Moral Responsibility and the Moral Community: Is Moral Responsibility Essentially Interpersonal?

Michael J. Zimmerman[1]

Abstract Many philosophers endorse the idea that there can be no moral responsibility without a moral community and thus hold that such responsibility is essentially interpersonal. In this paper, various interpretations of this idea are distinguished, and it is argued that no interpretation of it captures a significant truth. The popular view that moral responsibility consists in answerability is discussed and dismissed. The even more popular view that such responsibility consists in susceptibility to the reactive attitudes is also discussed, and it is argued that this view at best supports only an etiolated interpretation of the idea that moral responsibility is essentially interpersonal.

Keywords Answerability · Blameworthiness · Holding responsible · Interpersonal · Moral community · Moral responsibility · Prospective responsibility · Reactive attitudes · Retrospective responsibility

1 Introduction

T. M. Scanlon asks us to imagine the following scenario:

> Suppose I learn that at a party last week some acquaintances were talking about me, and making some cruel jokes at my expense. I further learn that my close friend Joe was at the party, and that rather than coming to my defense or adopting a stony silence, he was laughing heartily and even contributed a few barbs, revealing some embarrassing facts about me that I told him in confidence. (Scanlon 2008: 129)

✉ Michael J. Zimmerman
   m_zimme2@uncg.edu

[1] University of North Carolina, Greensboro, NC, USA

Scanlon states that Joe's behavior raises a question about the relationship between the two of them. Indeed it does. The particular question that Scanlon asks is whether he should still consider Joe to be his friend. Another question that might be asked is whether Joe is to blame for his shabby behavior. For Scanlon, these questions are intimately related. On his view, blame has essentially to do with impaired relationships. He writes: "To blame a person for an action…is to take that action to indicate something about the person that impairs one's relationship with him or her, and to understand that relationship in a way that reflects this impairment." (Scanlon 2008: 122–123) He states further that to blame someone is, in part, to judge him or her to be blameworthy (Scanlon 2008: 128), and he takes blameworthiness to be a form of moral responsibility.

Scanlon is not the first to suggest that moral responsibility has to do with personal relationships; on the contrary, he explicitly notes that P. F. Strawson likewise understands human relationships to be "the foundations of blame." (Scanlon 2008: 128) Indeed, Strawson regards such relationships to be the foundations of moral responsibility in general. In his celebrated essay "Freedom and Resentment," Strawson gives an account of moral responsibility according to which such responsibility consists in being the target of certain "non-detached," "reactive" attitudes, whether favorable (such as gratitude) or unfavorable (such as resentment), attitudes that involve seeing the person to whom they are directed "as a morally responsible agent, as a term of moral relationships, as a member of the moral community." (Strawson 1974: 17) Strawson's view has proven highly influential. One implication of it, upon which I will concentrate in this paper, might be expressed as follows:

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

Now, this slogan, as presented here, is my invention. I do not know of anyone who has stated it in just the way that I have done. Nonetheless, it seems to me to capture the spirit of an increasingly popular approach to moral responsibility, one according to which such responsibility is, in some important way, essentially interpersonal. This approach has been endorsed in one way or another not only by Strawson and Scanlon but also by many other philosophers, among whom are R. Jay Wallace (1994), Gary Watson (2004), Angela Smith (2005, 2008, 2012, 2015), Stephen Darwall (2006), David Shoemaker (2007, 2011, 2015), Michael McKenna (2012), and Derk Pereboom (2013).

As it stands, the Slogan is of course obscure. It needs interpretation. In what follows, I will argue that there is no interpretation of it that captures a significant truth.

2 How to Understand "Can"

First, we must ask how "can," as used in the Slogan, is to be interpreted. The answer is that it expresses *conceptual* possibility. The Slogan claims that it is *inconceivable* that there should be moral responsibility in the absence of a moral community. Some of the Slogan's supporters state this explicitly (e.g., Strawson

1974: 21; Darwall 2006: 11–12; McKenna 2012: 107f.), others implicitly (e.g., Watson 2004: 7; Scanlon 2008: 122–123; Smith 2008: 369). For still others, the evidence is not altogether unequivocal, and it may be that some of them are concerned with some other kind of possibility—for example, with what is "humanly" possible as opposed to what is conceptually possible. I will briefly discuss such an alternative interpretation in Sect. 8 below, but otherwise my focus will be on conceptual possibility and its complement, conceptual necessity.

