**Is Moral Responsibility Essentially Interpersonal? A Reply to Zimmerman**

1. **Introduction**

Michael J. Zimmerman (2016) argues that no interpretation of the idea that moral responsibility is essentially interpersonal captures a significant truth. He raises several worries about two popular views that imply the idea that moral responsibility is essentially interpersonal: (1) the view that moral responsibility consists in answerability and (2) the Strawsonian view that moral responsibility consists in susceptibility to the reactive attitudes.

Zimmerman does not argue directly against the idea that moral responsibility is essentially interpersonal; he only claims that two popular views that imply the idea are unconvincing, thus leaving open the possibility that some other, more convincing version of the idea can be developed. I do not think that a new version of the idea is needed. Although I agree, by and large, with Zimmerman’s points against the view that moral responsibility consists in answerability, I will argue that there are ways for Strawsonians to respond to his points against the view that moral responsibility consists in susceptibility to the reactive attitudes, and to his claim that the Strawsonian view at best supports only an etiolated interpretation of the idea that moral responsibility is essentially interpersonal .

Zimmerman’s initial presentation of the Strawsonian view is as follows:

*Responsibility-1:*

*P* is morally responsible for *x* = 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.*[[1]](#footnote-1)

Zimmerman admits that this view does not face any of the problems that make the answerability view unattractive. However, it has its own problems. Zimmerman mentions four problems for *Responsibility-1*. I agree that three of them (the fourth is discussed in section 2 of this paper) really are problems for *Responsibility-1,* fatal problems even, and that they force us to adapt *Responsibility-1* as follows, as Zimmerman suggests:

*Responsibility-2:*

*P* is 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 reactive attitude toward *P* in respect of *x.*

*Responsibility-2* solves several problems with *Responsibility-1.* However, Zimmerman thinks that *Responsibility-2,* in turn, faces three problems. I intend to show thatthere are ways for Strawsonians to respond to these problems*.* The first problem, posed by the existence of self-reactive attitudes, will be discussed in section two. Section three deals with Zimmerman’s second problem, the problem from significance: does the acceptance of *Responsibility-2* give us reason to say that moral responsibility is essentially interpersonal in a significant way? In section four, I discuss Zimmerman’s suggestion that *Responsibility-2* puts forward the wrong kind of conceptual priority relation between ‘*P* is morally responsible’ and ‘it is appropriate to adopt some reactive attitude toward *P*’. I conclude in section five that Strawsonians can respond to all three problems for *Responsibility-2* raised by Zimmerman, and that the Strawsonian view can support a significant interpretation of the idea that moral responsibility is essentially interpersonal.

1. **Self-Reactive Attitudes**

The first problem for *Responsibility-2* concerns the self-reactive attitudes. This problem is a problem for *Responsibility-1* as well as for *Responsibility-2.* The problem is that the existence of self-reactive attitudes, such as remorse or guilt, seems to be in tension with the idea that moral responsibility is essentially *inter*personal. A self-reactive attitude is an attitude that I take towards myself. In light of the existence of self-reactive attitudes, then, the stipulation in *Responsibility-1* and *Responsibility-2* that *Q* be distinct from *P,* that the one who adopts a reactive attitude is distinct from the one towards whom it is adopted,seems ill advised. But if the stipulation is dropped, the *inter*personal nature of moral responsibility vanishes with it, because our definition of moral responsibility would then only need to include one person, *P.*

Strangely enough, Zimmerman himself suggests a perfectly adequate response to this objection:

A more promising response, I would think, is this: if ever some self-reactive attitude is appropriate, then so too, in principle, is some other-reactive attitude. True, only *P* can feel genuine remorse for what *P* has done, but it may be perfectly appropriate for someone else, *Q,* to adopt some related attitude toward *P* in light of what *P* has done. (2016: 258-259)

Zimmerman does not give any indication as to why this response to the problem would be inadequate.[[2]](#footnote-2) If it is not, however, his objection based on the existence of self-reactive attitudes disappears.

There is much to say for the solution that Zimmerman proposes. Strawson (2009: 161-162) repeatedly emphasizes that his three types of reactive attitudes are closely, both logically and humanly, connected. The human connection consists in the fact that ‘in general, though within varying limits, we demand of others, as well as of ourselves for others, something of the regard which we demand of others for ourselves’ (2009: 162). According to Strawson, we can barely imagine a case of one or two of the three types being fully developed, without being accompanied by some trace of the remaining two or one. Personal reactive attitudes, such as resentment, have moral or vicarious *analogues* or *correlates* (Strawson’s terms!): if it would be appropriate for *Q*, a person distinct from *P* who satisfies certain conditions *C,* to adopt the personal reactive attitude of resentment toward *P* in respect of *x* (something that *P* did to *Q,* e.g. *P*’s having lied to *Q*), then, if there were a person *R* distinct from *Q* such that *R* satisfied certain conditions *C,* it would be appropriate for *R* to adopt the moral reactive attitude of indignation toward *P* in respect of *x.* If it would be appropriate for *Q* to adopt the personal reactive attitude of resentment toward *P* because *P* lied to *Q,* then it would be appropriate for *R* to adopt the moral reactive attitude of indignation toward *P* because *P* lied to *Q,* and *vice versa.* Similarly, if it would be appropriate for *Q* to adopt the personal reactive attitude of resentment toward *P* because *P* lied to *Q,* then it would be appropriate for *P* to adopt the self-reactive attitude of guilt or remorse toward him- or herself because he lied to *Q,* and *vice versa.*

It may be thought that the close connection between the three types of reactive attitudes, as I have described it, and as it is in Strawson, is not even necessary to refute the argument from self-reactive attitudes. The self-reactive attitudes *themselves,* one might say, apart from their connection to the other types of reactive attitudes, already possess an essentially interpersonal element. They could be (and have been) described as attitudes towards our own attitudes towards others in which the others’ demand for or expectation of goodwill or regard is acknowledged. An example: I feel guilty for my contemptuous attitude towards you, and my feeling guilty is a way of acknowledging your demand for or expectation of goodwill. It may seem to follow from this description that an*other*’s demand or expectation figures essentially in a description of what self-reactive attitudes are, and that would be a sufficient reason to say that they are essentially interpersonal.

This description of the self-reactive attitudes, however, does not take into account that we may feel guilty for destroying the environment, for example, or for beating a dog, that is, for behavior that does not affect other *persons*. The same problem arises with respect to certain characterizations of the moral reactive attitudes. A characterization of the moral reactive attitudes as attitudes towards attitudes of others towards third *persons* leaves no room for saying that we may be indignant at someone’s attitude towards a dog or towards the environment. Strawson does not consider such cases. However, it seems rather uncontroversial, and it is admitted by authors sympathetic to Strawson (Bennett 1980: 45; Shabo 2012: 104), that we may be indignant at someone’s attitude towards a dog or towards the environment. In response to this problem, Strawsonians have suggested 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. 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 (Bennett 1980: 45; Shabo 2012: 104).

Zimmerman discusses what he calls ‘wrongdoing that appears not to involve other persons’ (2016: 255, section 7.1 of his paper), wrongdoing towards a dog or towards the environment, for example. He correctly points out that wrongdoing that appears not to involve other persons constitutes a problem for the answerability view of moral responsibility. He explicitly admits, however, that it does not constitute a problem for the Strawsonian view (2016: 257). The reason why it does not is that ‘even if non-human animals and the environment are not members of any moral community, it may well be appropriate for others to reproach me for mistreating them – or to praise me for treating them well’. Or, in other words, neither *Responsibility-1* nor *Responsibility-2* implies that one can only be morally responsible for behavior that affects otherpersons;there is no requirement that the *x* in both definitions be construed as some kind of behavior towards other persons. Thus, I suspect that Zimmerman could agree with the way in which Strawsonians have adapted the descriptions of self-reactive and moral reactive attitudes in order to be able to account for cases of wrongdoing towards animals or towards the environment.

