully agrees with Strawson on this point:

It is often thought that scientific explanations are causal, while explanations of actions and mental affairs are not. I think almost exactly the reverse is the case: ordinary explanations of action, perception, memory, and reasoning, as well as the attribution of thoughts, intentions, and desires, is riddled with causal concepts ... (Davidson, 2004, p. 96)

We explain why someone ate by pointing out that they were hungry (Davidson, 2004, p. 96). Rather than a scientific explanation, this is an everyday one. I propose to accept the point made by Strawson and Davidson that there is room for causal explanation within ordinary discourse.

Thirdly, this should not lead us to think that non-ordinary, scientific causal explanations must be explanations in terms of strict law. While it is clear that, according to Sneddon and Ciurria, explanations in terms of inner psychological mechanisms are causal explanations belonging to scientific psychology and not to ordinary discourse (Sneddon, 2005, p. 246; Ciurria, 2014b, p. 273), the exact nature of causal explanations in scientific psychology can be left open. Sneddon and Ciurria do not specify whether they are thinking of explanations in terms of strict laws, or *ceteris paribus* laws, or even whether the causal explanations they have in mind appeal to laws at all, and there is a lot of controversy about the kind and scope of causal explanations in the special sciences (for a good overview, see Kincaid, 2009). The distinction that matters for our purposes is the distinction between ordinary kinds of causal explanation and *non-ordinary, scientific kinds of causal explanation.*

In light of the three remarks above, the argument made by Ciurria can now be put in the following form (which, as far as I can see, she would accept):

1. Internalism explains moral responsibility in terms of inner, psychological mechanisms (see the definition of internalism in the introduction).
2. Explanations in terms of inner, psychological mechanisms are non-ordinary, scientific causal explanations.
3. Scientific causal explanations have their natural home in scientific discourse.
4. [From (1), (2) and (3)] Internalism explains moral responsibility in terms at home in scientific discourse.
5. According to Strawson, moral responsibility has to be explained in terms at home in ordinary discourse, not in scientific discourse.
6. [From (4) and (5)] Strawson’s account of moral responsibility cannot be internalist.

Premise (3) can now be accepted. Because (1) is true by definition and (5) is also true, and (4) and (6) follow from these premises, (2) is the premise to be investigated. In comparison to its counterpart in the first argument, (2) has now been made more specific: the notion of cause employed in it is a scientific notion, not an everyday notion. This specification is needed: if an everyday notion would be meant, internalism would explain moral responsibility in terms of an everyday notion, which is perfectly compatible with (5), and the conclusion would not follow. Our question, then, is whether explanations in terms of inner, psychological mechanisms or mental states are non-ordinary, scientific causal explanations, or at least explanations at home in scientific discourse.

They are not always. Sneddon mentions that we could explain mental states in folk psychological terms, and folk psychology belongs to the ordinary and not to the scientific perspective. But this, he says, raises a dilemma for Strawson about the relationship between folk psychology and scientific psychology (Sneddon, 2005, p. 246). According to Sneddon, there are two possibilities. First, the relationship between folk psychology and scientific psychology is ‘very close’, ‘folk psychology gets a principled interpretation *within* the [scientific] perspective’ and can be ‘incorporated’ by it. Deep down, folk psychological explanation *is* scientific explanation, ultimately at home in scientific discourse. If that is the case, the possibility of folk psychological explanation poses no threat to the argument above.

The second possibility is that the relationship between folk psychology and scientific psychology is not close. But this way of securing the independence of folk psychology ‘robs folk psychology of its authority’ (Sneddon, 2005, p. 246). Explanations in terms of mental states can be folk psychological or scientific, but the *best* ones will be in terms of scientific psychology. This preserves a version of our argument in which premise (2) starts with ‘The *best* explanations …’, and (4) becomes that, according to internalists, the best explanations of moral responsibility are ones in terms at home in scientific theories. If (5) is reformulated as ‘According to Strawson, moral responsibility is *best* explained in terms at home in ordinary discourse’ (which seems plausible), the conclusion follows.