3 How to Understand "Moral Responsibility"

Next we must ask how the term "moral responsibility," as used in the Slogan, is to be understood. It is a commonplace that this term, along with its cognates (such as "morally responsible"), is ambiguous. As just noted, Strawson talks of morally responsible agents (period), but of course we also talk of agents being morally responsible *for* something. As I see it, it is the latter locution that is the more fundamental, although even here there is ambiguity. We can distinguish two main senses of "being morally responsible for something." There is a *prospective* sense, in which the thing for which one is responsible lies in the future. A lifeguard, for example, may be morally responsible for the continued safety of the swimmers under his care. Instead of saying that the lifeguard is responsible for the swimmers' safety, we might say that he has a responsibility to see to it that they remain safe. Here "responsibility" means the same as, or very close to the same as, "duty" or "obligation." The second main sense of "being morally responsible for something" is *retrospective*, since the thing for which one is responsible lies in the past. In this sense, the lifeguard might be responsible for the swimmers' deaths. We would not say that he *has a* responsibility for the deaths, but we might well say that he *bears* responsibility for them. In this case, of course, the responsibility in question is "negative"; it consists in culpability or blameworthiness. In other cases, such as when someone is morally responsible for having rescued a swimmer, the retrospective responsibility may be "positive," consisting in laudability or praiseworthiness.

The sorts of things for which one can have a prospective responsibility include bits of behavior, whether active or passive, and the outcomes thereof. (If the bit of behavior is passive, it will constitute an omission, but not every omission is passive. On occasion we perform acts of omission.) Thus the lifeguard has a responsibility to see to it that the swimmers remain safe, and he is (also) responsible for their safety. The sorts of things for which one can bear retrospective responsibility likewise include bits of behavior and the outcomes thereof. If the swimmers drown, the lifeguard may be responsible for having let them drown and (also) responsible for their deaths. According to some philosophers (e.g., Smith 2005; Scanlon 2008: 192f.), one can also be retrospectively responsible for one's attitudes (whether or not they result in behavior). These philosophers may also hold that one can have a prospective responsibility to see to it that one has or lacks certain attitudes, although this is a matter that is seldom addressed explicitly.

What is a morally responsible agent (period)? Here, too, there would seem to be two main senses of the term. In one sense, a morally responsible agent (period) is someone who takes his or her prospective responsibilities *seriously*. The contrary of "responsible" in this sense is "irresponsible," and the trait in question is scalar; for one can be more or less (ir)responsible. In another sense, a morally responsible agent is someone who is *capable* of being morally responsible for something, whether prospectively or retrospectively. It is this sense that Strawson has in mind (although he seems to think of morally responsible agents as individuals who not only have the capacity in question but also manifest it—that is, he seems to think of them as indeed being morally responsible for something). There is no common contrary of "responsible" in this sense, although sometimes the term "non-responsible" is used for this purpose. We tend to treat responsibility in this sense as an all-or-nothing affair—one either has the capacity in question or not—but matters tend to get murky when the moral status of children and other "marginal" individuals is in question. (Shoemaker 2007: 2015)

It is worth noting that, in the capacity sense of "morally responsible," there are two capacities at issue, one having to do with prospective responsibility, the other having to do with retrospective responsibility. It may be that these capacities always accompany one another, although in my view they do not. As I see it, less is required for having the former capacity (the capacity to do moral right or wrong) than for having the latter (the capacity to be morally praiseworthy or blameworthy), but I will not pursue the matter here. It is also worth noting that prospective and retrospective responsibility can come apart in another way. Even if one cannot have one capacity without having the other, still it is possible to have a prospective responsibility without bearing any retrospective responsibility. This will be the case when the occasion for fulfilling one's prospective responsibility lies in the future (as when the lifeguard has yet to go on duty), and it might also be the case when the occasion lies in the past. The latter possibility is less straightforward, but if, while he was on duty, no swimmers needed saving, it would be rather odd to claim that the lifeguard is morally responsible for what took place. Certainly, there would seem to be no call for either praising or blaming him, although we might want to say, for the sake of some kind of completeness, that he is morally responsible in a "neutral" way for what happened. (Smith 2015: 107) But note, too, that, even if the lifeguard allowed some of the swimmers to drown, thereby failing to fulfill his prospective responsibility, it might still be the case that he is not to blame for their deaths; for he might have an excuse for his failure.

So far we have been concerned with various ways of *being* responsible, but another distinction worth noting is that between being and *holding* responsible. One can hold someone responsible for something either prospectively or retrospectively. The lifeguard's boss might say, "I hold you responsible for the swimmers' safety." This would seem to mean something like, "Don't screw up! You'll be in trouble if you do," where the trouble in question would involve some form of blame or sanction. Or the boss might say, "I hold you responsible for the swimmers' deaths," in which case the lifeguard is already in trouble. [We can also *hold* someone *to be* responsible for something. (Watson 2004: 267) Here "hold" simply means the same

as "judge" and does not necessarily involve any prospect or instance of praise or blame.]