I suggest, therefore, that although an adequate account of the self-reactive attitudes should describe them in terms of attitudes towards our own attitudes, it should not require that these attitudes be attitudes towards our own attitudes towards other *persons*. It is in the first place because of their essential connection to the other (essentially interpersonal) types of reactive attitude, rather than because of their being attitudes towards our own attitudes towards others, that the self-reactive attitudes cannot speak against the interpersonal nature of moral responsibility. That means that an example in which *P* beat his dog or harmed the environment could do the same work as my example in which *P* lied to *Q.* If it would be appropriate for *R* to adopt the moral reactive attitude of indignation toward *P* in respect of *P*’s behavior towards a dog or towards the environment, then it would be appropriate for *P* to adopt the self-reactive attitude of guilt or remorse because of what he did, and *vice versa.* I can see no reason why this example would be more problematic, for Zimmerman or for Strawsonians, than the lying example, precisely because Zimmerman admits that wrongdoing that appears not to involve other persons does not constitute a problem for the Strawsonian view.

An adequate Strawsonian answer to the problem of self-reactive attitudes is, in short, that it would not be *wrong* to say that *P* is morally responsible for *P* if and only if it is the case that, if *P* satisfied certain conditions *C,* it would be appropriate for *P* to adopt some self-reactive attitude toward him- or herselfin respect of *x*. However, because the self-reactive attitudes are essentially connected to the other kinds of reactive attitude, and because these kinds of reactive attitude are essentially interpersonal (something that is not in dispute here), it cannot be maintained that the existence of self-reactive attitudes threatens the essentially interpersonal nature of moral responsibility. At most, the existence of self-reactive attitudes is an invitation to unpack *Responsibility-2* as follows:

*P* is 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 (personal or moral) attitude toward *P* in respect of *x, and it would be appropriate for P, if (s)he satisfied conditions C, to adopt some self-reactive attitude toward him-/herself in respect of* x*.*

This is nothing more than *Responsibility-2* made explicit, and it shows that, even if the existence of self-reactive attitudes is explicitly recognized in the definition of moral responsibility, a reference to some person *Q* distinct from *P* remains necessary. As long as such a reference is necessary, the interpersonal nature of moral responsibility is accounted for. Thus, a definition of moral responsibility in which the self-reactive attitudes figure explicitly can imply that moral responsibility is essentially interpersonal.

1. **The Significance Problem**

The second problem for *Responsibility-2* concerns the question of how significant the interpersonal nature of moral responsibility is, according to *Responsibility-2.* Zimmerman asks us to consider the fitting-attitude analysis of value:

*Value-1:*

*x* is good [neutral, bad] = df. If there were a person *P* such that *P* satisfied certain conditions *C,* it would be appropriate for *P* to favor [be indifferent toward, disfavor] *x.*

Zimmerman’s question is:

If *Value-1* were true, would that give us any reason to say that value is essentially *personal* in any significant way? I cannot see that it would. After all, *Value-1* is compatible with things having value even if no person ever existed. So, too, Responsibility-2 is compatible with someone’s being responsible for something even if no other person ever existed. (2016: 261)

I will make two points in response to Zimmerman’s argument. First, the main problem with the argument from *Value-1* is that it presupposes an analogy between the meanings of ‘personal’ and ‘interpersonal’. The supposed analogy is the following: if we cannot conclude from the fact that only one person *P* necessarily figures in a definition of something (value, for example) that that something is personal, then we cannot conclude from the fact that *P* and *Q* necessarily figure in a definition of something that that something is interpersonal.

It is true, I believe, that the fact that only *P* figures in a definition of something should not lead us to say that that something is essentially personal. Consider, for example, a popular view of color:

*Color-1:*

*x* is red [green, blue] = df. If there were a person *P* such that *P* satisfied certain conditions *C, x* would in some lighting conditions produce the sensation of redness [greenness, blueness] in *P.*[[3]](#footnote-3)

If this, or something similar, is an adequate definition of color, then color cannot be defined without reference to a person *P.* However, it seems odd to say that color is essentially personal in a significant way. Similarly, *Value-1* does not seem to give us a reason to say that value is essentially personal. The occurrence of a person *P* in *Color-1* and *Value-1* is the reason why these definitions are often characterized as *response-dependence* accounts of color and value, respectively. They are not called *personal* accounts of color and value, because that description would suggest what the definitions clearly do not imply: saying that color and value are personal could easily be taken to suggest, for example, that whether something is red or good is a matter of personal preference or taste.

At this point, questions about the meaning and use of ‘personal’ arise. The most common uses of ‘personal’, according to my *Oxford Dictionary of English,* are:

1. belonging to or affecting a particular person rather than anyone else: *her personal fortune was recently estimated at £37 million.*
2. of or concerning one’s private life, relationships, and emotions rather than one’s career or public life: *the book describes his sporting career and gives little information about his personal life.*
3. relating to a person’s body: *personal hygiene.*

If these are indeed the most common uses of ‘personal’, it can readily be seen why it would be misleading to say that color and/or value are essentially personal. The reason why it would be misleading is that the occurrence of exactly one person in the definition of something is not the criterion (or at least not a sufficient criterion) by which we commonly distinguish what is personal from what is not.

Let us now have look at the meaning of ‘interpersonal’. This is how the dictionary explains it: ‘relating to relationships or communication between people: *you will need good interpersonal skills*’. Although this definition confirms that multiple persons should be involved, it makes perfectly clear that the difference between ‘personal’ and ‘interpersonal’ is not just the difference between ‘one person involved’ and ‘multiple persons involved’; the move from ‘personal’ to ‘interpersonal’ is not a move from ‘one person’ to ‘more than one person’. We cannot substitute ‘two or more persons’ for ‘one person’ in the definitions of ‘personal’ mentioned above in order to get a definition of ‘interpersonal’. ‘Interpersonal’ does *not* mean ‘belonging to or affecting two or more persons rather than some others’, or ‘concerning two or more persons’ private lives, relationships and emotions’, or ‘relating to two or more persons’ bodies’. Conversely, ‘personal’ does not mean ‘relating to relationships or communication with oneself’. The concepts ‘personal’ and ‘interpersonal’ are not just quantitatively different concepts, as Zimmerman assumes (or must assume in order for his objection to *Responsibility-2* to work), they are qualitatively different. This disanalogy between ‘personal’ and ‘interpersonal’ means that the fact that we would not say that color and value are essentially personal if we accept *Color-1* and *Value-1* respectively, does not provide us with a reason to doubt that moral responsibility is essentially interpersonal if we accept *Responsibility-2.*

Whether responsibility is essentially interpersonal if we accept *Responsibility-2* can be established only by looking at the definition of ‘interpersonal’. Apart from the fact that multiple persons should be involved, what is key in this definition is the reference to *relationships* or *communication* between people (an aspect that is entirely absent, of course, in the definitions of ‘personal’). The question whether moral responsibility is essentially interpersonal if we accept *Responsibility-2* is the question whether *Responsibility-2* provides an essential link between moral responsibility and relationships or communication between people (rather than the question whether multiple persons are necessarily involved). If *that* is the question, I cannot see how one could escape the conclusion that *Responsibility-2* presents moral responsibility as essentially interpersonal. Although the words ‘communication’ and ‘relationship’ do not occur in the definition, the notion of a reactive attitude does the requisite work.

Strawson is very explicit about the link between reactive attitudes and relationships between people. He writes, for example, that ‘being involved in inter-personal relationships as we normally understand them precisely is being exposed to the range of reactive attitudes and feelings that is in question’ (2009: 158). Although the identification between involvement in interpersonal relationships and susceptibility to the reactive attitudes might go too far (in light of the view that moral responsibility consists in susceptibility to the reactive attitudes, it seems to lead to the idea that being involved in interpersonal relationships *is* being morally responsible and *vice versa*), it is clear that no adequate reading of Strawson and his followers could lead one to deny that the reactive attitudes are essentially interpersonal in the non-trivial sense specified by the dictionary definition.

