

Moral Responsibility and the Moral Community: Another Reply to Zimmerman

Benjamin De Mesel^{1,2} 

Received: 13 November 2017 / Accepted: 9 January 2018 / Published online: 16 January 2018
© Springer Science+Business Media B.V., part of Springer Nature 2018

Abstract Michael Zimmerman has recently argued against the twofold Strawsonian claim that there can be no moral responsibility without a moral community and that, as a result, moral responsibility is essentially interpersonal. I offered a number of objections to Zimmerman’s view, to which Zimmerman responded. In this article, I respond to Zimmerman’s responses to my criticisms.

Keywords P.F. Strawson · Moral responsibility · Interpersonal · Moral community · Reactive attitudes · Holding responsible · Michael Zimmerman

1 Introduction

Zimmerman (2016, 2017) addresses the following thesis:

Slogan: There can be no moral responsibility without a moral community.

Zimmerman focuses on an interpretation of the Slogan according to which (a) ‘can’ expresses conceptual possibility, (b) the kind of responsibility in question is retrospective and consists primarily in either culpability or laudability for some past event, and (c) the kind of community in question consists, roughly, in a group of persons who owe each other duties and who hold each other responsible in the event that these duties are not fulfilled. According to Zimmerman, the Slogan, so understood, implies that there is a significant sense in which the kind of responsibility at issue is essentially interpersonal (Zimmerman 2017: 251).

✉ Benjamin De Mesel
benjamin.demesel@kuleuven.be

¹ Centre for Ethics, Social and Political Philosophy, Institute of Philosophy, KU Leuven, Leuven, Belgium

² Vijfhoekstraat 55, 1770 Liedekerke, Belgium

Zimmerman claims that there are two popular views about moral responsibility that imply that the Slogan is true. The first is this (where P and Q are distinct persons and x is whatever sort of thing P might be responsible for):

Answerability:

P is retrospectively morally responsible to Q for $x = \text{df.}$ It is appropriate for Q to call on P to answer for (i.e., to defend or justify) x .

The other view is this:

Responsibility-1:

P is retrospectively morally responsible for $x = \text{df.}$ There is some person Q distinct from P for whom it is appropriate to adopt some reactive attitude toward P in respect of x .

According to Zimmerman, Responsibility-1 is endorsed by Strawson and many Strawsonians. He argues against both Answerability and Responsibility-1 (Zimmerman 2016).

I agree with Zimmerman that both Answerability and Responsibility-1 should be rejected, but I think that Strawsonians can nevertheless defend the claim that moral responsibility is essentially interpersonal in a significant way. They could do so by arguing for this:

Responsibility-2:

P is retrospectively morally responsible for $x = \text{df.}$ If there were a person Q distinct from P such that Q satisfied certain conditions C , it would be appropriate for Q to adopt some reactive attitude toward P in respect of x .

Zimmerman (2016, 2017) argues that there are reasons to doubt Responsibility-2, in which case we might have to retreat to the following:

Responsibility-3:

Necessarily, P is retrospectively morally responsible for x if and only if it is the case that, if there were a person Q distinct from P such that Q satisfied certain conditions C , it would be appropriate for Q to adopt some reactive attitude toward P in respect of x .

In contrast to Zimmerman, I think that there is a significant sense in which it might be said that, according to Responsibility-2 and Responsibility-3, moral responsibility is essentially interpersonal (De Mesel 2017). Zimmerman disagrees. In this article, I will respond to his objections.

2 Self-Reactive Attitudes

According to Zimmerman (2016), the existence of self-reactive attitudes fits uneasily with the idea that moral responsibility is essentially *interpersonal*. I noted that Strawson's three types of reactive attitudes (personal, moral, and self-reactive attitudes) are closely connected and suggested that, in light of the distinction

between self- and other-reactive attitudes, Responsibility-2 could be understood as follows:

Responsibility-2 (extended):

P is retrospectively morally responsible for x = df. If there were a person *Q* distinct from *P* such that *Q* satisfied certain conditions *C*, it would be appropriate for *Q* to adopt some other-reactive attitude toward *P* in respect of x , and it would be appropriate for *P*, if (s)he satisfied certain conditions *C*, to adopt some self-reactive attitude toward him-/herself in respect of x .

Zimmerman's problem with this is 'the assumption that, no matter what that thing, x , is for which *P* is morally responsible, there could always be some distinct person, *Q*, for whom it would be appropriate to adopt some reactive attitude toward *P* in respect of x ' (Zimmerman 2017: 255).

I agree that there are problems with the assumption. x could be a morally neutral act, one for which *P* is morally responsible but for which (s)he is neither blameworthy nor praiseworthy. Taking that possibility into account, Responsibility-2 would have to be adapted as follows: '*P* is retrospectively morally responsible for x in the sense of being praiseworthy or blameworthy for it = df. If there were ...' (see McKenna 2012: 16–17, 38). Another reason to doubt Responsibility-2 has to do with the notion of appropriateness. According to McKenna, the appropriateness must be understood to offer a *pro tanto* reason, one that could be defeated by other reasons, rather than an all-things-considered reason. Maybe *Q* does not have the moral standing to blame *P*, maybe blaming *P* would not be worth the cost, etc. Thus, it would be better to include in Responsibility-2 that 'it would be *pro tanto* appropriate for *Q* ...' (McKenna 2012: 36–38).