*2.2 Behaviour and its relation to humanistic and scientific perspectives*

It may now seem as if an externalist conclusion cannot be avoided. But that would be too quick. In ‘The Mental and the Physical’, a chapter in *Scepticism and Naturalism,* Strawson writes:

… it seems increasingly reasonable … that there is a system of physical laws such that all bodily movements of human beings … are the causal outcome of the stimulation of sensory surfaces together with the internal physical constitution of the organism … We have another and more familiar style of talking about ourselves and others, in which we speak of action and behaviour (in the ordinary sense of this latter word), rather than simply of limbs moving, and in which we freely use the language of sensations, perceptions, thoughts, memories, assertions, beliefs, desires, and intentions; in short, mentalistic or personalistic language. … we have the theoretical idea of the two histories, each complete in its own terms; we might call them the physical history and the personal history, admitting, of course, that the latter, the personal history, will include accounts of physical action, reaction and displacement as well as mental events and states. Each story will invoke its own explanatory connections, the one in terms of neurophysiological and anatomical laws, the other in terms of what is sometimes called, with apparently pejorative intent, ‘folk psychology’; i.e. the ordinary explanatory terms employed by diarists, novelists, biographers, journalists, and gossips, when they deliver their accounts of human behaviour and human experience – the terms employed by such simple folk as Shakespeare, Tolstoy, Proust and Henry James. (Strawson, 2008d, p. 46)

Strawson is dismissive of the term ‘folk psychology’. So-called ‘folk psychological’ explanations of mental states and behaviour are neither forms of scientific explanation, nor are they in any way inferior to scientific explanation. The relation between an everyday personal and a specialized scientific perspective is not a relation in which they basically come to the same thing (with the scientific perspective being the more principled one), or a relation in which the one is superior to the other. Sneddon’s two possibilities are the result of the assumption that ordinary and scientific explanations are basically in the same business. But they are not: according to Strawson, they are not ‘*in pari materia*’ (Strawson, 2008d, p. 51); they explain *different things in different ways.*

Let me explain. First, the different ways are mentioned in the quotation. Within the scientific perspective, things are explained in terms of ‘neurophysiological and anatomical laws’. Within the humanistic perspective, by contrast, things are explained in terms of ‘sensations, perceptions, thoughts, memories, assertions, beliefs, desires, and intentions; in short, mentalistic or personalistic language’. It is within this perspective that ‘accounts of [someone’s] mental life as a subject of conscious states’ (Strawson, 1998b, p. 169) are given. This speaks against the second premise of Ciurria’s argument, namely, that explanations in terms of mental states are non-ordinary, scientific causal explanations. Secondly, *what* is explained within the different perspectives is not the same thing (see Strawson, 2008d, p. 54). Within the humanistic perspective, we explain ‘action and behaviour’; within the scientific perspective, we explain ‘bodily movements’. Thus, according to Strawson, we have a humanistic perspective in which action and behaviour are explained in terms of (among other things) mental states, and a scientific perspective in which movements are explained in terms of neurophysiological and anatomical laws. This is confirmed in other work:

… on the one hand, the actual normal practice of explaining human *behaviour* in terms of the agent’s desires, perceptions, beliefs, and decisions and, on the other hand, the theoretical possibility of a causal explanation of gross *bodily movement* in purely physical and physiological terms. Of course these are, in a sense, complementary, since intentional action involves bodily movement. But they are in no sense competitors. (Strawson, 1998a, p. 87)

The difference between explanations of behaviour and explanations of movement is important for Strawson. (Again, this is a Wittgensteinian point. See Glock, 1996, p. 175.) Tracing physical causal routes through physical organisms cannot yield a causal explanation of human *action* or human *behaviour*: ‘Any explanatory account of a person’s behaviour (a much misused word) belongs firmly to the personal story, the biographer’s or diarist’s story, and mentions, as causally explanatory factors, … mental events and states (desires, perceptions, beliefs, decisions)’ (Strawson, 2008d, p. 52).[[9]](#endnote-9)

I would like to distinguish two ways of reading these passages from Strawson. First, there is a strong reading of Strawson, according to which he rejects premise (2) of Ciurria’s argument: explanations in terms of mental states are not scientific causal explanations. Strawson’s point, as outlined above, could be made in an even stronger way: explanations in terms of mental states *cannot be* scientific causal explanations. He denies a weak version of premise (2), according to which explanations in terms of mental states *can be* scientific causal explanations.

Given my aim to elucidate Strawson’s position, Strawson’s denial of (2) is significant. It could be objected, however, that Strawson’s rejection of premise (2) is not enough: in order to show that the argument is unsound, I have to establish that (2) is false, not just that Strawson rejects it. And why think that (2), especially its weaker version, is false? Why think that explanations in terms of mental states cannot be provided from a scientific perspective? Is this not straightforwardly contradicted in, e.g., scientific social psychology and cognitive neuroscience? There are tons of scientific theories that are formulated in mentalistic terms.