In short, Zimmerman rightly observes that we would not say that value is essentially personal*.* But because the difference between personal and interpersonal is not just a difference between ‘one person involved’ and ‘multiple persons involved’, the point about value has no bearing on the question whether moral responsibility is essentially interpersonal. That question can only be decided by looking at what ‘interpersonal’ means. A closer look at its meaning reveals that *Responsibility-2* represents moral responsibility as interpersonal, in the commonly accepted and non-trivial sense of that term. That is my first point in response to Zimmerman’s significance argument.

The second point is the following. According to Zimmerman (see the quotation at the beginning of this section), the truth of *Value-1* would not give us a reason to say that value is essentially personal, because *Value-1* is compatible with things having value even if no person ever existed. Similarly, the truth of *Responsibility-2* would not give us a reason to say that moral responsibility is essentially interpersonal, because *Responsibility-2* is compatible with someone’s being responsible for something even if no other person ever existed.

A difficulty with Zimmerman’s argument is, again, that it presupposes an analogy where there may in fact be a disanalogy. It is clearly possible to imagine that no person ever existed or, in other words, that there is a possible world in which no person ever existed. Had evolution taken a different turn, there might not have been any persons. The fact that *Value-1* is compatible with the possibility that no person ever existed casts doubts on the claim that, according to *Value-1,* value is essentially personal. But is it possible to imagine that one person exists while no other person ever existed?[[4]](#footnote-4) Is there a possible world in which that is the case?

The answer to these questions crucially hinges on what one takes a person to be. Although this is not the place to discuss and evaluate different conceptions of personhood (for a good overview, see Korfmacher 2006), one fundamental distinction should be noted. Some philosophers think that all persons are human beings, for example because they hold that the spatiotemporal continuity of a functioning *human* body constitutes personal identity. Others think that non-human animals could be persons too, provided that they have or could acquire certain kinds of abilities. Still others have a conception of personhood according to which Martians or computers could be persons.

If one’s conception of personhood is such that only human beings, or only human beings and certain kinds of animals, could be persons, it becomes doubtful whether the possibility that one person exists while no other person ever existed is a coherent idea.[[5]](#footnote-5) It seems essential to (human and non-human) animals that they are born of other animals. Moreover, what Zimmerman needs is not only the possibility that one person exists while no other person ever existed, he also needs the possibility of such a person being a morally responsible agent. He writes about Robinson Crusoe that ‘it is conceivable that he never was a member of any community and yet somehow developed the capacities necessary for being a morally responsible agent’ (2016: 260), and cites empirical research in support of that point:

Some recent research (e.g., Sloane et al. 2012; Smetana et al. 2013; Wynn and Bloom 2013) indicates that human beings may have an innate moral sense, which in turn suggests that they may be capable of developing the relevant capacities in isolation from other people, at least to some extent. (2016: 260).

I do not think that the empirical research that Zimmerman refers to can support his point. Some of the research indeed mentions the *possibility* that human beings have an innate moral sense, but this possibility is mentioned as one among *other* possible ways of explaining the experimental results (see Sloane et al. 2012: 203). And even if some of the research suggests that human beings have an innate moral sense, it does *not* suggest that they may be capable of *developing* this innate moral sense into the capacities that are relevant for moral responsibility in isolation from other people. Smetana’s social (!) domain approach, for example, explained in the paper that Zimmerman refers to, ‘views moral understanding as *constructed through social interactions* and building on predispositions toward empathy and a concern for others that are evident in infancy’ (Smetana et al. 2013: 38; my italics).

The empirical research does not refute Zimmerman’s point: after all, Zimmerman is concerned with conceivability, and what is actually the case cannot show that what is not the case is inconceivable. What I want to emphasize here is, rather, that some of the research is neutral on the question whether human beings can develop the capacities necessary for being a morally responsible agent in isolation (Sloane et al. 2012; Wynn and Bloom 2013), and that some of it explicitly puts forward a negative answer (Smetana et al. 2013). That means that, if the empirical research referred to by Zimmerman supports any answer to the question, it is the answer that human beings cannot develop the capacities for being a morally responsible agent in isolation from other human beings.

If only human beings, or only human beings and certain kinds of animals, could be persons, the idea of a morally responsible person who never was a member of any community is not as straightforwardly consistent as Zimmerman assumes. A way for Strawsonians to avoid Zimmerman’s objection, then, or at least a way for them to shed serious doubt on its coherence, is to adopt a certain conception of personhood. Even though such conceptions are not uncommon, as Korfmacher’s (2006) overview shows, this response may seem *ad hoc.* It is not, however, because the conception of personhood that Strawsonians need is also *Strawson’s* conception, as it appears in ‘Freedom and Resentment’ (2009 [1962]) and is worked out in his book *Individuals* (1959). In ‘Freedom and Resentment’, Strawson mostly uses ‘human being’ instead of ‘person’. He describes the reactive attitudes as ‘essentially natural human reactions’ (2009: 156), ‘given with the fact of human society’ (2009: 169). He writes about ‘the objective attitude to another human being’ (2009: 155) and claims that a sustained objectivity of attitude ‘does not seem to be something of which human beings would be capable’ (2009: 158). He uses ‘human relationship’ and ‘interpersonal relationship’ interchangeably. In his reply to critics of ‘Freedom and Resentment’, he emphasizes that ‘we are concerned with human behaviour’ (1980: 264). In short, there is no indication in ‘Freedom and Resentment’ that non-human beings could be persons, and the same can be said of *Individuals.*

It could be remarked that ‘no indication that non-human beings could be persons’ implies that Strawson does not explicitly exclude that possibility. That is true, but it does not mean that the possibility has to be taken seriously or that it should have any influence on our conception of personhood (an influence that Strawson clearly does not allow it to have). In ‘Strawson’s Concept of a Person’ (2002), Peter Hacker writes that, according to Strawson, the category of persons ‘includes only sentient, language-using creatures and their offspring’. He adds, in a Strawsonian vein:

Whether they [persons] must be biological creatures may be debated, but need not. Our concept of a person is tailored for the circumstances of our lives and the objects we encounter in them. Whether we should call ‘persons’ also manufactured beings with capacities and forms of behaviour akin to ours is a question which has not arisen, and, if it were to arise, would require a decision, not a discovery. (2002: 28)

In a later discussion of personhood, influenced by Strawson, Hacker (2010: 285-316) explains the matter in some detail. Because his conclusion is relevant for our purposes, I will quote it at some length:

Human beings are the only persons we know or are ever likely to know. Our concept of a person evolved above all as a concept applicable to human beings – social beings who are members of a moral community. […]

Although we know of no persons other than human beings, usage does not restrict the application of this categorial term to human beings alone. If there be other creatures possessing the appropriate range of language-dependent rational powers grafted on to an appropriate animal nature, then they too are persons. To be a person is not to be a certain kind of animal, but rather to be an animal of one kind or another with certain kinds of abilities. (This conception is sometimes characterized as ‘animalism’, but misleadingly so. The identity of a person depends upon the *kind* of animal that the person in question is, for ‘person’ is a qualification on a substance noun. We are, and the only persons we are ever likely to know are, *human persons.* Our identity as persons turns on our identity as *individuals* of the kind *human being.* If an ‘ism’ has to be bestowed, the appropriate and wholly unsurprising one is ‘humanism’.) The nature of a person is rooted in animality, but transformed by possession of intellect and will. […]