Zimmerman's problem is different. He notes that the assumption 'may seem safe, if we also assume that *P* is a member of some moral community, but of course the latter assumption is one that I am not prepared to make' (Zimmerman 2017: 255). I am inclined to think that there cannot be a person who is not a member of some moral community, but this is not something for which I can argue here, and it is clear that Zimmerman and I have different ideas about personhood (De Mesel 2017: 316–320; Zimmerman 2017: 260). More interesting for my purposes is what Zimmerman adds, namely, that even if *P* is a member of some moral community, the assumption seems questionable:

Suppose, for example, that *P* is responsible for the mistreatment of a wild dog or the wanton destruction of some wild plants. In such a case, it seems clear that it would not be appropriate for any other person, *Q*, to resent what *P* has done, given that resentment is appropriate only when one has oneself been mistreated in some way. It seems almost as clear that it would not be appropriate for anyone else to feel indignant about what *P* has done, given that indignation consists in a kind of vicarious resentment. (Zimmerman 2017: 255)

It is certainly true that resentment is appropriate only when one has oneself been mistreated in some way: the personal reactive attitudes are attitudes towards attitudes of others towards oneself. The self-reactive attitudes, as Strawson

describes them, are attitudes towards our own attitudes towards other persons (or towards ourselves), while the moral reactive attitudes are attitudes towards the attitudes of others towards third parties. However, as I remarked in my previous response to Zimmerman (De Mesel 2017: 312–313), it is admitted by authors sympathetic to Strawson (see especially Shabo 2012: 104) that self-reactive attitudes need not be described as attitudes towards our own attitudes towards *persons*, but can instead be described as attitudes towards our own attitudes towards persons, animals or things, and this modification in the description of the self-reactive attitudes explains why *P* may feel guilty for the mistreatment of a dog or the destruction of the environment. Similarly, moral reactive attitudes need not be described as attitudes towards others' attitudes towards third *persons*, but can instead be described as attitudes towards others' attitudes towards persons, animals or things, and this explains why *Q* may feel indignant about *P*'s mistreatment of a dog or his wanton destruction of the environment. Shabo claims that we may feel indignation toward someone who has carelessly despoiled a scenic area (Shabo 2012: 104), and it is difficult to understand why that would be impossible, as Zimmerman thinks.

Zimmerman may think that, within a Strawsonian framework, indignation toward someone who has carelessly despoiled a scenic area is impossible because, according to Strawson, indignation consists in a kind of vicarious resentment. If I have not been mistreated in some way, then resentment cannot be appropriate. If it is not appropriate for me to feel resentment in respect of *x*, and indignation is vicarious resentment, then it will not be appropriate for some third person to feel indignation in respect of *x*. However, if the self-reactive and moral reactive attitudes are redescribed as outlined in the previous paragraph, there will be cases in which indignation is appropriate while resentment is not. We will then have to understand Strawson's link between indignation and resentment in a way that is less strong than the way in which Zimmerman understands it (as obtaining in paradigm cases but not in derivative or marginal cases, for example). In any case, it seems ad hoc for Zimmerman to admit that we may feel guilty for the destruction of the environment, thereby allowing for the redescription of self-reactive attitudes as attitudes towards our own attitudes not only towards persons, but also towards animals or things, while excluding that the moral reactive attitudes could be redescribed in an analogous way, as attitudes towards others' attitudes not only towards persons, but also towards animals or things.

Responsibility-2 (extended) does not imply that one can only be morally responsible for behavior that affects other persons; there is no requirement that the *x* be construed as some kind of behavior towards other persons. Responsibility-2 (extended) implies that if *P* is morally responsible (in the sense of being blameworthy or praiseworthy) for his behavior towards a dog or towards the environment, then it would be *pro tanto* appropriate for *P* to adopt a self-reactive attitude (of guilt, for example), and it would be *pro tanto* appropriate for *Q* to adopt a moral reactive attitude (of indignation, for example) toward *P* in respect of *P*'s behavior, and *vice versa*.

3 The Priority Problem

Zimmerman argues that there may be reason to adopt Responsibility-3 rather than Responsibility-2. The reason is that

[...] if (a) it is correct to say that, if it is ever appropriate for someone Q to adopt some reactive attitude toward someone P in respect of x , that will be *because* P is responsible for x , and (b) this statement is correctly interpreted as saying that the appropriateness of Q 's attitude *supervenies on* (i.e., *is grounded by*) P 's responsibility, then (c) Responsibility-2 must be rejected. (I take (c) to follow from (a) and (b) because a decompositional analysis presupposes that analysans and analysandum are identical, whereas supervenience is asymmetrical.) (Zimmerman 2017: 256)

Zimmerman's point is that the because-statement (if it is ever appropriate for Q to adopt some reactive attitude toward someone P in respect of x , that will be *because* P is responsible for x) suggests that the relation between the analysans and the analysandum of Responsibility-2 is asymmetrical, while a proper analysis requires the relation between analysans and analysandum to be symmetrical.