4 How to Understand "Moral Community"

How is the term "moral community" to be understood? In the passage quoted above, Strawson seems to regard "a morally responsible agent," "a term of moral relationships," and "a member of the moral community" as synonymous. I suspect that by "a term of moral relationships" Strawson means someone who not only has the capacity to have, but actually has, moral relationships with other agents, and that it is in virtue of having these relationships that he takes the agents together to form a moral community. (This is why I stated that Strawson seems to think of morally responsible agents as individuals who not only have the capacity to be responsible for something but also manifest this capacity.) The relationships in question presumably involve having moral responsibilities, both prospective and retrospective, that somehow concern some or all members of the community in question.

Strawson talks of "the" moral community and not simply, as I have done in my formulation of the Slogan, "a" moral community. Others have followed his example (e.g., Darwall 2006: 7; Shoemaker 2007, 2015: 5—but contrast Shoemaker 2015: ix). "The" is of course more restrictive than "a," and so the Slogan, as I have formulated it, is weaker than the thesis that some of its supporters might have in mind. Perhaps the distinction between "the" and "a" moral community matters for some purposes, but I will not pursue the matter here. I will stick with the weaker formulation, since that makes the denial of its significance all the stronger.

5 How to Understand the Slogan

Here again is the Slogan:

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

I have just stated that I understand a moral community to be a group of individuals interrelated by responsibilities, both prospective and retrospective, that somehow concern one another. This is still rough, but I think it will do for present purposes. The question remains, though, how to understand the Slogan's use of "moral responsibility." Let us consider the distinctions drawn in Sect. 3 in reverse order.

5.1 Being Responsible vs. Holding Responsible

Given the distinction between being responsible and holding responsible, we may distinguish between these two versions of the Slogan, where the variables range over individuals:

> *Version 1*: Necessarily,[1] if *P* is morally responsible, then *P* is a member of a
>   moral community;
> *Version 2*: Necessarily, if *Q* holds *P* morally responsible, then both *Q* and
>   *P* are members of a moral community.

Version 2 could of course be qualified in certain ways—having to do with whether *Q* and *P* are members of the same moral community or of different communities, with whether both of them or only one of them is a member of such a community, and so on—but I will not pursue the issue. For although supporters of the Slogan are typically very much concerned with what is involved in holding someone morally responsible for something, I think it is clear that neither Version 2 nor any such qualification of it is acceptable, and also clear that it is not any such thesis that the supporters are intent on supporting.

Oliver Sacks tells the memorable tale of a man who mistook his wife for a hat. (Sacks 1998) Consider someone, *Q*, with an inverse form of visual agnosia who mistakes a hat for his wife and holds it/her morally responsible for many imagined misdeeds, and suppose that *Q* is a hermit, a member of no community at all.[2] If he is a member of no community at all, then *Q* is not a member of any moral community. Nonetheless, *Q* holds his hat, *P*, morally responsible for the imagined misdeeds. Moreover, *P* is of course also not a member of any moral community. Hence Version 2 of the Slogan must be rejected.

To meet this objection, someone might propose the following qualification of Version 2:

> *Version 2′*: Necessarily, if *Q* correctly holds *P* morally responsible, then both
>   *Q* and *P* are members of a moral community.

Since the man does not correctly hold the hat morally responsible for any misdeeds, his case provides no counterexample to this version of the Slogan. But if, as seems evident, it is correct to hold someone morally responsible only if that person is morally responsible, Version 2′ implies Version 1, and so we might as well attend directly to the latter.

5.2 Being Responsible for Something vs. Being Responsible (Period)

At this point, the distinction between being morally responsible for something and being morally responsible (period) comes into play. Should Version 1 be understood as follows:

> *Version 3*: Necessarily, if *P* is morally responsible for something, then *P* is a
>   member of a moral community;

or should it rather be understood as follows:

---

[1] In keeping with what was stated in Sect. 2, "necessarily" here expresses *conceptual* necessity.

[2] *Q* will of course still be a member of the *set* of persons, but that does not make him a member of a *community* whose members have significant interrelations.

Version 4: Necessarily, if *P* is morally responsible (period), then *P* is a member of a moral community?

Let us consider Version 4 first.

As I noted earlier, the sense of "being morally responsible (period)" that is at issue here is the capacity sense (and not the sense that concerns taking one's prospective responsibilities seriously). But, as I also noted, Strawson seems to think of morally responsible agents as individuals who not only (i) have the capacity to be morally responsible (whether prospectively or retrospectively) for something but also (ii) manifest this capacity. Let us treat (i) and (ii) separately.