It is a moot point whether the idea of mechanical, artefactual persons is intelligible. Science fiction is replete with androids. But not everything that is, in this sense, imaginable, is logically possible. The issue turns not on artefactuality, but on biology. If advanced kinds of life can be artificially made, then, in principle there is nothing *logically* awry with the thought of manufactured animals with the necessary endowment to be or become persons. But the idea of androids is far more problematic. Such imaginary beings are not merely manufactured, they are machines. So they presumably do not grow, or go through the phases of life – knowing no childhood, youth, maturity or old age. […] Since they do not reproduce, they presumably have no sexual character or drive; hence too they neither lust nor enjoy sexual intercourse. […] Lacking parents and bereft of procreative drives and powers, do they have a capacity for love? Can androids feel passions at all? In what sense, if any, are they by nature social creatures, belonging to a moral community? That depends on their author’s tale and its coherence – which is rarely adequately elaborated. […] We stray here far beyond the bounds of application of our concept of a person. It is patent that it matters little what we say, since the rules for the use of the word ‘person’ do not extend to such cases. If any such cases were to arise, we should need to modify the rules in the light of logical, practical and ethical considerations. But they do not, and we need not. (2010: 310-314)

This passage, to be sure, does not contain a full argument in favor of a humanist conception of personhood; such an argument can be found in Hacker’s text. It does contain, however, some elements that are crucial to understand the Strawsonian view on the matter. First, manufactured animals could, logically speaking, be persons, provided that they share with human beings an ‘appropriate animal nature’, which includes, among other things, the capacities to engage socially with others and to reproduce. If these capacities are required, the possibility that one person exists while no other person ever existed becomes, again, difficult if not impossible to imagine. It should be taken into account, moreover, that the idea of a manufactured animal is logically connected to the idea of a *manufacturer,* that is, that manufactured animals could not possibly exist if no manufacturer ever existed.

Secondly, there is the idea of androids. According to a Strawsonian conception of personhood, they could not be persons if they lacked an appropriate animal nature. It is not clear whether the idea of an android is compatible with the idea of having an animal nature, let alone with the idea of a person. Our concept of a person, tailored for the beings we encounter in our lives, simply does not extend to such cases. If we were to encounter an android with many of the capacities relevant for personhood, we would have to *decide* whether or not to *modify* our concept of personhood in such a way as to include the android. It is not inconceivable that this would happen, and it is perfectly normal for concepts to evolve. The *present* concept of a person, however, is not made for application to such cases. Their conceivability matters little, if at all, and it should not influence our definitions of ‘person’ and, consequently, of ‘moral responsibility’.

One might insist at this point that, although some conceptions of personhood exclude the conceivability of a non-human person, the Strawsonian conception does not rule it out entirely: it still allows for the conceivability of a non-human person (and adds that this conceivability matters little), and that is enough for Zimmerman’s objection to retain its force. After all, Zimmerman is concerned with ‘bare’ conceptual possibility only, not with the question whether this conceptual possibility matters. But why is he concerned with conceptual possibility? He starts his paper by introducing what he calls the Slogan: ‘There can be no moral responsibility without a moral community’ (2016: 248). The Slogan, he says, captures the spirit of the Strawsonian idea that moral responsibility is, in some important way, essentially interpersonal. The first question to ask about the Slogan is how ‘can’ is to be interpreted, what kind of possibility the Slogan expresses, because that is the kind of possibility with which Zimmerman wants to deal. Zimmerman’s answer is that the Slogan ‘expresses *conceptual* possibility. It claims that it is *inconceivable* that there should be moral responsibility in the absence of a moral community’ (2016: 248). Zimmerman adds that Strawson (2009: 167) states this explicitly.

A closer look at Strawson’s text reveals that matters are more complicated than Zimmerman presents them to be. The word ‘inconceivable’ occurs twice in Strawson text, not in the passage that Zimmerman refers to, but in the following one:

And our question reduces to this: could, or should the acceptance of the determinist thesis lead us always to look on everyone in this [objective] way? […] It does not seem to be self-contradictory that this might happen. So I suppose we must say that it is not absolutely inconceivable that this might happen. But I am strongly inclined to think that it is, for us as we are, practically inconceivable. The human commitment to participation in ordinary interpersonal relationships is, I think, too thoroughgoing and deeply rooted for us to take seriously the thought that a general theoretical conviction might so change our world that, in it, there were no longer any such things as inter-personal relationships as we normally understand them […]. (2009: 157-158)

Strawson suggests in this passage that the disappearance of interpersonal relationships, though not absolutely inconceivable, is practically inconceivable for us as we are. This practical inconceivability is enough for him to say that the thought need not be taken seriously. ‘Bare’ conceptual possibility, not rooted in life as it is ‘for us as we are’, or as it normally is, does not interest him (consider, in this respect, his frequent use of ‘normal(ly)’ and ‘ordinary’/‘ordinarily’). That point is confirmed by some other passages. First, he describes the reactive attitudes as ‘an essential part of the moral life as we know it’ (2009: 168), but refuses to claim that they are ‘essential features of the concept of morality in general’ (2009: 170). Secondly, he writes the following about the reactive attitudes:

One who manifested the personal reactive attitudes in a high degree but showed no inclination at all to their vicarious analogues would appear as an abnormal case of moral egocentricity, as a kind of moral solipsist. Let him be supposed fully to acknowledge the claims to regard that others had on him, to be susceptible of the whole range of self-reactive attitudes. He would then see himself as unique both as one (*the* one) who had a general claim on human regard and as one (*the* one) on whom human beings in general had such a claim. This would be a kind of moral solipsism. But it is*barely more than a conceptual possibility* [my italics]; if it is that. In general, though within varying limits, we demand of others for others, as well as of ourselves for others, something of the regard which we demand of others for ourselves. Can we imagine, besides that of the moral solipsist, any other case of one or two of these three types of attitude being fully developed, but quite unaccompanied by any trace, however slight, of the remaining two or one? If we can, then we imagine something far below or far above the level of our common humanity – a moral idiot or a saint. (2009: 161-162)

The idea of someone who manifested the personal reactive attitudes in a high degree but showed no inclination to their analogues is not excluded as absolutely inconceivable or conceptually impossible, but the possibility at issue here is ‘barely more than a conceptual possibility’. *Strawson does not allow such a bare conceptual possibility to have any influence on his thesis*, and the bare conceptual possibility does not prevent him from claiming, as we have seen, that the three kinds of reactive attitude are *logically* connected (2009: 161).

In light of these passages from Strawson’s text, we can ask Zimmerman’s question again: how is ‘can’ in the Slogan to be interpreted, what kind of possibility does it express? Zimmerman is confronted with a dilemma. If he takes the first horn and attributes the Slogan to Strawson, it should be clear that ‘bare conceptual possibility’ or ‘absolute conceivability’ cannot be the answer. Rather, the answer would be something like ‘practical conceivability’, ‘possibility for us as we are’ or ‘possibility in life as we know it’. The possibility that one person exists while no other person ever existed is, especially according to a Strawsonian conception of personhood, barely more than a conceptual possibility, if it is that. If Zimmerman takes the second horn of the dilemma, he can maintain that the Slogan expresses bare conceptual possibility or absolute conceivability, but then the Slogan can no longer be attributed to Strawson. (A Strawsonian reading of the Slogan ‘There can be no moral responsibility without a moral community’ would be ‘There can be no moral responsibility, in life as we know it, without a moral community’.)

A possible objection will help to make my argument more explicit. Could it not be objected here that, if Strawson admits that it is conceivable (even if only absolutely and not practically) that one morally responsible person exists while no other person ever existed, Zimmerman’s point is confirmed?[[6]](#footnote-6)

But what exactly *is* Zimmerman’s point? Zimmerman wants to question ‘how *significant* the interpersonal nature of moral responsibility is, according to *Responsibility-2*’ (2016: 261). His argument (which is more complicated than it may seem) can be reconstructed as follows:

1. Strawsonians (and some others) think that moral responsibility is essentially interpersonal.
2. The view that moral responsibility is essentially interpersonal, endorsed by Strawsonians (and some others), is expressed by the Slogan: ‘There can be no moral responsibility without a moral community.’
3. The Slogan, as endorsed by Strawsonians and others, expresses conceptual possibility.
4. *Responsibility-2* expresses the Strawsonian view and implies the Slogan.
5. *Responsibility-2* is compatible with the possibility of someone’s being responsible for something even if no other person ever existed.
6. The Slogan is incompatible with the possibility of someone’s being responsible for something even if no other person ever existed.
7. It follows from (5) and (6) that *Responsibility-2* is compatible with a possibility that is incompatible with the Slogan.
8. Therefore, *Responsibility-2* cannot imply a significant version of the Slogan.
9. Because the Slogan expresses the view that moral responsibility is essentially interpersonal (2), *Responsibility-2* cannot imply a significant version of the view that moral responsibility is essentially interpersonal.