I do not think that Strawson is committed to deny this. His characterization of the scientific perspective is not descriptive, that is, he does not attempt to describe what scientists are actually doing (or prescribe what they should be doing). Rather, his characterization is stipulative: insofar as one is seeking explanations of mere bodily movement in purely physical and physiological terms, one is approaching things from the scientific perspective. If that is Strawson’s point, it is compatible with the claim that scientific social psychology and cognitive neuroscience cannot be located within the scientific perspective as he characterizes it. Maybe they, or parts of them, can be located somewhere in between these perspectives, or on the border between them, or within the humanistic perspective. What Strawson denies when he denies that explanations in terms of mental states can be provided from the scientific perspective, is that mental states can be fully explained in neurophysiological and/or anatomical terms. Strawson’s scientific perspective (a term which may be ill-chosen) is much narrower than the perspective that scientists actually adopt, which means that his denial of (2) is less problematic than could be supposed.

According to a strong reading of Strawson, (2) can be denied and Ciurria’s argument will be unsound. But some will now feel that (2) is denied almost by conceptual fiat: if we understand ‘scientific causal explanations’ in the narrow way in which Strawson understands it, as more or less equivalent to ‘causal explanations in terms of neurophysiological and/or anatomical terms’, (2) might be thought to become almost trivially false. It could be argued, in defense of Strawson, that the narrow interpretation of ‘scientific causal explanations’ seems quite close to Ciurria’s own understanding of these explanations. Although she does not provide a lot of information about the form that explanations in terms of inner psychological mechanisms take, she suggests that we should think of them as explanations in ‘neurobiological’, ‘neurophysiological’ and ‘neurological’ terms (Ciurria, 2014a, p. 555; Ciurria, 2014b, pp. 273-274; Ciurria, 2015, p. 604).

Let us assume, however, that a broader interpretation of ‘scientific causal explanations’ is available, so that explanations in terms of mental states can be scientific causal explanations. Such a broader interpretation seems plausible, and if we interpret (2) in this way, Ciurria’s argument still stands.

A weaker reading of Strawson is helpful here. The point of the weaker reading is not that explanations in terms of mental states cannot belong to the scientific perspective (if the scientific perspective is the perspective that scientists in social psychology and cognitive neuroscience are actually adopting, this point seems false), but that mental states, on the one hand, and behaviour and external circumstances, on the other hand, are *in the same boat* with regard to the possibility of scientific causal explanations. Consider the following argument:

(1\*) Externalism explains moral responsibility in terms of external factors such as behaviour and external circumstances.

(2\*) Explanations in terms of behaviour and external circumstances are non-ordinary, scientific causal explanations.

(3) Scientific causal explanations have their natural home in scientific discourse.

(4\*) [From (1\*), (2\*) and (3)] Externalism explains moral responsibility in terms at home in scientific discourse.

(5) According to Strawson, moral responsibility has to be explained in terms at home in ordinary discourse, not in scientific discourse.

(6\*) [From (4) and (5)] Strawson’s account of moral responsibility cannot be externalist.

This argument against Strawson’s externalism mirrors the structure of Ciurria’s argument against internalism. In parallel with the strong reading of Strawson, premise (2\*) has to be rejected, but what has been said about the rejection of (2) applies equally to (2\*): it seems plausible to think that there is a broad sense of ‘scientific causal explanations’ which allows for the idea that explanations in terms of behaviour and external circumstances can be scientific causal explanations. A question suggested by the weak reading of Strawson can be put as follows: *Why should we treat mental states and behaviour differently? If we reject (2), then why not (2\*)? If we accept (2), then why not (2\*)?* Externalists and internalists have to explain why they accept only Ciurria’s argument and not the mirror argument or *vice versa*. If explanations in terms of mental states can be scientific causal explanations, then why could explanations in terms of behaviour and external circumstances not be scientific causal explanations?