I accept Zimmerman's conditional statement that if (a) and (b) are true, then Responsibility-2 must be rejected. Zimmerman is inclined to accept (a) and (b), but he acknowledges that the conjunction of (a) and (b) is open to challenge, which is why he couches it in conditional terms (Zimmerman 2017: 256). In order to provide us with an argument against Responsibility-2, however, (a) and (b) must be true. I accept (a), but there are reasons to doubt (b).

I think that Zimmerman's because-statement (a) is true. My question is whether the truth of this statement indeed implies that the relation between analysans and analysandum in Responsibility-2 is asymmetrical. It does so only, I would say, if the converse because-statement is false, that is, if it is not *also* true that, if P is responsible for x , that will be *because* it is appropriate for Q to adopt some reactive attitude toward someone P in respect of x . If *both* because-statements are true, the relation between analysans and analysandum will not be asymmetrical, and Zimmerman has not offered an objection to Responsibility-2.

But how can both because-statements be true? Zimmerman suggests that his because-statement is correctly interpreted as saying that the appropriateness of Q 's attitude *supervenies on* (i.e., *is grounded by*) P 's responsibility. Can both because-statements be true if they are understood in this way? Not if grounding is understood as it is commonly understood, namely as an asymmetric relation: if φ grounds ψ , then it cannot be the case that ψ grounds φ .

Are there other ways to understand the because-statements? And if it is neither the case that being responsible grounds the appropriateness of adopting a reactive attitude nor that the appropriateness of adopting a reactive attitude grounds being responsible, then what grounds being responsible and what grounds the appropriateness of adopting a reactive attitude? These questions are insightfully discussed by Manata (2016) in relation to McKenna's (2012) interdependence thesis, which is

basically the idea that the analysans and analysandum of Responsibility-2 are symmetric and mutually dependent. Neither is more basic.

Manata makes clear that McKenna never wished to view the interdependence thesis as an intergrounding thesis (Manata 2016: 68). According to the interdependence thesis, both being morally responsible and holding morally responsible are ‘significantly implicated in a direct metaphysical explanation of the other’ (McKenna 2012: 81). The relation between grounding and metaphysical explanation is complex, but their formal properties are similar (Manata 2016: 69), so metaphysical explanation is asymmetric too. Thus, McKenna may be taken to suggest that being responsible explains holding responsible and the other way round, but this suggestion violates a formal principle of metaphysical explanation and should therefore be rejected.

The key to Manata’s solution is the following:

It is open for McKenna to deny that being and holding morally responsible metaphysically explain one another at all. This is so because, as it seems to me, the claim that φ is ‘implicated in a direct metaphysical explanation’ of ψ does *not* commit one to the claim that φ metaphysically *explains* ψ . (Manata 2016: 70)

Thus, being responsible and holding responsible do not ground each other, and they do not explain each other. This leaves us with two questions. (1) What, if anything, grounds being and holding morally responsible? (2) How are being and holding morally responsible each implicated in metaphysical explanations of the other? Manata refers to (1) as the *missing ground question* and (2) as the *role-playing question*.

In brief, Manata’s answers are as follows. With regard to (1), he suggests that we see being morally responsible and holding morally responsible as elements in a complex moral responsibility system, and it is the whole system that grounds them. While it is often assumed that wholes are grounded in their parts, this is questionable when dealing with systems (Manata 2016: 79).

A system is a set of interconnected things that produce a pattern of behaviour (Meadows 2008: 2). As I shall mean it, then, our moral responsibility system is constituted by what Vargas (2013: 109–110) calls ‘core phenomena’ or ‘aspects’ such as ‘being responsible, holding responsible, responsible agency, exculpation, blameworthiness, blame, and desert’. Other aspects of the ‘system’ Vargas includes are ‘epistemic conditions’ and ‘some sense of the agent being active with respect to the action’. (Manata 2016: 78)

Manata’s answer to (2) is connected to his answer to (1): both being morally responsible and holding morally responsible are features of a system, and it is the system that does the metaphysical explaining, which is why being and holding morally responsible are ‘implicated’ in a metaphysical explanation of each other. However, Manata emphasizes, ‘they do not *do* the grounding [the metaphysical explaining], since on the view presented above, entangled systems are not reducible to or identical to their parts’ (Manata 2016: 80; on the relation between grounding and metaphysical explanation, see Manata 2016: 69).

I have only sketched Manata's views, and I refer to his paper for elaboration. What is important for our purposes is that Manata offers a way to respond to Zimmerman's worry. He offers what I regard as a convincing argument for accepting (a) and rejecting (b). Zimmerman's because-statement will be true, according to Manata's view, because holding morally responsible is grounded in and explained by a complex system of which being morally responsible is a part. The converse statement will also be true, because being responsible is grounded in and explained by a complex system of which holding morally responsible is a part. Thus, there is a way to accept (a) and reject (b), and a way to avoid the conclusion that Responsibility-2 has to be rejected.