I have started my discussion of Zimmerman’s argument by focusing on (5): the coherence of the possibility of someone’s being responsible for something even if no other person ever existed can be doubted. However, I admitted that the possibility cannot be ruled out entirely. If (5) is Zimmerman’s point, there is a sense (the absolute conceivability sense) in which Strawsonians can agree with it. But the truth of (5) is not enough for Zimmerman to make his argument work. I propose, therefore, to have a look at (3). As we have seen, Zimmerman *wants the Slogan to capture the spirit of the Strawsonian view* that moral responsibility is essentially interpersonal. But how can the Slogan be endorsed by Strawsonians if Strawson *admits* that it is (absolutely) conceivable that someone is responsible for something even if no other person ever existed? The Slogan can *only* be taken to express the Strawsonian view if the possibility expressed in the Slogan is *not* absolute conceivability. If the Slogan would express absolute conceivability, and Strawson admits that one responsible person could exist while no other person ever existed, he would undermine his own Slogan. An alternative reading of the Slogan is suggested by Strawson himself: it does not express absolute conceivability, but practical conceivability. *The Slogan can only be an adequate expression of the Strawsonian view if it is taken to express practical conceivability.*

What happens with Zimmerman’s argument in light of the distinction between absolute and practical conceivability? (3) will have to be read as follows:

(3.1) The Slogan, as endorsed by Strawsonians, expresses practical conceivability.

If (3) has to be read as (3.1), (6) becomes false. Instead, we get

(6.1) The Slogan is compatible with the absolute conceivability of someone’s being responsible for something even if no other person ever existed.

Now (7) no longer follows. The link between *Responsibility-2* and the Slogan is not broken and, *contra* (8), *Responsibility-2* still implies a significant version of the Slogan. In a sense, Zimmerman’s point has been confirmed: according to (5), *Responsibility-2* is indeed compatible with the possibility that Zimmerman highlights. In another, more important sense, however, Zimmerman’s point loses its force, because it no longer helps to establish Zimmerman’s conclusion (9).

In short, the problem is as follows. There are two possibilities. First, the Slogan expresses practical conceivability. If it does, it is compatible with the possibility that Zimmerman mentions in (5), and the possibility cannot undermine the Slogan (which is Zimmerman’s aim). Secondly, one could insist that the Slogan expresses absolute conceivability, and that (3) is to be read as

(3.2) The Slogan, as endorsed by Strawsonians, expresses absolute conceivability.

If the Slogan expresses absolute conceivability, Zimmerman’s conclusion (9) follows, but then the Slogan can no longer be taken to express Strawson’s view. As we have seen, Strawson does not allow ‘bare conceptual possibility’ to have any influence on his views. In light of Strawson’s distinction between practical and absolute conceivability, (3.2) is false. Zimmerman cannot have the combination of ideas that he needs: (3) that the Slogan expresses the Strawsonian view, and (6) that the Slogan is incompatible with the possibility that he mentions in (5).

I conclude that Zimmerman has not offered a powerful reason for doubting the significance of the interpersonal nature of moral responsibility, as specified in *Responsibility-2*. (1) The difference between ‘personal’ and ‘interpersonal’ is not just a difference between ‘one person involved’ and ‘multiple persons involved’. There is a significant, qualitative difference between our uses of ‘personal’ and ‘interpersonal’; they are not as analogous as Zimmerman takes them to be. That means that the force of his point about value (*Value-1* is compatible with things having value even if no person ever existed, so it would be strange to say that *Value-1* gives us reason to say that value is essentially personal) does not carry over to the point about responsibility (*Responsibility-2* is compatible with someone’s being responsible for something even if no other person ever existed, so it would be strange to say that *Responsibility-2* gives us reason to say that moral responsibility is essentially interpersonal). Whether moral responsibility is essentially interpersonal, according to *Responsibility-2,* has to be decided by looking at the definition of ‘interpersonal’. Because ‘interpersonal’ means ‘relating to relationships or communication between people’, and because *Responsibility-2* clearly relates moral responsibility to relationships or communication between people, *Responsibility-2* expresses, in a significant way, the interpersonal nature of moral responsibility.

(2) Strawsonians have resources to respond to Zimmerman’s observation that *Responsibility-2* is compatible with someone’s being responsible for something even if no other person ever existed, and to his claim that this compatibility casts doubt on the significance of the interpersonal nature of moral responsibility, according to *Responsibility-2*. First, the idea that one person exists while no other person ever existed may not be coherent. According to certain established conceptions of personhood, including the Strawsonian one, personhood requires an animal nature, and it may be essential to animals that they are born of other animals. If the idea that one person exists while no other person ever existed is not coherent, then neither is the idea of someone’s being responsible for something even if no other person ever existed, and Zimmerman’s objection loses its force. Secondly, our concept of a person is tailored for the beings we encounter in our lives. Even if there is a coherent possibility that one person exists while no other person ever existed, this possibility is what Strawson would call ‘barely more than a conceptual possibility’. It is absolutely conceivable, but not practically conceivable, because it is not rooted in ‘life as we know it’. *Responsibility-2* may be compatible with someone’s being responsible for something even if no other person ever existed, as Zimmerman claims. But Zimmerman suggests that, given such a compatibility, *Responsibility-2* no longer implies the Slogan and the idea expressed by it, namely that moral responsibility is, in a significant way, essentially interpersonal (the link between the Slogan and the idea that moral responsibility is essentially interpersonal is not in dispute here). The link between *Responsibility-2* and the Slogan would be broken, and *Responsibility-2* could no longer be taken to imply that moral responsibility is essentially interpersonal in a significant way. That conclusion only holds, however, if the Slogan is taken to express bare conceptual possibility. In that case, *Responsibility-2* would be compatible with a possibility that would be incompatible with the Slogan, and this result would indeed make it questionable how significant the interpersonal nature of moral responsibility is, according to *Responsibility-2.* The Strawsonian Slogan, however, expresses practical conceivability, which means that both *Responsibility-2* and the Slogan are compatible with someone’s being responsible for something even if no other person ever existed. Thus, *Responsibility-2* stillimplies the Slogan and the idea that moral responsibility is essentially interpersonal.

1. **We Adopt Reactive Attitudes *Because P* Is Responsible**

This is Zimmerman’s third problem with *Responsibility-2:*

Finally, just as some philosophers […] have argued that *Value-1* is to be rejected, on the grounds that, if it is ever appropriate to favor something *x,* that will be *because x* is good, so too it may be argued that *Responsibility-2* is to be rejected, on the grounds that, if it is ever appropriate for someone *Q* to adopt some reactive attitude toward someone *P* in respect of something *x,* that will be *because P* is responsible for *x.* (2016: 261)

Zimmerman’s problem can be spelled out as follows. According to him, it could be argued that there is a certain priority relation between ‘*P* is responsible’ and ‘it is appropriate to adopt some reactive attitude toward *P*’. If it is ever appropriate to adopt some reactive attitude toward *P,* that will be *because P* is responsible; and that means that ‘*P* is responsible’ is in some sense more fundamental or basic than ‘it is appropriate to adopt some reactive attitude toward *P*’. If indeed there is such a priority relation between ‘*P* is responsible’ and ‘it is appropriate to adopt some reactive attitude toward *P*’ or, in other words, if being responsible is prior to holding responsible, *Responsibility-2* is in trouble.