So even if the strong reading of Strawson, including his rejection of premise (2), is not accepted (because one believes that explanations in terms of mental states can be scientific causal explanations), the passages from Strawson contain an important lesson: the argument against internalism will only be successful if the argument against externalism fails. Because of the structural similarity between the two arguments, the only way to establish that the argument against externalism is unsound without making the argument against internalism unsound, is to show that (2) is true and (2\*) is false, that is, to show that explanations in terms of behaviour and external circumstances are not (or, if the argument against internalism uses the weaker version of (2), *cannot* be) scientific causal explanations. If (2) and (2\*) are true and *both* arguments are sound, the conclusion must be that Strawson’s account can be neither internalist nor externalist.

Thus, as long as mental states and behaviour are in the same boat (whether it be the humanistic or the scientific one), as long as no wedge is driven in between them, both arguments will either be unsound (which is suggested by the strong reading of Strawson, according to which (2) and (2\*) are false) or both will be sound (which is made possible by the weak reading of Strawson, allowing for the possibility that (2) and (2\*) are true). If they want to appeal to a version of the arguments above, internalists andexternalists need an account of how and why mental states, behaviour and circumstances are not in the same boat with respect to scientific causal explanations. They have not provided such an account, and I find it difficult to imagine what it might look like. As long as it has not been provided, the arguments give us no reason to think that Strawson was either an externalist or an internalist about moral responsibility.

**3. Conclusion**

Strawson rejects the distinction between internalism and externalism about moral responsibility, as made by Ciurria: mental states are not ‘internal’ and behaviour is not ‘external’. They are much more closely linked: mental states infuse behaviour, behaviour manifests or expresses mental states. While Strawson frequently mentions mental states and behaviour in ‘Freedom and Resentment’, external circumstances play a relatively marginal role. They can be said to excuse or exempt insofar as they are logically connected to certain mental states and certain forms of behaviour. The fact that Strawson refers to peculiarly unfortunate formative circumstances does not indicate that he was an externalist. Neither does an argument against Strawson’s internalism provided by Ciurria, mainly because it fails to explain why a structurally similar argument against Strawson’s externalism would be unsound. I conclude that Strawson was not an externalist about moral responsibility, and not an internalist either.[[10]](#endnote-10)

**References**

Anscombe, G.E.M. (1975). Causality and determination. In E. Sosa (Ed.), *Causation and conditionals* (pp. 63-81).Oxford: Oxford University Press.

Avramides, A. (2001). *Other minds.* London: Routledge.

Avramides, A. (2019). Other minds. In E. Zalta (Ed.), *The Stanford encyclopedia of philosophy,* Summer 2019 edition.

Beebee, H., Hitchcock, C., & Menzies, P. (Eds.). (2009). *The Oxford handbook of causation.* Oxford: Oxford University Press.

Campbell, J. (2017). P.F. Strawson’s free will naturalism. *International Journal for the Study of Skepticism*,7, 26-52.

Cassam, Q. (2007). *The possibility of knowledge.* Oxford: Oxford University Press.

Child, W. (2017). The inner and the outer. In H.-J. Glock, & J. Hyman (Eds.). *A companion to Wittgenstein* (pp. 465-477)*.* Chichester: Wiley-Blackwell.

Ciurria, M. (2014a). Moral responsibility. Justifying Strawson and the excuse of peculiarly unfortunate formative circumstances. *Ethical Theory and Moral Practice*,17, 545-557.

Ciurria, M. (2014b). The case of Jojo and our pretheoretical intuitions. An externalist interpretation. *Review of Philosophy and Psychology,* 5, 265-276.

Ciurria, M. (2015). Moral responsibility ain’t just in the head. *Journal of the American Philosophical Association,* 1, 601-616.

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

Davidson, D. (1985). Replies to essays I-IX. In B. Vermazen, & M. B. Hintikka (Eds.). *Essays on Davidson. Actions and events* (pp. 195-229)*.* Oxford: Clarendon.

Davidson, D. (2004 [1990]). Representation and interpretation. In D. Davidson. *Problems of rationality* (pp. 87-99)*.* Oxford: Clarendon.

De Mesel, B. (2017). Is moral responsibility essentially interpersonal? A reply to Zimmerman. *The Journal of Ethics,* 21, 309-333.

De Mesel, B. (2018a). Moral responsibility and the moral community. Another reply to Zimmerman. *The Journal of Ethics,* 22, 77-92.

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

De Mesel, B., & Heyndels, S. (2019). The facts and practices of moral responsibility. *Pacific Philosophical Quarterly,* 100, 790-811.

Ellis, J., & Guevara, D. (2012). *Wittgenstein and the philosophy of mind.* Oxford: Oxford University Press.