This response to Zimmerman's problem of priority differs from the one that I gave earlier (De Mesel 2017). I had understood Zimmerman's because-statement as follows. The because-statement suggests that being responsible is prior to holding responsible. By contrast, Responsibility-2 suggests that holding responsible is prior to being responsible, at least if it is understood as providing a decompositional analysis of the concept of moral responsibility, that is, an analysis that purports to analyze a complex concepts into more basic or fundamental concepts. Because Responsibility-2 suggests that holding morally responsible is prior to being morally responsible, while the because-statement makes clear that being morally responsible is prior to holding morally responsible, Responsibility-2 must be rejected. My reply was that the objection only holds if Responsibility-2 is understood as providing a decompositional analysis. However, Strawson prefers connective analysis over decompositional analysis (De Mesel 2017: 327–328), and a connective analysis does not put forward any kind of priority relation between analysans and analysandum. It is therefore compatible with the fact that being morally responsible is prior to holding morally responsible. If we read Responsibility-2 as a connective analysis, Zimmerman's objection vanishes.

Zimmerman's answer to my reply is, roughly, that I have misunderstood him: he was not talking about conceptual priority, but about metaphysical priority. I have provided an answer to the metaphysical priority objection in the previous paragraphs, but I would like to point out here, however briefly, that I do not regard my earlier reply as fundamentally mistaken. Roughly, the reason is that, in my view and in the Strawsonian view, metaphysical and conceptual priority are much closer related than Zimmerman assumes. Zimmerman's response relies on a sharp distinction between conceptual and metaphysical priority: 'Mutual elucidation is one thing, *interdependence* is another' (Zimmerman 2017: 257–258). I do not think that Strawson clearly separates the two (for more detailed support of this, see Hacker 2001 on Strawson's method of descriptive metaphysics), and I am not convinced that they have to be sharply distinguished. But no defense of that point is required here: the reply outlined in this section does not need it, and that is an advantage in comparison to the reply that I gave earlier.

4 The Circularity Problem

I have been writing about being the appropriate target of some reactive attitude and being appropriately held responsible, as if the two were interchangeable. Zimmerman suggests that being the target of some reactive attitude (the formulation of Responsibility-2) does not seem to imply being held responsible,

[...] since the concept of being held responsible would seem to involve the concept of being responsible, whereas the concept of being the target of some reactive attitude does not involve the concept of being responsible. And that is surely all to the good. *Any* analysis, whether decompositional or merely connective, is surely defective if one and the same concept is involved on both sides of the analysis. No concept can be elucidated in terms of itself. (Zimmerman 2017: 257)

Zimmerman adds that it is open to Strawsonians to claim that Responsibility-2 ‘could just as well be couched in terms of being held responsible as in terms of being the target of some reactive attitude, precisely because (what they understand as) the concept of being held responsible does *not* involve the concept of being responsible’ (Zimmerman 2017: 258). He goes on to doubt, however, whether this is what Strawsonians typically have in mind, and quotes McKenna, who writes that ‘holding a person morally responsible for an action comes to *more than* merely judging ... [that] the person is morally responsible for that action’ (McKenna 2012: 36; Zimmerman’s emphasis). According to Zimmerman, McKenna suggests that holding responsible includes judging responsible, which itself involves being responsible. If holding responsible involves being responsible, then Responsibility-2 and Responsibility-3 are in trouble, because they would then provide an analysis of a concept in terms of itself.

Strawsonians have indeed argued that the concept of being held responsible does not involve the concept of being responsible (see, for example, Maher 2010). But suppose that it does, as the quote from McKenna seems to imply. Do we then have to reject Responsibility-2 and Responsibility-3 because they are circular? Manata anticipates the circularity charge and discusses it at some length (Manata 2016: 71–73). He notes that issues surrounding what makes circular analyses unacceptable are quite complex (Walton 2006). The idea is, roughly, that circularity is not always vicious. The difference between decompositional and connective analysis is important here. McKenna’s project is one of connective analysis. He aims to provide an elucidation of the concept of moral responsibility, not a reductive explanation in terms of simpler concepts. Manata comments, as if he foresees Zimmerman’s worry:

This non-reductivist project might be seen to invite the charge of problematic circularity again. But here, McKenna might appeal to Woodward’s (2003: 20–22) defence of non-reductive explanations in his ‘manipulability’ theory of causation. Woodward notes that non-reductivist theories can be generally illuminating even if they elucidate the concept of cause with concepts that do not lie outside the interrelated circle of causal concepts. They may be

illuminating, for example, if they conflict with other reductivist or non-reductivist theories of some concept φ (Woodward 2003: 21). Secondly, non-reductivist theories may ‘face nontrivial choices about exactly how the various notions in this circle should be connected with or used to elucidate one another – choices that can be made in more or less defensible ways’ (Woodward 2003: 21). (Manata 2016: 72–73)

The idea that not all circularity is vicious and that some may even be benign is, of course, not new. It has been defended at some length by David Wiggins in his defense of fitting-attitude analyses of value, and Zimmerman (2016: 314) has rightly observed that such analyses are analogous to Responsibility-2 and Responsibility-3 in important ways. A fitting-attitude analysis, according to Wiggins, wants

[...] to persuade us that, when we consider whether or not x is good or right or beautiful, there is no appeal to anything that is more fundamental than actually possible human sentiments [...] Circularity as such is no objection to it, provided that the offending formulation is also *true*. But what use (I shall be asked) is such a circular formulation? My answer is that, by tracing out such a circle, the subjectivist hopes to elucidate the concept of value by displaying it in its actual involvement with the sentiments. One would not, according to him, have sufficiently elucidated what value is *without* that detour. (Wiggins 1998b: 188–189)

It is not a coincidence that Wiggins uses the word ‘elucidation’ here. He explicitly recognizes that he uses it in contrast to ‘analysis’ (Wiggins 1998b: 188, footnote 4; see also Wiggins 1998a: §§2–4), and it is clear that by ‘analysis’ he means ‘decompositional analysis’.