I will make two points about Zimmerman’s objection. The remarks are not meant to refute the objection or to prove anything about the relation between being responsible and holding responsible. Rather, they are meant to show that several aspects of the objection can be and have been challenged (something that Zimmerman, given his tentative formulation of the objection, is not committed to deny).

First, Zimmerman takes the ‘because’-statement to suggest that being responsible is prior to holding responsible. He does not comment on the kind of priority relation suggested by his objection, but some of his other articles are helpful in this regard: the ‘because’-statement is taken to suggest that *P*’s being responsible *grounds* and *explains* the appropriateness of adopting some reactive attitudetoward *P* (Zimmerman 2007: 348, 351; 2010: 108-110; 2015: 49)*.* The debate about the relation between being responsible and holding responsible is complicated, and I will not add anything to it here (for an overview, see Clarke, McKenna and Smith 2015: 6-7). It should be noted, however, that there is a lot of disagreement about the priority relation, even among Strawsonians, and that all parties in the debate recognize that it may often be appropriate to say that someone is *held* responsible because she *is* responsible. That is, disagreement about the priority relation persists in the face of Zimmerman’s observation.

Many Strawsonians believe that holding responsible is prior to being responsible. Patrick Todd (2016) calls this the ‘reversal’ thesis, and he has provided a helpful overview of how the thesis has been understood. He argues that, although the reversal thesis has not yet been developed in a satisfactory way, the case for a satisfactory development is anything but lost. Moreover, Todd shows that not all Strawsonians accept the reversal thesis. First, he remains neutral on the question whether any reversal thesis was intended by Strawson himself and claims that, ‘such a thesis is, at most, implicit in Strawson’s essay’ (Todd 2016: 210). Secondly, some prominent Strawsonians, such as R. Jay Wallace (1994), do not articulate any clear reversal thesis (Todd 2016: 231-234).

Todd’s judgment about Strawson is confirmed by Michael McKenna:

[…] the explanatory relation between holding and being morally responsible is, at best, mutually supporting. And there is good reason to think that, on Strawson’s own view, *if* one direction has a place of privilege, it is being morally responsible; being morally responsible explains holding morally responsible. (McKenna 2005: 172)

McKenna himself has advanced a so-called ‘interdependence thesis’. He claims that neither being responsible nor holding responsible is more basic. In the second chapter of his *Conversation and Responsibility,* McKenna argues to the conclusion that ‘Strawsonians should forgo a commitment to a single direction of explanatory priority, and tracking this, should also forgo any thesis that there is a single order of metaphysical priority […] there is an irreducible relation of interdependence between being and (pertinent norms pertaining to) holding responsible’ (2012: 53-54; for a defense of McKenna’s interdependence thesis in the light of recent literature on grounding, see Manata 2016).

These remarks on the relation between being responsible and holding responsible do not prove that Zimmerman’s objection does not succeed. However, they do seem to confront Zimmerman with a dilemma similar to the one discussed in the previous section. Zimmerman mentions Strawson, McKenna and Wallace as proponents of the Strawsonian view (2016: 248). On the first horn of the dilemma, *Responsibility-2* is understood as implying that holding responsible is prior to being responsible. This is the interpretation of *Responsibility-2* that Zimmerman needs in order for his objection to work: his suggestion that being responsible is prior to holding responsible only constitutes an objection to *Responsibility-2* if *Responsibility-2* is understood as implying the converse priority relation. According to this interpretation of *Responsibility-2,* however, it can no longer be claimed that *Responsibility-2* expresses the Strawsonian view, because prominent Strawsonians mentioned by Zimmerman do not subscribe to the reversal thesis. Zimmerman’s objection may then be an objection to *Responsibility-2,* but it is not an objection to the Slogan and loses its force against the Strawsonian claim, endorsed by Wallace and McKenna, that moral responsibility is essentially interpersonal. In view of the disagreement among Strawsonians regarding the priority relation between being and holding responsible, it would be better to express the Strawsonian view in such a way that it does not put forward a metaphysical or explanatory priority relation between the two. This is the second horn of the dilemma, on which *Responsibility-2* is understood as not implying such a priority relation at all. On this horn, however, Zimmerman’s objection no longer works: if *Responsibility-2* is compatible with the priority of being responsible, the suggestion that being responsible may be prior to holding responsible does not constitute an objection to it. In short: if *Responsibility-2* is taken to express the Strawsonian view, Zimmerman’s ‘because’-statement does not constitute an objection to it. If Zimmerman’s objection works, *Responsibility-2* can hardly be taken to express the Strawsonian view.

Thus, Zimmerman’s objection only works if *Responsibility-2* is taken to imply that holding responsible is prior. My first point is that Strawsonians, including Strawson himself, do not unequivocally accept this priority, which makes it doubtful whether *Responsibility-2* adequately expresses the Strawsonian view. Secondly, one can ask why *Responsibility-2* would imply that holding responsible is prior to being responsible. Why would the components of the *definiens* have to be prior to those of the *definiendum*? Zimmerman (2015: 49) contains an explicit statement of his methodological commitments: he wants to provide an analysis of the concept of responsibility that ‘involves a ‘deconstruction’ of a complex concept into simpler concepts’, an analysis that reveals what responsibility *consists in.* It is not difficult to see why, according to this kind of analysis, *Responsibility-2* is irreconcilable with the priority of being responsible: if being responsible is prior to holding responsible, *Responsibility-2,* as a definition of being responsible, does not provide an analysis in terms of concepts that are simpler or prior to it, and therefore it has to be rejected.

Do we have to read *Responsibility-2* as providing the kind of deconstructive analysis that Zimmerman favors? If *Responsibility-2* is to be interpreted as an expression of the Strawsonian view, this is at least doubtful. Zimmerman suggests that *Responsibility-2* can be read as an expression of the Strawsonian view because Strawsonians are concerned with conceptual analysis (2016: 248-249). The *kind* of conceptual analysis that they are concerned with, however, differs significantly from Zimmerman’s deconstructive analysis. Zimmerman refers to three Strawsonian authors in his discussion of conceptual analysis. In the passage by Stephen Darwall (2006: 11-12) that he refers to, I cannot discern a commitment to deconstructive conceptual analysis. The two other authors, Strawson and McKenna, are very clear about their commitment to *another* kind of analysis.

In *Analysis and Metaphysics,* Strawson contrasts two kinds of conceptual analysis. The first ‘may be thought of as a kind of breaking down or decomposing of something’ (1992: 2), ‘the resolution of something complex into elements and the exhibition of the ways in which the elements are related in the complex’, and the aim is ‘to get a clear grasp of complex meanings by reducing them, without remainder, to simple meanings’ (1992: 17-18). The second kind of conceptual analysis, favored by Strawson, is called the ‘connective model’ (1992: 21). The aim of the connective model is to explain concepts and describe conceptual connections by way of implication, presupposition and exclusion. Strawson remarks that it might be better to use the word ‘elucidation’ rather than ‘analysis’, because analysis ‘strongly suggests the dismantling model’, but he decides to stick with ‘analysis’ (1992: 20).

In a response to his critics, McKenna writes about his philosophical method that ‘it sometimes goes under the name *elucidation*’ (2013: 79). Elucidation, he claims, does not purport to show that one thing is metaphysically prior to another, and the aim is not to offer a reductive account of one concept in terms of another, more basic concept. He refers to his method as ‘soft-core conceptual geography’ and contrasts it with ‘hard-core conceptual analysis’. Another way to put the contrast is in terms of ‘same-level analysis’ versus ‘decompositional analysis’ (for an overview of both kinds of analysis and their relation to different philosophical traditions, see Glock 2009: 153-160 and 218).