Feigl, H. (1958). Other minds and the egocentric predicament. *Journal of Philosophy,* 55, 978-987.

Glock, H.-J. (1996). *A Wittgenstein dictionary.* Oxford: Blackwell.

Gomes, A. (2011). Is there a problem of other minds? *Proceedings of the Aristotelian Society,* 111, 353-373.

Hacker, P.M.S. (2010). *Human nature. The categorial framework.* Oxford: Wiley-Blackwell.

Hacker, P.M.S. (2013). *The intellectual powers. A study of human nature.* Oxford: Wiley-Blackwell.

Hacker, P.M.S. (2019). *Wittgenstein. Meaning and mind.* Oxford: Wiley-Blackwell.

Heyndels, S. (2019). Strawson’s method in ‘Freedom and Resentment’. *The Journal of Ethics,* 23, 407-423.

Kenny, A. (1978). *Freewill and responsibility.* London: Routledge & Kegan Paul.

Kincaid, H. (2009). Causation in the social sciences. In H. Beebee, C. Hitchcock, & P. Menzies (Eds.), *The Oxford handbook of causation* (pp. 726-743)*.* Oxford: Oxford University Press.

Lau, J., & Deutsch, M. (2016). Externalism about mental content. In E. Zalta (Ed.), *The Stanford encyclopedia of philosophy,* Winter 2016 edition.

McDowell, J. (1998a [1978]). On ‘the reality of the past’. In J. McDowell, *Meaning, knowledge, and reality* (pp. 295-313)*.* Cambridge, MA: Harvard University Press.

McDowell, J. (1998b [1982]). Criteria, defeasibility, and knowledge. In J. McDowell. *Meaning, knowledge, and reality* (pp. 369-394)*.* Cambridge, MA: Harvard University Press.

McKenna, M. (2012). *Conversation and responsibility.* Oxford: Oxford University Press.

Overgaard, S. (2007). *Wittgenstein and other minds. Rethinking subjectivity and intersubjectivity with Wittgenstein, Levinas, and Husserl.* New York: Routledge.

Overgaard, S. & Zahavi, D. (2009). Understanding (other) minds. Wittgenstein’s phenomenological contribution. In E. Zamuner, & D.K. Levy (Eds.). *Wittgenstein’s enduring arguments* (pp. 60-86)*.* London: Routledge.

Moyal-Sharrock, D. (Ed.) (2007). *Perspicuous presentations. Essays on Wittgenstein’s philosophy of psychology.* Basingstoke: Palgrave Macmillan.

Reinhardt, L. R. (1969). Wittgenstein and Strawson on other minds. In P. Winch (Ed.), *Studies in the philosophy of Wittgenstein* (pp. 152-165)*.* London: Routledge & Kegan Paul.

Roessler, J. (2011). Perceptual causality, counterfactuals, and special causal concepts. In C. Hoerl, T. McCormack, & S. R. Beck (Eds.), *Understanding counterfactuals. Understanding causation. Issues in philosophy and psychology* (pp. 75-89)*.* Oxford: Oxford University Press.

Sneddon, A. (2005). Moral responsibility. The difference of Strawson, and the difference it should make. *Ethical Theory and Moral Practice,* 8, 239-264.

Schaffer, J. (2016). The metaphysics of causation. In E. Zalta (Ed.), *The Stanford encyclopedia of philosophy,* Fall 2016 edition.

Schroeder, S. (Ed.) (2001). *Wittgenstein and contemporary philosophy of mind.* Basingstoke: Palgrave.

Strawson, P. F. (1959). *Individuals. An essay in descriptive metaphysics.* London: Routledge.

Strawson, P. F. (1985). Causation and explanation. In B. Vermazen, & M. B. Hintikka (Eds.), *Essays on Davidson. Actions and events* (pp. 115-135)*.* Oxford: Clarendon.

Strawson, P. F. (1992). *Analysis and metaphysics. An introduction to philosophy.* Oxford: Oxford University Press.

Strawson, P. F. (1998a). Reply to E. M. Adams. In L. E. Hahn (Ed.), *The philosophy of P. F. Strawson* (pp. 86-90)*.* LaSalle: Open Court.

Strawson, P. F. (1998b). Reply to Simon Blackburn. In L. E. Hahn (Ed.), *The philosophy of P. F. Strawson* (pp. 168-174)*.* LaSalle: Open Court.