It seems clear that, if (i) alone is at issue, then Version 4 of the Slogan is false, but also that, once again, it is not this version that the Slogan's supporters have in mind. The capacity in question is a property that is intrinsic to its bearer (whereas being a member of a community is not).[3] The unfortunate man who mistook a hat for his wife may well have had the capacity to be morally responsible for something, but he was not a member of any moral community. Thus it cannot be (i) alone that is at issue. But if (ii) is included, then, given that no one can be morally responsible for something without having the capacity to be morally responsible for it, Version 4 is equivalent to Version 3. In either case, then, it is on Version 3 that we should focus.

5.3 Prospective Responsibility vs. Retrospective Responsibility

At this point, the distinction between prospective and retrospective moral responsibility comes into play. Should Version 3 be understood as follows:

Version 5: Necessarily, if *P* is prospectively morally responsible for something, then *P* is a member of a moral community;

or should it rather be understood as follows:

Version 6: Necessarily, if *P* is retrospectively morally responsible for something, then *P* is a member of a moral community?

If, as Strawson seems to think, being a morally responsible agent is the same as being a term of moral relationships, and if being such a term involves having both prospective and retrospective responsibilities, then Strawson's own answer to this question would appear to be "Both." And I think that this is the answer that most supporters of the Slogan would give. Still, it seems clear that it is with Version 6 in particular that they have been primarily concerned, and it is on this version that I will focus in the remainder of this paper.

---

[3] Saying that the capacity is intrinsic to its bearer is of course compatible with saying that the manifestation of the capacity may depend on something extrinsic to the bearer (such as the presence of alternate possibilities, say)—just as, for example, the exertion of some external force may be required for a glass to manifest its fragility.

6 Responsibility as Answerability

Consider again the case of Joe, Scanlon's friend (or supposed friend). Scanlon mentions three ways in which he might respond to what Joe has done. (Scanlon 2008: 129–130) He might wonder whether he should continue to regard Joe as a friend; he might revise his attitude toward Joe, ceasing to value spending time with him and to confide in him; or he might complain to Joe, demanding an explanation for or justification of his conduct. Each of these responses would appear to indicate that Scanlon already regards his relationship with Joe to have been impaired, at least to some extent, by Joe's behavior and already blames him for what he has done. But might not such blame be premature? What if, in responding to Scanlon's demand that he account for his behavior, Joe provides a convincing justification of it? Then, it seems, Scanlon's blaming him would be unfounded. (Pereboom 2013: 220) (What might the justification be? I will leave that to your imagination.) Even if Scanlon's blame turns out to be unfounded, though, it certainly seems appropriate for him to call upon Joe to answer for what he has done. Joe's being *answerable* in this way to Scanlon for his behavior is, it might be thought, what his retrospective moral responsibility for it consists in. This, indeed, is precisely how Scanlon himself conceives of such responsibility (Scanlon 1998: 272, 2008: 130, 193), and it is a conception shared by many others, including Watson (2004: Part III), Darwall (2006: 82–83), and especially Angela Smith, who has developed and defended this conception in detail in a number of places. (Smith 2005, 2008, 2012, 2015) Here is a representative statement of her view: "To say that an agent is morally responsible for something…is to say that that thing reflects her rational judgment in a way that makes it appropriate, in principle, to ask her to defend or justify it." (Smith 2008: 369, cf. 2012: 577–578, 2015: 103) One feature of this view, not in fact emphasized by Smith but stressed by others, is that it construes responsibility as a *triadic* relationship: an individual is answerable *to* others *for* something. (Watson 2004: 7, 274; Darwall 2006: 68–69) Stressing this feature serves to highlight the interpersonal nature of moral responsibility, on this conception. Accordingly, I suggest the following formulation of the view (where P and Q are distinct persons and x is whatever sort of thing P might be responsible for, whether an attitude, a bit of behavior, or the outcome of a bit of behavior):

> *Answerability*:
> P is retrospectively morally responsible to Q for x = df. It is appropriate for Q to call on P to answer for (i.e., to defend or justify) x.