I believe that, although not all Strawsonians are equally explicit about their methodological commitments, there is every reason to think that, generally speaking, Strawsonians prefer connective analysis to decompositional analysis and understand themselves as offering the former rather than the latter. They do not offer, and they do not aim to offer, the kind of decompositional definition of moral responsibility that Zimmerman is looking for. Again, there seems to be a dilemma for Zimmerman. His ‘because’-objection can only work if *Responsibility-2* is understood as offering a decompositional analysis of moral responsibility. If it is so understood, however, it does no longer express the Strawsonian view: it is not the way in which Strawsonians would express the Slogan or the idea that moral responsibility is essentially interpersonal. It would be better to express the Strawsonian view in a way that makes clear that the analysis on offer is connective rather than decompositional. This is the second horn of the dilemma: if *Responsibility-2* is understood as offering a connective analysis of moral responsibility, the Strawsonian view is adequately expressed by it, but Zimmerman’s observation no longer constitutes an objection: because a connective analysis does not purport to show that one element is metaphysically prior to another, it is perfectly compatible with the observation that being responsible may be prior to holding responsible.

The two points I have made about Zimmerman’s objection may not suffice to refute it. They suggest that *Responsibility-2* may not be an adequate expression of the Strawsonian view that moral responsibility is essentially interpersonal, the view expressed by the Slogan that is (ultimately) the target of the objection. This view would better be expressed in such a way that it does not put forward a metaphysical or explanatory relation between being responsible and holding responsible and makes clear that the analysis on offer is connective rather than decompositional.

Because my remarks about Zimmerman’s objection do not (and are not meant to) decisively refute it, I propose to have a look at what happens if the objection succeeds. First, it should be noted that the success of the objection seems to come with a price. If the objection succeeds, it threatens not only *Value-1* and *Responsibility-2,* but *all* fitting attitude analyses. Many contemporary proponents of fitting attitude analyses think that these analyses are more plausible with respect to the funny and the shameful, for example, than they are with respect to good and bad (Jacobson 2011; Todd 2016). However, Zimmerman’s ‘because’-statements seem equally plausible in both cases: just as we can say that it is appropriate to favor *x* because *x* is good, we can say that it is appropriate to be amused by *x* because *x* is funny, or that it is appropriate to feel shame about *x* because *x* is shameful. If Zimmerman’s objection succeeds, it seems to follow that all fitting attitude analyses have to be rejected. In the light of the popularity of such analyses (at least with respect to some values), that may seem like a heavy price to pay.

Secondly, and more importantly, Zimmerman himself indicates what, according to him, follows if his objection succeeds:

If this objection succeeds, then we would have to retreat from *Responsibility-2* in turn. The most that we could say would be this:

*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.*

Now, someone might insist that, even on *Responsibility-3,* moral responsibility is essentially interpersonal, but, on such an understanding of what it is for moral responsibility to be essentially *inter*personal, consider again what it would take for something to be essentially *personal.* Propositions would be essentially personal (it is appropriate under certain conditions for people to accept or reject them). Numbers would be essentially personal (it is appropriate under certain conditions for people to add or subtract them). […] Sunsets, doorknobs, black holes, alligators – *everything* would be essentially personal. And so everything having to do with *persons* would be essentially *inter*personal. On this interpretation, the thesis that moral responsibility is essentially interpersonal is robbed of all import. (2016: 261-262)

I agree with Zimmerman’s claim that, if his objection succeeds, Strawsonians would have to retreat from *Responsibility-2*. I believe, however, that there is nothing wrong with *Responsibility-3.* First of all, recall what has been concluded about the Strawsonian view and its relation to *Responsibility-2* in my discussion of Zimmerman’s objection: instead of being expressed by *Responsibility-2,* it may be thought that the Strawsonian view would better be expressed in such a way that it does not put forward a metaphysical or explanatory relation between being responsible and holding responsible and makes clear that the analysis on offer is connective rather than decompositional. These requirements are clearly met by *Responsibility-3* (and Zimmerman implicitly admits this: if *Responsibility-3* did not meet these requirements, the ‘because’-objection would be effective against *Responsibility-3* too).[[7]](#footnote-7) Many Strawsonians offer biconditional analyses such as *Responsibility-3* instead of decompositional definitions (Wallace’s ‘normative interpretation’ (1994: 91) may be the most famous example), and we have seen that the aim of Strawsonian connective analysis is to explain concepts and describe conceptual connections by way of *implication*, presupposition and exclusion. So why not accept *Responsibility-3* as an adequate expression of the Strawsonian view?

Zimmerman’s answer is that *Responsibility-3* only implies that moral responsibility is essentially interpersonal in a trivial, non-significant way. In his argument for that answer, however, he seems to do nothing more than restate the significance problem discussed in section three of this paper.[[8]](#footnote-8) The argument construes the relation between ‘personal’ and ‘interpersonal’ as a relation between quantitatively (rather than qualitatively) different concepts. As we have seen, the relation between ‘personal’ and ‘interpersonal’ is more complex. Just as the occurrence of a person in the definition of something is not the criterion by which we distinguish what is personal from what is not, the occurrence of a person in the necessary and sufficient conditions for something does not give us a reason to say that that something is essentially personal. The fact that a person *P* figures in the necessary and sufficient conditions of propositions does not give us a reason to say that propositions are essentially personal. However, we would have a reason to say that propositions are essentially personal if, for example, it would be a necessary and sufficient condition of propositions that they concern one’s private life, that they relate to a person’s body, or that they affect one particular person rather than anyone else. Because these are clearly *not* necessary and sufficient conditions of propositions, propositions are not essentially personal. Similarly, accepting that *Responsibility-3* implies that moral responsibility is essentially interpersonal should not lead us to think, *pace* Zimmerman, that everything to do with persons is essentially interpersonal. For something to be essentially interpersonal, it does not suffice that two distinct persons *P* and *Q* figure in its necessary and sufficient conditions. ‘Interpersonal’ means ‘relating to *relationships* or *communication* between people’, and, as we have seen, the notion of a reactive attitude can do the requisite work in the account of moral responsibility provided by Strawsonians.

I conclude that, even if Zimmerman’s ‘because’-objection succeeds and Strawsonians have to retreat to *Responsibility-3,* they should not be afraid of doing so. As long as the *necessary and sufficient* conditions of moral responsibility relate moral responsibility to relationships or communication between people, moral responsibility will be, in a significant way, *essentially* interpersonal. The significance of its being essentially interpersonal lies in the fact that not everything, and not everything having to do with persons, will be essentially interpersonal in this way, as Zimmerman suggests. The reference to communication between people is a necessary and sufficient condition, and therefore an ineliminable aspect, of the account of moral responsibility offered by *Responsibility-3. Responsibility-3* expresses the Strawsonian view that one would not have sufficiently elucidated what moral responsibility is without an appeal to the reactive attitudes (the appeal is a necessary aspect of any adequate account of moral responsibility), and implies the idea that moral responsibility is essentially interpersonal.

1. **Conclusion**

According to Zimmerman, *Responsibility-2* faces three problems. I have proposed ways to respond to these problems*.* First, Zimmerman himself has suggested an adequate solution to the problem of self-reactive attitudes: if ever some self-reactive attitude is appropriate, then so too, in principle, is some other-reactive attitude. This solution is true to the Strawsonian view of moral responsibility as susceptibility to the reactive attitudes and saves the interpersonal nature of moral responsibility*.*

With regard to the second problem, the problem of significance, I have made two points. According to Zimmerman, the analogy between *Value-1* and *Responsibility-2* brings out several difficulties with the latter. First, I have argued that the difference between ‘personal’ and ‘interpersonal’ is not merely quantitative; it is not simply a difference between ‘one person involved’ and ‘multiple persons involved’. Because Zimmerman makes the relation between ‘personal’ and ‘interpersonal’ seem simpler than it is, the fact that we would not say that value is essentially personal has no immediate consequences for the question whether moral responsibility is essentially interpersonal. There is no problem in denying the former and asserting the latter. Secondly, Zimmerman finds it problematic that *Responsibility-2* is compatible with someone’s being responsible for something even if no other person ever existed: if *Responsibility-2* is compatible with that idea, then how can it still imply the Slogan and the idea that moral responsibility is essentially *interpersonal*? I believe that it is doubtful whether it is possible to imagine that one person exists while no other person ever existed. If it is possible, it is, as Strawson would formulate it, barely more than a conceptual possibility. Because in order to threaten the Slogan, as Strawson understands it, such a conceptual possibility is not enough, *Responsibility-2* can still imply the Slogan in light of this possibility. It can still imply that moral responsibility is essentially interpersonal, in a sense that is as significant as the sense in which, on the Slogan, moral responsibility is essentially interpersonal.