Strawson, P. F. (1998c). Reply to David Pears. In L. E. Hahn (Ed.), *The philosophy of P. F. Strawson* (pp. 256-262)*.* LaSalle: Open Court.

Strawson, P. F. (2008a [1962]). Freedom and Resentment. In P. F. Strawson, *Freedom and resentment and other essays* (pp. 1-28)*.* London: Routledge.

Strawson, P. F. (2008b [1974]). Causation in perception. In P. F. Strawson, *Freedom and resentment and other essays* (pp. 73-93)*.* London: Routledge.

Strawson, P. F. (2008c [1958]). Wittgenstein’s ‘Philosophical Investigations’. In P. F. Strawson, *Freedom and resentment and other essays* (pp. 147-185)*.* London: Routledge.

Strawson, P. F. (2008d [1985]). *Scepticism and naturalism. Some varieties.* London: Routledge.

Strawson, P. F. (2011 [1952]). *Introduction to logical theory.* New York: Routledge.

ter Hark, M. (1990). *Beyond the inner and the outer. Wittgenstein’s philosophy of psychology.* Dordrecht: Kluwer.

ter Hark, M. (2001). The inner and the outer. In H.-J. Glock (Ed.), *Wittgenstein. A critical reader* (pp. 199-223)*.* Oxford: Blackwell.

von Wright, G. H. (1971). *Explanation and understanding.* Ithaca: Cornell University Press.

von Wright, G. H. (1974). *Causality and determinism.* New York: Columbia University Press.

Watson, G. (1987). Responsibility and the limits of evil. Variations on a Strawsonian theme. In F. Shoeman (Ed.), *Responsibility, character, and the emotions* (pp. 374-394)*.* Cambridge: Cambridge University Press.

Witherspoon, E. (2011). Wittgenstein on criteria and the problem of other minds. In O. Kuusela, & M. McGinn (Eds), *The Oxford handbook of Wittgenstein* (pp. 472-498)*.* Oxford: Oxford University Press.

Wittgenstein, L. (1992). *Last writings on the philosophy of psychology. The inner and the outer. Volume 2.* Oxford: Blackwell.

Wittgenstein, L. (2009 [1953]). *Philosophical investigations.* Chichester: Wiley-Blackwell.

Wolf, S. (1987). Sanity and the metaphysics of responsibility. In G. Watson (Ed.), *Free will* (pp. 372-387)*.* Oxford: Oxford University Press.

Woodward, J. (2016). Causation and manipulability. In E. Zalta (Ed.), *The Stanford encyclopedia of philosophy,* Winter 2016 edition.

Zahavi, D. (2014). *Self and other. Exploring subjectivity, empathy, and shame.* Oxford: Oxford University Press.