A number of points are to be noted about this view of responsibility. First, as Smith observes, it gains some support from considerations of etymology; it conceives of responsibility as response-ability. (Smith 2015: 103, note 13) Second, as she also points out, if P does respond to Q's demand for P to defend or justify x, such a response "is the key to opening the door to the further moral responses that *may* (depending on the details of the case) appropriately follow upon the answer she

gives." (Smith 2015: 103)[4] On this view, then, responsibility is seen as involving a kind of *conversation* between members of a moral community, an aspect of responsibility emphasized not only by Smith but also by Darwall (2006: 75f.) and, especially and in great detail, by McKenna. This conversational aspect of moral responsibility, McKenna states, is precisely what renders such responsibility "*essentially* interpersonal." (McKenna 2012: 110) Third, it is not a presupposition of this view that the relation between *P* and *Q* must be a "personal" one of the sort that obtains between Scanlon and Joe; it may simply be the kind of "moral" relationship that obtains between members of a moral community, in virtue of which they are co-members of that community. (Scanlon 2008: 139) And fourth, so understood, this view clearly implies Version 6 of the Slogan.

7 Objections to Answerability

Plausible as this view of responsibility may seem, however, it faces a number of problems, as I will now explain.

7.1 Wrongdoing that Appears not to Involve other Persons

One problem with Answerability is that it seems to overlook responsibility for wrongdoing that appears not to involve other persons. Think of mistreating animals or destroying the environment. What does such behavior have to do with other members of some moral community?

One response is to say that Answerability is not supposed to account for responsibility for such wrongdoing; it is supposed only to concern wrongs done to other persons. At times, Scanlon (1998: 178f., 2008: 124) seems to suggest this response, and Darwall does too (Darwall 2006: 28), but I have to say that it seems to me quite ill advised, since it belies the claim, to which their allegiance to Answerability commits them, that retrospective moral responsibility is essentially interpersonal.

A second response, one that is wholly in keeping with Answerability, is to say that the obligation that we have not to mistreat animals or destroy the environment is one that we owe to other members of our moral community. Both Scanlon (2008: 166) and Darwall (2006: 28–29) suggest this response too, but it faces difficulties. It is surely the case that I have a moral obligation not to mistreat my dog or destroy the local habitat, but it is not at all clear to me that I owe it to anyone else not to do these things. Still, perhaps I do, in which case these examples do not impugn Answerability after all. If I do not owe it to others not to do these things, however, Answerability is in trouble. In the case of my dog, I think I probably do owe it to him not to mistreat him, but I am not sure that that provides any support for Answerability. Even if the notion of a moral community is understood expansively, so as to include animals such as my dog, the fact remains that my dog cannot call on

---

[4] Why "defend *or* justify"? It might be that, although *P* cannot justify *x*, *P* can furnish an adequate excuse for *x*. Or it might be, of course, that *P* can neither defend nor justify *x*. *P*'s success, or lack thereof, in answering for *x* will dictate the nature of *Q*'s counter-response to *P*.

me to answer for my mistreating him. Maybe a proxy for my dog could do so, however, in which case Answerability might still be thought to apply to such a case. Even if so, this kind of account cannot be applied to destruction of the environment. Although I have an obligation not to destroy the local habitat, I do not owe it to the habitat not to destroy it, and it makes no sense at all to think of the habitat as a member of any moral community.[5]

7.2 Blameworthiness vs. Praiseworthiness

Another problem with Answerability is that it seems more suited to playing a role in an account of culpability or blameworthiness in particular than in an account of retrospective moral responsibility in general. As noted in Sect. 3, such responsibility can on occasion take the form of laudability or praiseworthiness and, whereas it is often appropriate to call on someone to answer for some wrongful act, it is rarely, if ever, appropriate to call on someone to answer for doing what was right.

Smith is aware of this difficulty. At one point she writes this:

> This feature of negative moral appraisal [viz., calling upon an agent to answer for something] is more difficult to extend to its positive analogue, but I think we can say something like the following: positive moral appraisal, though it does not itself address a demand to its target, by its nature involves an acknowledgment that the agent has exceeded in some way the legitimate moral demands that apply to her. (Smith 2008: 381, note 15)

This response to the difficulty strikes me as inadequate. Smith appears to conflate two kinds of demand. First, there are the demands that *morality* makes of an agent, P. These consist in the moral obligations or duties that P has to behave in certain ways. (To call these demands is to speak metaphorically. Morality is not literally capable of demanding anything; only individual persons can do that.) Then there is the demand that *another person*, Q, makes of P when Q calls upon P to answer for having done something that apparently violated some obligation. (To call this a demand is to speak literally.) It is the latter kind of demand that is at issue in Answerability. That P has gone beyond what morality demands of her does not render it appropriate to demand that she answer for her behavior.

7.3 Being Responsible to Someone

A third problem with Answerability is that, even if its application is restricted to the negative aspect of responsibility that is blameworthiness, the fit is still awkward. If we substitute "blameworthy" for "responsible" in the statement of the thesis, we end up with the nonsensical phrase "P is blameworthy to Q for x." This is an indication that the connection between responsibility as understood in

---

[5] In making this claim, I am presupposing that we can have moral obligations *regarding* certain entities (including people) that we do not owe *to* these things, and hence that these things do not have a moral *right* to the fulfillment of the obligations, even though it would of course be morally *wrong* not to fulfill them. [Cf. Feinberg (1980: 143f. and 161f.) although Feinberg's emphasis tends to be on legal, rather than moral, rights and obligations].