I believe, in short, that Zimmerman’s charge of circularity loses much of its force when Responsibility-2 and Responsibility-3 are read as connective analyses. Maybe Responsibility-3 is a better candidate for connective analysis than Responsibility-2, and I have indicated that I am happy to accept Responsibility-3 instead of Responsibility-2 (De Mesel 2017: 328–331). Responsibility-3 may be circular, but what has to be shown (and what I do not think Zimmerman has shown) is that the circularity is vicious instead of benign.

In my reply to Zimmerman’s circularity problem, I have relied on work by Manata, Wiggins, and McKenna. But why think that my reply is accurate from a Strawsonian point of view? Here is Strawson himself:

For there is a certain form of words which the analytical philosopher hates to hear and which his opponent in argument, also an analytical philosopher, delights to pronounce: viz. the words, ‘Your analysis is circular’. This means, of course, that included in the elements of his analysis, though perhaps covertly included and only to be revealed by further steps of the same kind, is the very concept which the philosopher is claiming to analyse.

Now why should this formula be felt to be so damaging? Well, of course, the formula ‘Your analysis is circular, it suffers from circularity’ really is damaging, indeed fatally damaging, to the pretended analysis if we are

thinking in terms of that model of analysis which represents it as a kind of dismantling of a complex structure into simpler elements, a process which terminates only when you reach pieces which cannot be further dismantled; for this process has not even begun if one of the alleged pieces turns out to be, or to contain, the very thing, the very concept, that was to be dismantled.

But now let us consider a quite different model of philosophical analysis. This new model [of connective analysis] I am going to declare more realistic and more fertile than the one just discussed. [...] Let us imagine, instead, the model of an elaborate network, a system, of connected items, concepts, such that the function of each item, each concept, could, from the philosophical point of view, be properly understood only by grasping its connections with the others, its place in the system – perhaps better still, the picture of a set of interlocking systems of such a kind. If this becomes our model, then there will be no reason to be worried if, in the process of tracing connections from one point to another of the network, we find ourselves returning to, or passing through, our starting-point. We might find, for example, that we could not fully elucidate the concept of knowledge without reference to the concept of sense perception; and that we could not explain all the features of the concept of self perception without reference to the concept of knowledge. But this might be an unworrying and unsurprising fact. So the general charge of circularity would lose its sting, for we might have moved in a wide, revealing, and illuminating circle. This is not to say that the charge of circularity would lose its sting in every case. Some circles are too small and we move in them unawares, thinking we have established a revealing connection when we have not. But it would be a matter for judgement to say when the charge was damaging and when it was not. (Strawson 1992: 18–20)

This response makes quite clear that Strawsonians need not think of circularity as necessarily problematic.¹ It is still open to Zimmerman, of course, to argue that the circularity in question is problematic, for example because the circle is too small, but the burden of proof is on Zimmerman's side here.² A way for Zimmerman to

¹ Note also that Strawson explicitly asks us to imagine the model of a 'system of connected items' (for another reference to the idea of a system or network, see Strawson 2011: 184). This reference seems to support Manata's interpretation of McKenna's Strawsonian interdependence thesis. Elsewhere, Strawson writes that the major task of analytical philosophy may well be 'to establish the connections between the major structural elements of our conceptual scheme—to exhibit it, not as a rigidly deductive system, but as a coherent whole whose parts are mutually supportive and mutually dependent, interlocking in an intelligible way' (Strawson 1985: 18). Manata thinks of our responsibility system in exactly this way: a system whose parts are mutually supportive and mutually dependent, interlocking in an intelligible way.

² In a comment on this paper, Zimmerman suggests a way to distinguish between problematic and unproblematic circularity. According to him, a circular analysis of the following form is problematic: (A) x is $F = \text{df.} \dots F \dots$. A wider circle of the following form is not necessarily problematic: (B) x is $F = \text{df.} \dots G \dots$; x is $G = \text{df.} \dots H \dots$; x is $H = \text{df.} \dots I \dots$; x is $I = \text{df.} \dots J \dots$; x is $J = \text{df.} \dots F \dots$. Zimmerman suggests that McKenna's proposal has the (problematic) form of (A). I cannot discuss Zimmerman's proposal in detail here. I suspect that the distinction between problematic and unproblematic circularity is not equivalent to the distinction between (A) and (B), that is, that some (A)-type circularities are illuminating and some (B)-type circularities are problematic. When Strawson writes that 'it would be a matter for judgment' whether circularity in a given case is damaging, I take him

avoid having to bear this burden is to insist that, for him, circularity is always problematic, and it does not matter much what Strawson or Strawsonians think about it. Does he not have the right to stipulate that the answer to his central question should not be circular in any way? That question will be discussed in the last section of this paper.