1. Sneddon (2005) also claims that Strawson was an externalist about moral responsibility. First, however, his distinction between internalism and externalism, which is similar to the distinction between internalism and externalism about mental content (see Lau and Deutsch, 2016), seems different from Ciurria’s. For him, internalism is the view that moral responsibility is determined *solely* by internal factors, while externalism is the view that external factors *also* play a role. If the distinction is made in this way, then Strawson was indeed an externalist. Secondly, Sneddon puts a lot of weight on the idea that, according to Strawson, moral responsibility has to be understood in social or interpersonal rather than in individualistic terms. I agree, but I will not discuss that claim here. Thirdly, Sneddon is not so much seeking to *interpret* Strawson, his aim is rather to *develop* Strawson’s account or, as he puts it, to defend a *neo-*Strawsonian view (Sneddon, 2005, p. 246). [↑](#endnote-ref-1)
2. For other examples of how the rest of Strawson’s *oeuvre* helps to understand ‘Freedom and Resentment’, see Campbell (2017), Coates (2017), Heyndels (2019), and some of my own work (De Mesel, 2017; De Mesel, 2018a; De Mesel, 2018b; De Mesel and Heyndels, 2019). [↑](#endnote-ref-2)
3. Ter Hark (2001, p. 209) explains how Wittgenstein argues against a point made by Herbert Feigl (1958), namely that a certain kind of [outer] behaviour must be regarded as a probabilistic indicator of a certain kind of [inner] mental state in the manner in which symptoms in medicine are regarded as probabilistic indicators of diseases. Ciurria (2015, p. 605) makes exactly the point that Feigl makes, thus illustrating what I hope to make clear: she accepts the inner-outer picture that Wittgenstein and Strawson explicitly reject. [↑](#endnote-ref-3)
4. There has been a lot of controversy about what exactly Wittgenstein meant by ‘criterion’. I cannot go into these debates here (for a list of references, see Avramides, 2019). Avramides (2019) situates the idea of criterial relations within contemporary discussions in philosophy of mind, and she helpfully summarizes it as follows: ‘What one can say is that a criterial relation between mind and behaviour is designed to be distinct from, on the one hand, logical entailment, and on the other, inductive inference.’ On Wittgenstein’s idea of criteria, see Glock (1996, pp. 93-97) and Hacker (2019, Ch. 15, ‘Criteria’). Hacker (2019, p. 283) notes that the term ‘criterion’, as used by Wittgenstein, indicates a relation of *presumptive implication* or *defeasible support.* He claims that, partly because of the formalization of logic, reflection on criteriafalls outside the received scope of reflection on logical relations. Strawson, who explicitly warned against the limitations of formalization in logic (see, in particular, Strawson, 2011 [1952]), would have been sympathetic to this claim. Criteria have to be distinguished from symptoms, which are pieces of inductive evidence. ‘That *p* is a symptom for *q* presupposes the possibility of an independent identification of *q,* since the empirical determination of *p* as symptomatic evidence requires the inductive correlation of two distinct, externally related phenomena’ (Hacker, 2019, pp. 290-291). By contrast, if *p* is a criterion for *q,* then it is part of the *meaning* of ‘*q’* that *p*’s being the case is a ground or reason for the truth of *q.* To say that *p* is a criterion for *q* is to give a partial explanation of the meaning of ‘*q’.* We do not know what pain *is* unless we know how it is normally manifested. As Hacker (2019, p. 302) puts it: ‘It is not an empirical discovery that people scream when they are in severe pain.’ In that sense, pain and pain-behaviour are not externally, but internally related. Yet pain without pain behaviour is possible, the behavioural criteria for pain are neither necessary nor sufficient conditions. That does not make it illegitimate to justify ascriptions of pain by reference to behaviour: the claim that another person is in pain ‘cannot be undermined by the fact that the grounds supporting the claim are defeasible, but only by countervailing grounds that defeat them’ (Hacker, 2019, p. 304; see also Glock, 1996, p. 97). Strawson not only repeats Wittgenstein’s point about the criterial relation between the inner and the outer in *Individuals*, he also explicitly leaves room for the idea of ‘presumptive implication’, a logical relation that is weaker than entailment, in his work about perception (Strawson, 2008b [1974], p. 74). The idea here is that our experience of perceiving something *normally* or *generally* implies that we are perceiving it, but does not entail it, as the existence of hallucinations shows. [↑](#endnote-ref-4)
5. Compare Hacker on Wittgenstein’s use of ‘criterion’: ‘The wince of pain, the shriek of agony, the careful nursing of the injured limb are, of course, not themselves sensations. But they are not ‘mere’ behaviour either, but *pain-*behaviour, logically or grammatically bound up with the concept of pain. There is no ‘gap’ between criterial grounds for inner states and the propositions about inner states which they support’ (Hacker, 2019, p. 302). [↑](#endnote-ref-5)
6. Something similar could be said about the circumstantial factors that Strawson mentions in the group of excuses and the first subgroup of exemptions. The fact that someone was pushed, for example, excuses only insofar as it makes it the case that the agent could not help what she did. Similarly, the fact that someone was acting under post-hypnotic suggestion exempts only insofar as it makes it the case that the agent was not herself. [↑](#endnote-ref-6)