Answerability and responsibility understood as consisting in either praiseworthiness or blameworthiness (or, possibly, the worthiness of some neutral response [see Sect. 3]) is indirect at best.

The diagnosis is not hard to come by. The phrase "*P* is responsible to *Q* for *x*" is intelligible only when the form of responsibility at issue is *prospective*; it implies that *P* owes it to *Q* to do something in respect of *x*. And, of course, this is precisely what Answerability maintains; it maintains that *P* owes it to *Q* to answer for *x*. But *why* does *P* owe it to *Q* to answer for *x* (if, indeed, *P* does owe this to *Q*)? The answer, I suggest, is that *P* owes this to *Q* at least in part because *P* is a morally responsible agent (period), in the sense of being capable of being *retrospectively* morally responsible (more particularly, blameworthy) for *x*. As I see it, therefore, Answerability does not *provide* an account of retrospective moral responsibility; rather, it *presupposes* some such account.

8 Responsibility as Susceptibility to Reactive Attitudes

Smith claims that Answerability gains support from considerations of etymology. If there is any such support, it seems to me weak. I suggest that such considerations provide stronger support for the view that responsibility consists in the *worthiness* of some response. (Compare admirability, desirability, reprehensibility, and so on.) This observation of course aligns perfectly with understanding praiseworthiness and blameworthiness to be particular forms of responsibility.[6] It is important to recognize that such a conception of retrospective moral responsibility might still be understood to imply that such responsibility is essentially interpersonal. It is, after all, this conception that Strawson invokes when he characterizes such responsibility in terms of the appropriateness of certain reactive attitudes. (Strawson 1974: 6) Many such attitudes constitute forms of praise and blame, and they clearly qualify as a kind of response. Consider, then, this thesis:

> *Responsibility-1*:
> *P* is retrospectively 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*.[7]

On the assumption that the appropriateness of *Q*'s reaction toward *P* presupposes that *P* and *Q* are members of the same moral community, Responsibility-1, like Answerability, also implies Version 6 of the Slogan, but it does not face any of the problems that have been raised so far for Answerability (not even the first: 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

---

[6] I should note, however, that it is not clear how much weight such etymological considerations can bear. It is difficult to see how the construal of responsibility as the worthiness of some response could be correctly applied to prospective responsibility or causal responsibility.

[7] Two important questions that arise here are these. Which attitudes qualify as reactive attitudes? And what is the relation between the appropriateness and the worthiness of a response? I will have nothing to say about the former. I will have something to say about the latter in Sect. 8.4 below.

praise me for treating them well). Nonetheless, Responsibility-1 does face problems of its own.

8.1 Appropriateness in Fact vs. Appropriateness in Principle

One problem is that there may be no person Q for whom it is in fact appropriate to adopt some reactive attitude toward P in respect of x. It may be, for example, that no one knows the pertinent facts about P's relation to x, or that no one has the requisite moral standing or authority to adopt the pertinent reactive attitude (or set of attitudes) toward P in respect of x, and so on.

Smith in particular has addressed this problem [and others are aware of it (e.g., Scanlon 2008: 175f.)]. It is for this reason that, as noted in Sect. 6, she claims, not that it is *in fact* appropriate for someone Q to respond to P in the relevant way, but that it is *in principle* appropriate for someone Q so to respond. This idea may seem promising—for one thing, it rules out what would otherwise be a very peculiar and surely unacceptable form of moral luck—but I do not think that it can be cashed out in a way that lends support to Responsibility-1.

8.2 Self-Reactive Attitudes

To begin with, there is the question whether there need be someone Q distinct from P for whom it is appropriate, even if only in principle (whatever exactly that is supposed to mean), to adopt some reactive attitude toward P. In his discussion of the reactive attitudes, Strawson explicitly acknowledges that some such attitudes are "self-reactive" (rather than "other-reactive"), citing remorse as an example. (Strawson 1974: 15) This seems quite right, and it is a point that has been echoed by others (e.g., Watson 2004: 274; Scanlon 2008: 154–155), yet it would seem to be in tension with the idea that retrospective moral responsibility is essentially *inter*personal.[8]

Jonathan Bennett has this to write on the matter:

> A self-reactive attitude does involve an important 'interpersonal' relation: remorse, for instance, can be represented as a confrontation—with an accusing glare on one side and downcast eyes on the other—between one's present self and some past self. (Bennett 1980: 44)

Darwall approves of this passage (Darwall 2006: 74), but I find it forced: a "community" of one, composed of temporal slices of oneself! In any case, if all it takes is one person to make a moral community, then of course the Slogan is true, but only trivially so.[9]

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

---

[8] A related question could of course have been raised regarding Answerability. Alongside of the concern that has to do with obligations regarding animals and the environment, there is a concern that has to do with self-regarding obligations.

