

## Precis for Responsibility from the Margins

David Shoemaker<sup>1</sup>

Published online: 2 February 2018

© Springer Science+Business Media B.V., part of Springer Nature 2018

The motivation for *Responsibility from the Margins* comes from the sense of ambivalence many of us tend to feel in appraising the responsibility of several types of “marginal” agents. These include individuals with clinical depression, obsessive-compulsive disorder, autism, psychopathy, intellectual disability, dementia, or terrible upbringings. For example, some individuals with autism can make good decisions about what to do once presented with all the facts, but they have a very hard time recognizing what some of those facts are, especially if they involve reading other people. Those in the moderate to late stages of Alzheimer’s dementia, alternatively, seem to lack the ability to make good decisions about their futures, but they nevertheless may still be *kind*, and this can be reflected in their attitudes or actions.

Previous theories of responsibility tend to view these individuals in an all-or-nothing fashion, as either wholly inside or (more likely) wholly outside the community of morally responsible agents. This cut-and-dried regimentation belies our actual more nuanced reactions to such people. Instead, our ambivalent feelings suggest that we think they are responsible in *some* ways but not others.

My aim in the book is to make sense of, and indeed to vindicate, this ambivalence. I undertake this task by seeing how far I can pursue a *quality of will* approach to responsibility, an approach originating in P.F. Strawson’s “Freedom and Resentment.” In that article, Strawson investigates our interpersonal reactive attitudes in order to see how in attending to them we might “recover from the facts as we know them a sense of... *all* we mean, when, speaking the language of morals, we speak of desert, responsibility, guilt, condemnation, and justice” (Strawson 2003: 91; emphasis in original). Strawson believes these attitudes are sensitive exclusively to the quality of will of our fellow members in the moral community. He also thinks that our array of reactive attitudes constitutes the framework of our

---

✉ David Shoemaker  
dshoemakr@gmail.com

<sup>1</sup> Tulane University, New Orleans, LA, USA

interpersonal lives, as our susceptibility to such attitudes is a crucial part of what makes our relationships *personal*. Consequently, this array has no justification external to this framework. Giving up the set of attitudes is also no realistic option, so worries about what theoretical conclusions about determinism imply for them are simply misplaced.

In the book, I am only concerned to explore the first of Strawson's theses, that attending to the reactive attitudes may help account for all we mean by responsibility and so forth. The second thesis—that there is no external justification, etc.—is very strong, and I set it aside. Nevertheless, if I turn out to be successful in defending the first thesis, I will have at least pointed the way toward a strategy for defending the second as well.

There are two problems with Strawson's own quality of will view. The first is that it's just unclear what he means by *will*. Indeed, there have been three distinct interpretations of it over the years: *character*, *judgment*, and *regard*. The second problem is that Strawson's is a monistic view—either you've got the relevant sort of will or you don't—in which case some agents don't fit easily under his rubric. For example, while those with some mental illnesses are appropriately exempted from resentment—grounding Strawson's excluding them from the moral responsibility community—they may still seem to exhibit poor quality of will in *some* sense. And this holds true of many other sorts of marginal agents, as I noted above. As Gary Watson puts it, “[A] psychotic can be hostile, a sociopath indifferent, a person under great strain can be rude, a woman or man ‘unfortunate in formative circumstances’ can be cruel” (Watson 2004: 228). My aim in the book is to develop a quality of will theory that can withstand these challenges.

In Part I, I develop a new *tripartite* theory of responsibility, built out of a full appreciation of our wide range of emotional responsibility responses. There are, as it turns out, three distinct qualities of will, each roughly corresponding to one of the interpretations of “will” noted above. Our emotional responsibility responses break down into three categories, each appraising the quality of a different type of will, and so, to the extent that the responses are apt, they implicate three distinct types of responsibility. In Part II, I explore what this tripartite theory predicts about the responsibility of *specific types* of marginal agents, thus testing the theory against our intuitions in these matters, rough hewn as they may be, once important empirical details about the agents have been filled in.

The first category of human emotional responsibility responses (*sentiments*) has as its paradigmatic pair *admiration* and *disdain* (I will here only talk about admiration to illustrate the basic ideas). When someone evinces a recognizable pattern of behavior or attitudes expressive of some aretaic ideal, we tend to respond to her with a feeling of uplift, an appraisal of her as admirable, and an action tendency of emulation. This is what I call *agential admiration*.<sup>1</sup> What we are appraising as admirable is the agent's excellent quality of character.