5 The Significance Problem

According to Zimmerman (2016), Responsibility-2 does not imply that moral responsibility is essentially interpersonal in a significant sense. I argued that it does: because ‘interpersonal’ means ‘relating to relationships or communication between people’, and because the notion of a reactive attitude in Responsibility-2 provides a necessary link between moral responsibility and relationships/communication between people, Responsibility-2 implies that moral responsibility is interpersonal in the commonly accepted sense of that term (De Mesel 2017: 314–316). Moral responsibility is, according to Responsibility-2, essentially interpersonal, because one would not have sufficiently elucidated what moral responsibility is without an appeal to the reactive attitudes.

In his reply, Zimmerman helpfully provides some examples of phenomena that are essentially interpersonal in the commonly accepted sense of the term, phenomena that he takes to differ significantly from moral responsibility. His examples include marriage, citizenship, community service, the playing (rather than practicing) of team sports, the playing of many non-team sports (such as ‘singles’ tennis) and other games (such as chess), and cooperation and collaboration in general (Zimmerman 2017: 260). Three questions are important here:

- (1) Is it necessary that another person is actually present?
- (2) Is it necessary that another person is alive?
- (3) Is it necessary that another person has ever been alive?

Let us take the case of marriage. I am still married when I am abroad for a conference, so the answer to (1) is no. I am no longer married when my partner is no longer alive, so the answer to (2) is yes, and that implies that the answer to (3) is also yes. It is not obvious that all examples lead to the same answers. Maybe I could still be a citizen when all other people have died, and the answer to (2) might be no in the case of citizenship while the answer to (3) is yes. It is possible to play chess against a computer that does not count as a person, so the answer to (1) and (2) is no in the case of chess.

I questioned in my earlier reply whether it is possible for a person P to be responsible for x even if no other person Q has ever existed. I argued that it is

Footnote 2 continued

to suggest that there is no way to formalize the distinction. I also want to recall, as noted in the beginning of this section, that Strawsonians could argue that the concept of holding responsible does not involve the concept of being responsible (see, for example, Maher 2010). If that is true, there is no circularity at all in Responsibility-2 or Responsibility-3.

doubtful whether it is possible for a person P to exist even if no other person Q has ever existed, because it might be essential to persons that they are born of other persons (De Mesel 2017: 316–320). Zimmerman disagrees (Zimmerman 2017: 260). I do not want to start a discussion about personhood here, so let us assume for the sake of the argument that Zimmerman is right and that it is possible for a person P to exist even if no other person Q has ever existed. If that is possible, and if it is possible for such a person to have the capacities for being morally responsible, as Zimmerman thinks, then I do not see why it would be impossible for that person to play chess against a computer. So here we have a case in which something (chess) is essentially interpersonal, according to Zimmerman, while the answer to (1), (2) and (3) is no. What I want to point out is that Zimmerman's examples of essentially interpersonal phenomena lead to very different combinations of answers to questions (1)–(3). Three times yes is possible, three times no seems also possible. And that just means, in my view, that these questions do not allow us to distinguish what is essentially interpersonal from what is not.

The chess analogy is quite useful when thinking about moral responsibility. It could be remarked that in the case of chess we need two players. These players need not be persons, but that does not matter much. What matters is that there are two roles. We could marry robots that are not persons, we could play football against a team of robots that are not persons, we could even marry ourselves or play chess against ourselves, but still marriage and the playing of team sports are essentially interpersonal in Zimmerman's view (and in mine).³ When a shift from persons to roles is allowed for, our question becomes: is the existence of something that plays the second role necessary for moral responsibility if we accept Responsibility-2? With respect to that question, I would like to come back to this:

Responsibility-2 (extended):

P is retrospectively morally responsible for x = df. If there were a person Q distinct from P such that Q satisfied certain conditions C , it would be appropriate for Q to adopt some other-reactive attitude toward P in respect of x , and it would be appropriate for P , if (s)he satisfied certain conditions C , to adopt some self-reactive attitude toward him-/herself in respect of x .

I argued that this is nothing more than Responsibility-2, unpacked in the light of the distinction between self- and other-reactive attitudes. Suppose that there is no other person Q distinct from P . In that case, P will be morally responsible if and only if it would be appropriate to adopt some self-reactive attitude toward him-/herself. To adopt a self-reactive attitude towards oneself *is* to allow for the role of the other, the role of the moral community. That is what McKenna means when he says that Robinson Crusoe 'carried with him in his head, so to speak, the moral community of his earlier life' (McKenna 2012: 108).

My reply to Zimmerman's examples of essentially interpersonal phenomena can be summarized as follows. Most of his examples do not require the existence of another *person*, they only require two roles. However, there is nothing that, in principle, excludes that these roles be played by the same person. If what makes

³ On self-marriage or sologamy, see <https://en.wikipedia.org/wiki/Sologamy>.

something essentially interpersonal is that there are different roles to be played, then moral responsibility will be essentially interpersonal according to Responsibility-2: even if no person Q distinct from P ever existed (a possibility that still seems doubtful to me), the notion of a self-reactive attitude does the required work.