Zimmerman’s third problem with *Responsibility-2* is that it may be taken to suggest the wrong kind of conceptual priority relation between ‘*P* is morally responsible for *x*’ and ‘it is appropriate for *Q* to adopt some reactive attitude toward *P* in respect of *x*’. While *Responsibility-2* presents ‘it is appropriate, etc.’ as more fundamental, ‘it is appropriate to adopt some reactive attitude *because P* is morally responsible’ shows that ‘*P* is morally responsible’ is more basic. I have shown that there are reasons to doubt that being responsible is prior to holding responsible, that *Responsibility-2* implies this thesis, and that the converse thesis is part of the Strawsonian view. But even if Zimmerman’s objection succeeds, Strawsonians are not in trouble. They could retreat to *Responsibility-3.* Zimmerman’s argument for rejecting *Responsibility-3* is based on a restatement of the significance argument discussed in section three. It cannot do the requisite work, because it makes the relation between ‘personal’ and ‘interpersonal’ seem simpler than it is.

I conclude that Strawsonians have ways to respond to Zimmerman’s objections. They can maintain thatthe Strawsonian view supports a significant interpretation (and as we have seen in section three, an interpretation in line with our common use of the term ‘interpersonal’) of the idea that moral responsibility is essentially interpersonal.

**References**

Bennett, Jonathan. 1980. Accountability. In *Philosophical subjects: Essays presented to P.F. Strawson,* ed. Zak van Straaten, 14-47. Oxford: Clarendon Press.

Büttner, Kai Michael. 2017. Is there such a thing as relative analyticity? *Ratio* 30: 47-56.

Clarke, Randolph, Michael McKenna, and Angela M. Smith. Introduction. 2015. In *The nature of moral responsibility. New essays,* eds. Randolph Clarke, Michael McKenna, and Angela M. Smith, 1-16. Oxford: Oxford University Press.

Darwall, Stephen. 2006. *The second-person standpoint.* Cambridge (MA): Harvard University Press.

Glock, Hans-Johann. 2009. *What is analytic philosophy?* Cambridge, Cambridge University Press.

Hacker, Peter. 2002. Strawson’s concept of a person. *Proceedings of the Aristotelian Society* 102: 21-40.

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

Jacobson, Daniel. 2011. Fitting attitude theories of value. In *The Stanford Encyclopedia of Philosophy* (Spring 2011 edition), ed. Edward Zalta. https://plato.stanford.edu/entries/fitting-attitude-theories/

Korfmacher, Carsten. 2006. Personal identity. In *The Internet Encyclopedia of Philosophy*. http://www.iep.utm.edu/person-i/

Manata, Paul. 2016. The conversational theory of moral responsibility and McKenna’s interdependence thesis. *The Philosophical Quarterly* 67: 61-83.

McKenna, Michael. 2005. Where Frankfurt and Strawson meet. *Midwest Studies in Philosophy* 29: 163-180.

McKenna, Michael. 2012. *Conversation and Responsibility.* Oxford: Oxford University Press.

McKenna, Michael. 2013. Defending conversation and responsibility. Reply to Dana Nelkin and Holly Smith. *Philosophical Studies* 171: 73-84.

Shabo, Seth. 2012. Where love and resentment meet. Strawson’s intrapersonal defense of compatibilism. *The Philosophical Review* 121: 95-124.

Sloane, Stephanie, et al. 2012. Do infants have a sense of fairness? *Psychological Science* 23: 196-204.

Smetana, Judith G., et. al. 2013. The social domain approach to children’s moral and social judgments. In *Handbook of moral development,* ed. Melanie Killen and Judith G. Smetana, 23-45. London: Routledge.

Strawson, P.F. 1980. P.F. Strawson replies. In *Philosophical subjects: Essays presented to P.F. Strawson,* ed. Zak van Straaten, 260-296. Oxford: Clarendon Press.

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

Strawson, P.F. 2009 [1962]. Freedom and resentment. In *Free Will,* ed. Derk Pereboom, 148-171.Indianapolis: Hackett.

Todd, Patrick. 2016. Strawson, moral responsibility, and the ‘order of explanation’. An intervention. *Ethics* 127: 208-240.

Wallace, R. Jay. 1994. *Responsibility and the moral sentiments.* Cambridge (MA): Harvard University Press.

Wynn, Karen and Paul Bloom. 2013. The moral baby. In *Handbook of moral development,* ed. Melanie Killen and Judith G. Smetana, 435-453. London: Routledge.

Zimmerman, Michael J. 2007. The good and the right. *Utilitas* 19: 326-353.

Zimmerman, Michael J. 2010. Responsibility, reaction, and value. *The Journal of Ethics* 14: 103-115.

Zimmerman, Michael J. 2015. Varieties of moral responsibility. In *The nature of moral responsibility. New essays,* eds. Randolph Clarke, Michael McKenna, and Angela M. Smith, 45-64. 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.

1. Strictly speaking, Zimmerman offers a definition of *retrospective* moral responsibility. Because the difference between prospective and retrospective moral responsibility is of no importance for my argument, I will leave ‘retrospective’ out of the discussion. [↑](#footnote-ref-1)
2. A reviewer has remarked that charity would seem to demand that I inquire further into why Zimmerman thinks that self-reactive attitudes might nonetheless pose a problem for *Responsibility-2.* I must confess that my inquiries have not been successful. Neither in Zimmerman’s paper nor in his other works have I been able to find a reason why this solution to the problem of self-reactive attitudes would be inadequate. It should be remarked, moreover, that it is not entirely clear *that* Zimmerman thinks that this solution is inadequate; after all, he describes it as ‘more promising’ than another solution (2016: 258). One reason why he may nevertheless be reluctant to accept it is that the proposed solution, though promising, is nothing more than a suggestion that has to be worked out. If that is the case, I hope that my discussion goes some way towards explaining why the solution’s promise is not empty. [↑](#footnote-ref-2)
3. There are many versions of the view of color expressed in this definition, and many more or less similar definitions have been proposed to capture that view. The subtleties of the color discussion are of no importance for my argument. [↑](#footnote-ref-3)
4. I am grateful to \*\*\* for suggesting this point to me. [↑](#footnote-ref-4)
5. It should be noted that, in his presentation of the fitting-attitude analysis of value, Zimmerman uses the word ‘person’. However, most proponents of fitting-attitude analyses take these analyses to bring out that values cannot be understood independently of *human* responses (Jacobson 2011). This move is not innocent: if Zimmerman had formulated *Responsibility-2* in terms of human beings rather than persons, his objection would have seemed less compelling: the possibility that one human being exists while no other human being ever existed does not seem to be a coherent idea. [↑](#footnote-ref-5)
6. I would like to thank an anonymous reviewer for pressing this point. [↑](#footnote-ref-6)
7. The requirements are also met by *Responsibility-2,* provided that one understands the idea of a definition in a certain way. According to Zimmerman, a definition offers a decompositional conceptual analysis. According to a recent contribution on definitions by Kai Büttner, however, definitions only determine ‘that the conditions for the applicability of the *definiens* are also the conditions for the applicability of the *definiendum*’ (Büttner 2017: 53). I will not go into the theory of definitions here. [↑](#footnote-ref-7)
8. He does *not* argue that only decompositional definitions of *x*, and not accounts of necessary and sufficient conditions of *x*, can provide information about what is essential to something. [↑](#footnote-ref-8)