[9] "Of course" is perhaps too strong. Perhaps one's temporal slices will fail to be interrelated in the appropriate way. Cf. note 2, above.

appropriate for someone else, *Q*, to adopt some related attitude toward *P* in light of what *P* has done.

8.3 The Time at which Reactive Attitudes are Appropriate

But there is still a problem. Just *when* is it appropriate for *Q* to adopt some reactive attitude toward *P*? Suppose that Martha commits murder on Monday in such a way that we would not hesitate to say that she is morally responsible—more particularly, culpable—for what she has done. (Her quality of will is defective in whatever way Strawson et al. require.) But suppose that on Tuesday Martha takes leave of the moral community: she lapses into insanity, or into a coma, or undergoes some other kind of transformation that provides her with some kind of exemption from moral responsibility. Would it be appropriate for Ralph on Wednesday to adopt some reactive attitude toward Martha—that of reprimanding her severely, say—in light of her behavior on Monday? One answer is "Yes." But this is a troubling answer, since the target—or, rather, the nature of the target—of Ralph's attitude has changed drastically. The other answer is "No." But then what does Responsibility-1 imply? Should we infer that, since it is not appropriate for Ralph to reprimand Martha for the murder, Martha is not responsible for it after all? This is troubling, too, since it once again introduces a very peculiar form of moral luck and, besides, seems at odds with the facts about Martha's quality of will on Monday; moreover, it overlooks the apparent fact that, had Ralph reacted sooner and reprimanded Martha for the murder prior to her lapse, his reaction would have been perfectly appropriate.

There may seem to be an obvious solution to this problem, and that is to take care to specify the times involved. Just as it would have been appropriate for Ralph to reprimand Martha for her behavior at any time between the murder and the lapse, so too we can and should say that Martha was morally responsible for the murder during that same period of time, but not afterwards. This is a tempting response, but I do not think it solves the problem. For what if Martha's lapse had occurred *immediately* after her commission of the murder? There would have been no time for Ralph to adopt an appropriate reactive attitude, in which case Responsibility-1 appears to imply that there was no time at which Martha was responsible for her behavior. But that contradicts the initial hypothesis that the manner in which Martha engaged in her behavior sufficed for her incurring responsibility for it.

Perhaps the best that can be said here on behalf of Responsibility-1 is this: Martha's behavior (and the quality of will that underlay it) was such that it was in principle appropriate for another member of her moral community to reprimand her immediately. That no one could in fact have reacted so quickly does not undermine this fact.

8.4 Responsibility in Isolation

But what if Martha committed mass murder on a global scale, and there was no one left to reprimand her? Would she then have luckily escaped responsibility for her behavior after all? Surely not. That would be a grotesque conclusion to draw, and

yet Responsibility-1 seems to mandate it. For if no person exists (other than Martha; but perhaps she committed suicide at the same time as committing mass murder), then there is and was no one for whom it is or was appropriate even in principle to reprimand Martha.

McKenna has discussed a similar case, one modeled on Daniel Defoe's *Robinson Crusoe*. He imagines Robinson living on a desert island, a member of no community, moral or otherwise, mercilessly beating his dog. Surely Robinson is blameworthy for his behavior; yet there is no one to blame him. McKenna replies that, in Defoe's story, Robinson was an Englishman who, though a castaway, "carried with him in his head, so to speak, the moral community of his earlier life…. Were we to encounter him, there is no reason to think that we would not have (at least *pro tanto*) reason to blame him." (McKenna 2012: 108) Maybe so, but this does not suffice as a defense of Responsibility-1. Surely, Robinson's blameworthiness, like Martha's, is not contingent on our or anyone else's actually being in a position to blame him for what he has done. Nor is Robinson's blameworthiness contingent on his having been a member of some community; 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 (period).

To say that it is conceivable that Robinson developed the relevant capacities in isolation is not to say that it is easy to imagine the details of his doing so. Human beings typically rely on the help of role models to develop these capacities, but it does not follow from this fact that they are essentially incapable of developing them in isolation from other people. 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. But even if it should turn out that for some reason humans are indeed essentially incapable of doing this, it does not follow that it is conceptually impossible that some individual do so.