This may be too quick. After all, there seems to be an important difference between chess and marriage on the one hand, and moral responsibility on the other. According to Responsibility-2 (extended), moral responsibility requires the *appropriateness* of a reactive attitude, that is, the appropriateness of some kind of response from an 'other'. What is not required, however, is the actual existence of such a response: P could be morally responsible without ever adopting a self-reactive attitude. It is clear that, in the case of marriage, the actual existence of a partner is required (even if one is married to oneself). Things are less clear in the case of chess, though. Suppose that Alice sets up a chessboard and makes her first move according to the rules of chess. Suppose that she turns out to be the only person who has ever existed. Was she playing chess? I think we can say (or, at least, I think that Zimmerman could say) that she was. We have a situation where some kind of response from an 'other' is appropriate, although there is no such actual response, and this situation is relevantly similar, in my view, to a situation in which someone is morally responsible while no other person has ever existed.

The chess analogy could be accepted without leading to the conclusion that moral responsibility is essentially interpersonal. Zimmerman could admit that there is a relevant analogy between chess and moral responsibility, but maintain that what the analogy shows is not that moral responsibility is interpersonal, but rather that chess is not essentially interpersonal, as he initially thought.⁴ I would accept, at this point, that if chess is not essentially interpersonal, then moral responsibility will not be essentially interpersonal either; but if chess will no longer count as essentially interpersonal, then we are no longer speaking of 'interpersonal' in the commonly accepted sense of that term, but in a restricted sense, while the interesting question is whether moral responsibility is essentially interpersonal in the commonly accepted sense.

Of course, the sense in which we can say that Alice was playing chess is a derivative sense (as are cases of self-marriage and playing chess against ourselves): we cannot understand what she is doing if we do not understand what it would be like to play chess in the presence of another player. Because an understanding of the latter case is fundamental for an understanding of what it *is* to play chess, we can still say, as Zimmerman says, that chess is essentially interpersonal. Similarly, the sense in which P is responsible for beating his dog even if no other person has ever existed is a derivative sense. But that is precisely the Strawsonian point: cases in which a person is morally responsible in the absence of others are parasitic upon cases in which others *are* present, and can be understood only in relation to what happens in these paradigm cases. A useful analogy might be drawn, in this respect, between the concept of moral responsibility and the concepts of a question and an invitation. It is possible, even common, that questions are not answered. That does

⁴ In a comment on this paper, Zimmerman takes this route and retracts his claim that chess is essentially interpersonal.

not mean that the concept of a question is not essentially interpersonal (that is, does not essentially relate to communication between people), because one would not have sufficiently elucidated what a question is without an appeal to the notion of an answer.

6 Does It Matter What Strawson Thinks?

It is necessary now to take up a distinction that I have hitherto neglected. Zimmerman wants to answer what he calls The Question, and The Question consists of two parts:

- (1) Is it possible to *be* morally responsible in the absence of a moral community?
- (2) Is it possible to *understand the concept* of moral responsibility without recourse to the concept of a moral community?

My answers and arguments should be more or less clear from what I have said above, but I will repeat them here for the sake of clarity. My answer to (2) is negative (see previous paragraph). As for (1), it depends on what is meant by the question. If the question is equivalent to ‘Is it possible to be morally responsible when no other person is actually present?’, my answer is yes. If it means ‘Is it possible to be morally responsible when no other person is alive?’, my answer is, again, yes. If it means ‘Is it possible to be morally responsible when no other person has ever been alive?’, my answer is more complicated. First, I have argued in my previous reply to Zimmerman that it might not be possible for one person to exist if no other person ever existed. Second, if it is possible for one person to exist if no other person ever existed, it is doubtful whether such a person could ever develop the capacities required for moral responsibility in the absence of other persons (see De Mesel 2017: 317–318). If there were no doubts about these points, I would give a positive answer to (1), but I still tend to answer (1) in the negative.

In what I just said about (1), I assumed that the question expresses what Strawson would call ‘bare’ conceptual possibility. However, Strawson explicitly distinguishes ‘bare’ or ‘absolute’ conceivability from ‘practical’ conceivability, and I have argued that Zimmerman (2016) has overlooked that distinction (De Mesel 2017: 320–322). Strawson seems primarily concerned with practical possibility, that is, with possibility *for us as we are*, not with bare conceivability. In the light of that observation, Strawson’s question may not be The Question (1), but rather the following:

- (1*) Is it possible *for us as we are* to be morally responsible in the absence of a moral community?

According to Strawson, the answer to *this* question is no, and Zimmerman tends to agree (Zimmerman 2017: 261). He points out, and rightly so, that a negative answer to Strawson’s question (1*) is compatible with a positive answer to The Question (1): that it is impossible for us as we are to be morally responsible in the absence of

a moral community does not imply that it is impossible for *someone* to be morally responsible in the absence of a moral community.

What this could mean, then, is that the disagreement between Zimmerman and Strawson is not a substantive disagreement about the answer to an agreed upon question. The questions that they set out to answer turn out to be different, and the answers may well be compatible. The real disagreement is to be situated at another level, and I take that conclusion to be a significant result of my discussion with Zimmerman. First, there is disagreement about whether (1) or (1*) is, philosophically speaking, the most important and interesting question. Maybe there are arguments to be made in favor of (1) or (1*), but maybe both authors just emphasize, and are struck by, a different fact. For Strawson, this is the fact that it is impossible for us as we are to be morally responsible in the absence of a moral community. For Zimmerman, it is the fact that it may be possible for a person who is not one of us to be morally responsible in the absence of a moral community. Second, there is disagreement about how to answer the important question [(1) for Zimmerman, (1*) for Strawson]. Strawson and Zimmerman agree that their respective answers require a conceptual analysis of moral responsibility. They disagree, however, about the form that such an analysis should take (decompositional or connective), and about what would make the analysis defective and/or illuminating (see the circularity problem).