7. My point in this section has been that Ciurria’s epistemological problem can be avoided *if Strawson’s view is correct.* But is it plausible to think that there is a logical connection between the inner and the outer? The *Stanford Encyclopedia* article ‘Other Minds’ (Avramides, 2019) provides an overview of standard positions on the inner-outer relation, and it treats the Wittgensteinian/Strawsonian view, as well as the idea of criteria associated with it (see endnote 4), at some length. It outlines some of the difficulties and advantages of this view, distinguishes it from behaviourism (see also Hacker, 2019, Ch. 7, ‘Behaviour and Behaviourism’; ter Hark, 2001), and compares the way in which it makes us think about knowledge of other minds to prominent alternatives, such as the argument from analogy and inference to the best explanation arguments. Avramides rightly notes that there is a close connection between criterial views of the inner-outer relation and the idea that we have direct knowledge of other minds. This idea may not be mainstream in contemporary analytic philosophy, but it has been defended by many philosophers in the phenomenological tradition, such as Edmund Husserl, Max Scheler, Edith Stein, Martin Heidegger, Maurice Merleau-Ponty, Emmanuel Levinas and Jean-Paul Sartre, and it is still prominent in that tradition (Zahavi, 2014). It stands in contrast to the idea held by defenders of the arguments from analogy and from best explanation that any knowledge we may have of other minds must be indirect because the minds of others are not directly observable. For Strawson’s view on inner and outer, see Strawson (1992, Ch. 6). For explanation and defense of Wittgenstein’s (1992; 2009) view, see ter Hark (1990; 2001), Glock (1996, pp. 174-179), Moyal-Sharrock (2007), Witherspoon (2011), Child (2017) and Hacker (2019, Ch. 8, ‘Knowledge of other minds. The inner and the outer’). On the relation between Wittgenstein’s philosophy of mind and contemporary approaches, see Overgaard and Zahavi (2009) (who also situate Wittgenstein’s view within contemporary debates on social cognition, such as theory-theory and simulation theory), Schroeder (2001), and Ellis and Guevara (2012). For contemporary defenses of Strawsonian/Wittgensteinian views on other minds, see McDowell (1998a; 1998b), Avramides (2001), Cassam (2007), Overgaard (2007), Gomes (2011), Hacker (2013, pp. 401-404). Avramides (2001) argues that there are profound affinities between Strawson’s work on inner and outer and the work of Donald Davidson. See, for example, the following quote: ‘Though our beliefs, intentions, fears, and other feelings are private and subjective if anything is, they cannot be identified or explained except by tying them from the start to external objects and events’(Davidson, 2004, p. 96). [↑](#endnote-ref-7)
8. One might worry that Strawson’s conception of causation, as presented in this section, is implausible from a contemporary perspective. If it is, then the point that there is a problem with Ciurria’s argument if we adopt a Strawsonian conception of causation might have some exegetical significance, but will not be very significant philosophically. I cannot discuss the philosophical merits of Strawson’s view on causation here, but I want to indicate that it is being taken seriously in contemporary discussions. The *Oxford Handbook of Causation* (Beebee, Hitchcock and Menzies, 2009) distinguishes between five standard approaches to causation: (1) regularity theories, (2) counterfactual theories, (3) probabilistic theories, (4) causal process theories and (5) manipulation-based theories (agency theories and interventionist theories). Strawson’s (1985; 1992, Ch. 9) view, though described as ‘influential’ (Beebee, Hitchcock and Menzies, 2009, p. 343; see also the references to Strawson in Schaffer, 2016), cannot easily be characterized as one of the standard approaches. It is most closely related to manipulation-based views (see Strawson 1992, p. 118), in particular to those of his contemporaries Anscombe (1975), whose view he sympathetically refers to (Strawson, 1992, p. 119), and von Wright (1971; 1974). On manipulability views, see Woodward (2016). While being closest to manipulation-based theories, Strawson’s view also bears affinities to regularity theories. Strawson himself (1985; 1992, Ch. 9) discusses the relation of his position to Humean and Millian regularity views. Davidson (1985) discusses similarities between his view and Strawson’s, and they seem to agree on many fundamental points. For discussion of Strawson’s position on causation, and some ways in which it differs from regularity theories and standard manipulation-based theories, see Roessler (2011). For a contemporary defense of a Strawsonian position in relation to alternative views, see Hacker (2010, Ch. 3, ‘Causation’). [↑](#endnote-ref-8)
9. The views expressed in this paragraph are not unique to Strawson. Davidson (2004, pp. 94-97) makes very similar points. [↑](#endnote-ref-9)
10. This work was supported by the Research Foundation – Flanders (FWO). An earlier version of this paper was presented as the Annual Wittgenstein Lecture (‘Strawson’s Wittgensteinian Approach to Moral Responsibility’) at the 55th Meeting of the Welsh Philosophical Society in April 2019. I am grateful for the many helpful questions and comments I received on this occasion, and especially to David Cockburn for the invitation to deliver the lecture. Special thanks are owed to an anonymous referee for this journal, and to Stefaan Cuypers for very helpful advice on causation and the inner/outer distinction. [↑](#endnote-ref-10)