9 Conclusion

That a person *P*, whether isolated like Martha and Robinson or, like you and me, a member of some community, is retrospectively morally responsible for something *x* has to do, I suggest, not with its being appropriate, whether in fact or in principle, for some actual person *Q* to adopt some reactive attitude toward *P* in respect of *x*. It has rather to do with *P*'s being *deserving* or *worthy* of such a reaction, and whether *P* is so deserving or worthy has to do, in turn, only with *P* and not with anyone else who happens to exist. Of course, someone else must exist if *P* is to receive the reaction he deserves (unless the reaction in question involves a "self-reactive" attitude), but his responsibility does not turn on his actually receiving it. Responsibility-1 is therefore false. The most we can say is something along the following lines:

Responsibility-2:
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 reactive attitude toward *P* in respect of *x*.

(I will not venture to specify just what *C* might consist in.)

It might be said that, even on Responsibility-2, there is truth in the claim that retrospective moral responsibility is essentially interpersonal. But here I would make three points.

First, given the concern raised above about self-reactive attitudes, the stipulation in Responsibility-2 that *Q* be distinct from *P* seems ill advised. If that stipulation is dropped, though, then of course the essentially *inter*personal nature of moral responsibility vanishes with it.

Second, even if the stipulation remains, it is questionable how *significant* the interpersonal nature of moral responsibility is, according to Responsibility-2. To appreciate this point, consider a now-popular parallel thesis regarding the nature of value. Many philosophers (e.g., Ewing 1948: 152; Brentano 1969: 18; Chisholm 1986: 52; Scanlon 1998: 96–97) endorse what has come to be called the fitting-attitude analysis of value and which may be rendered thus:

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*.

(Again, I will not concern myself with just what *C* might consist in.) Now, consider this question: 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.

Finally, just as some philosophers (e.g., Blanshard 1961: 284–286) 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*. 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,[10] *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*.

---

[10] Here, "necessarily" cannot express conceptual necessity. (Cf. note 1, above.) Rather, it expresses logical or metaphysical necessity.

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). Trees would be essentially personal (it is appropriate under certain conditions for people to climb them, or to cut them down, or to sit in their shade). 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.[11]

References

Bennett, Jonathan. 1980. Accountability. In *Philosophical subjects*, ed. Zak Van Straaten, 14–47. Oxford: The Clarendon Press.

Blanshard, Brand. 1961. *Reason and goodness*. London: George Allen and Unwin.

Brentano, Franz. 1969. *The origin of our knowledge of right and wrong*. London: Routledge and Kegan Paul.

Chisholm, Roderick M. 1986. *Brentano and intrinsic value*. Cambridge: Cambridge University Press.

Darwall, Stephen. 2006. *The second-person standpoint: Morality, respect, and accountability*. Cambridge: Harvard University Press.

Ewing, A.C. 1948. *The definition of good*. London: Macmillan and Co.

Feinberg, Joel. 1980. *Rights, justice, and the bounds of liberty*. Princeton: Princeton University Press.

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

Pereboom, Derk. 2013. Moral responsibility without desert. In *Free will and moral responsibility*, ed. Ishtiyaque Haji, and Justin Caouette, 213–228. Newcastle upon Tyne: Cambridge Scholars Publishing.

Sacks, Oliver. 1998. *The man who mistook his wife for a hat and other clinical tales*. New York: Simon and Schuster.

Scanlon, T.M. 1998. *What we owe to each other*. Cambridge: Harvard University Press.

Scanlon, T.M. 2008. *Moral dimensions*. Cambridge: Harvard University Press.

Shoemaker, David. 2007. Moral address, moral responsibility, and the boundaries of the moral community. *Ethics* 118: 70–108.

Shoemaker, David. 2011. Attributability, answerability, and accountability: Toward a wider theory of moral responsibility. *Ethics* 121: 603–632.

Shoemaker, David. 2015. *Responsibility from the margins*. Oxford: Oxford University Press.

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.

Smith, Angela M. 2005. Responsibility for attitudes: Activity and passivity in mental life. *Ethics* 115: 236–271.

Smith, Angela M. 2008. Character, blameworthiness, and blame: Comments on George Sher's *In Praise of Blame*. *Philosophical Studies* 137: 31–39.

Smith, Angela M. 2012. Attributability, answerability, and accountability. In defense of a unified account. *Ethics* 122: 575–589.

Smith, Angela M. 2015. Responsibility as answerability. *Inquiry* 58: 99–126.

Strawson, P.F. 1974. *Freedom and resentment*. London: Methuen.

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

Watson, Gary. 2004. *Agency and answerability*. Oxford: The Clarendon 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.