---

<sup>1</sup> In general, there are three features constitutive of a sentiment's syndrome: affect, appraisal, and action-tendency. They are primarily identified and differentiated by their action-tendencies. Also, I use the modifier “agential” here to identify just the familiar version of admiration (and disdain) we have toward agents.

But what does her character consist in? I take this to be the agent's *deep self*, a privileged subset of psychic elements representing the "true" agent, and in contrast to previous monistic accounts of the deep self, I argue for a pluralistic view. The deep self is located, I argue, in *both* the agent's cares (emotional dispositions) *and* her commitments (evaluative stance), and what we view as character traits are typically the agent's *care-commitment clusters*, groupings of cares and commitments that bolster and temper one another.

Agential admiration (and its many emotional family members, including disdain, contempt, hatred, shame, esteem, awe, and veneration) appraises the quality of an agent's character, defined in these terms. One has a reason of *fit* to feel it when its appraisal is correct. When someone has the capacity for a deep self, and so is fittingly susceptible to admiration, disdain, etc., then that agent is *attributability-responsible*, i.e., certain attitudes and actions may appropriately be attributed to her deep self. This is the first type of responsibility.

Agential regret and pride are the paradigmatic sentiments from which we can learn about the second type of responsibility. Agential regret (on which I will focus here) is a painful feeling appraising some of one's decisions or judgments as poor, and it has an action-tendency to change one's judgmental policies. It is fitting to the extent that its appraisal is correct. Its target is a judgment about the worth of reasons, something one may, in principle, be asked to answer for. But one can only answer for these judgments relative to what I call "instead of" reasons. That is, when I demand an answer from you about why you did X, implicit in my demand is the question, "Why did you do X *instead of not-X*?" To the extent that you lack any access to any reasons counseling against X (known as "normative competence"), you lack the capacity for this type of responsibility, called *answerability*. However, if you do have this evaluative capacity, you are eligible to be the fitting target of regret and its emotional family members (which includes disapproval, disappointment, irritation, approval, and appreciation).

Agential anger and gratitude make up the paradigmatic sentimental pair illustrating the third type of responsibility. Agential anger (on which I focus here) is the emotional constituent of resentment and indignation (which are the typical foci in the literature), and its syndrome includes feelings of heat and aggression, appraisals of agents as having *slighted* someone, and an action-tendency to communicate itself to the offending party, a communication which involves a demand for acknowledgment of what the offending agent did. Anger, that is, appraises one as having poor quality of *regard*. In order to be an apt target for anger, therefore, one must be able to regard others, to take them seriously. This is a function of empathy (both cognitive and emotional). At its core, then, regard is a *perceptual* stance, and so is not sensitive to evaluative judgment. To have the capacity to regard others in this way is to be *accountable* to them, which is the third type of responsibility.

Each of these three types of responsibility is independent from—and so unnecessary for—each of the others. Indeed, this is the precise source of our ambivalence in the marginal cases: Some of our sentimental responses are appropriate for some agents while others are not, and that's because these agents may have the capacities required for some types of responsible agency but lack

them for others. In Part II, then, I show how the tripartite theory predicts different types of ambivalence for different types of marginal agents, once we are armed with precise empirical details.

One could, for instance, be an accountable and attributable agent without being answerable, insofar as one is capable of the kind of empathy needed for regard and has the cares/commitments constituting character, but one is impaired for evaluative judgment. Indeed, as I argue, this is true of many adults with mild intellectual disabilities. What this means is that while two categories of sentimental responses may be apt for them (e.g., anger/gratitude and admiration/disdain), the third set (regret/pride, and their third-person analogues) is not.

Alternatively, one could be attributable and answerable without being accountable, given that one is incapable of regard (due to empathic impairments) but capable of certain evaluative judgments and has cares/commitments. This is true in different ways, I argue, for both psychopaths and those with high-functioning autism, once we get clear on the relevant empirical details. For them, a different set of sentimental responses is fitting: While anger/gratitude are unfitting, regret/pride and admiration/disdain may be fully appropriate.

In the remainder of Part II, I continue this approach in discussing the responsibility of those with clinical depression, OCD, and Alzheimer's dementia, as well as those who come from poor formative circumstances.

The tripartite theory is a qualities of will account of responsibility, and only in this pluralistic form can it capture all we mean when we speak of responsibility, desert, and so forth, as Strawson hoped.

## References

- Strawson, P. F. (2003). Freedom and resentment. In Gary Watson (Ed.), *Free will* (2nd ed., pp. 72–93). Oxford: Oxford University Press.
- Watson, Gary. (2004). *Agency and answerability*. Oxford: Oxford University Press.