Zimmerman could happily accept these points. He has the right to stipulate that his question is (1) and not (1*), that it has to be answered by way of a decompositional analysis, and that the answer should not be circular in any way. If these stipulations cannot be defended from a Strawsonian perspective, then Strawson has not defended the Slogan as Zimmerman understands it. Thus, Zimmerman's general claim that the Slogan as he understands it has not been convincingly defended still stands. What should have been avoided is, first, the impression that the Strawsonian view does involve a defense of the Slogan as Zimmerman understands it. It should be remarked that this impression is not given by Zimmerman's (2016) article alone, and there is much to say for the idea (though I will not defend it here) that some Strawsonians have read Strawson as a defender of the Slogan as Zimmerman understands it, which may be due to the fact that, in contrast to 'Freedom and Resentment', Strawson's methodological work is often neglected.

So Zimmerman could say that the Strawsonian view, as the view defended by some Strawsonians, does involve a defense of the Slogan, and I could agree. I only want to emphasize that there is also a Strawsonian view, as the view defended by Strawson himself, that does *not* involve a defense of the Slogan as Zimmerman understands it, but that *does* involve a significant interpretation of the idea that moral responsibility is essentially interpersonal. There are other ways to read the Slogan, and it goes too far to say that 'no interpretation of it [...] captures a significant truth' (Zimmerman 2016: 248). These ways are expressed by Responsibility-2 and/or Responsibility-3, understood as Strawson understands them. I have not provided a full defense of Responsibility-2 and/or Responsibility-3; rather, I have attempted to show that, when understood in a genuinely Strawsonian way, they (1) are not vulnerable to Zimmerman's objections, and (2) express a significant

sense in which moral responsibility is essentially interpersonal. Moral responsibility is, according to Responsibility-2 and Responsibility-3, essentially interpersonal, because they imply that one would not have sufficiently elucidated what moral responsibility is without an appeal to the reactive attitudes. Because the notion of a reactive attitude is strongly linked to relationships and communication between people (De Mesel 2017: 316), and ‘interpersonal’ means ‘relating to relationships/communication between people’ (De Mesel 2017: 315), Responsibility-2 and Responsibility-3 imply that moral responsibility is interpersonal in the commonly accepted sense of the term. I am content to leave you to judge how significant that sense is.

Acknowledgements This research was supported by the Research Foundation—Flanders (FWO). I am grateful to Michael Zimmerman and Michael McKenna for their comments.

Funding Funding was provided by the Research Foundation - Flanders (FWO).

References

- De Mesel, Benjamin. 2017. Is moral responsibility essentially interpersonal? A reply to Zimmerman. *The Journal of Ethics* 21: 309–333.
- Hacker, P.M.S., ed. 2001. On Strawson’s rehabilitation of metaphysics. In *Wittgenstein. Connections and controversies*, 345–370. Oxford: Clarendon.
- Maher, Chauncey. 2010. On being and holding responsible. *Philosophical Explorations* 13: 129–140.
- Manata, Paul. 2016. The conversational theory of moral responsibility and McKenna’s interdependence thesis. *The Philosophical Quarterly* 67: 61–83.
- McKenna, Michael. 2012. *Conversation and responsibility*. Oxford: Oxford University Press.
- Meadows, Donella H. 2008. *Thinking in systems. A primer*. White River Junction: Chelsea Green Publishing.
- Shabo, Seth. 2012. Where love and resentment meet. Strawson’s intrapersonal defense of compatibilism. *The Philosophical Review* 121: 95–124.
- Strawson, P.F. 1985. *Scepticism and naturalism. Some varieties*. New York: Routledge.
- Strawson, P.F. 1992. *Analysis and metaphysics. An introduction to philosophy*. Oxford: Oxford University Press.
- Strawson, P.F., ed. 2011. Review of Paul Grice, *Studies in the ways of words*. In *Philosophical writings*, 178–182. Oxford: Oxford University Press.
- Vargas, Manuel. 2013. *Building better beings. A theory of moral responsibility*. Oxford: Oxford University Press.
- Walton, Douglas. 2006. Epistemic and dialectical models of begging the question. *Synthese* 152: 237–284.
- Wiggins, David., ed. 1998a. Truth, and truth as predicated of moral judgments. In *Needs, values, truth*, 139–184. Oxford: Clarendon.
- Wiggins, David., ed. 1998b. A sensible subjectivism? In *Needs, values, truth*, 185–211. Oxford: Clarendon.
- Woodward, James. 2003. *Making things happen. A theory of causal explanation*. Oxford: Oxford University Press.
- Zimmerman, Michael J. 2016. Moral responsibility and the moral community. Is moral responsibility essentially interpersonal? *The Journal of Ethics* 20: 247–263.
- Zimmerman, Michael J. 2017. Strawson or straw man? More on moral responsibility and the moral community. *The Journal of Ethics* 21: 251–262.