Sanity and Moral Responsibility

1. The Deep Self View: Susan Wolf surveys the state of things so far:

What we learned from Frankfurt: Moral responsibility requires more than mere freedom of action. A hypnotized person’s actions may be free: She wants to cluck like a chicken, and nothing hinders her from doing so, so she does. And yet, she does not seem responsible for her action (rather, the hypnotist is). What is needed is freedom of will. This is not the mere freedom to act on our desires, but rather the freedom to govern the desires themselves.

Wolf calls this the “Deep Self View”. This is the view that, “in order to be responsible for one’s actions, one must be responsible for the self that performs these actions. ... [I]f we are responsible agents, it is not just because our actions are within the control of our wills, but because, in addition, our wills are not just psychological states in us, but expressions of characters that come from us, or that at any rate are acknowledge and affirmed by us.” (375)

In short, a responsible agent is not simply one who acts on desires—but is one who possesses a deeper “self” who is able to reflect on her desires, and shape herself into the kind of person she wants to be by directing and controlling those desires.

This view explains why the following individuals are not responsible for their actions:

- Kleptomaniacs, addicts, the brainwashed, etc. are not morally responsible for their actions because those actions do not flow from their deeper self. She says their “selves are alienated from their actions” (376)

- Toddlers, the severely mentally disabled, most animals, etc. are not morally responsible for their actions because they do not HAVE deeper selves.

What we learned from Strawson: However, the worry arises: How can we shape our “selves” if the motivations and reasons for doing so will themselves be a product of factors outside of our control (genetics, upbringing, etc.)? Wolf writes,

“Though I can step back from the values my parents and teachers have given me and ask whether these are the values I really want, the “I” that steps back will itself be a product of the parents and teachers I am questioning.” (378)

If the causal chain goes back to a time before we were born, it seems that we can never get that initial, ultimate “foothold” that is needed for ultimate control of our “selves”.

What we learned from Chisholm: It seems that the only way we could be ultimately responsible, or in charge of, who we ARE is to be the FIRST cause of our “selves”.

But then, Wolf says, the only way to be TRULY responsible for what we are is to be self-causing, which is impossible/nonsense.

The Deep-Self View: If nothing short of self-creation will ease Strawson’s worry, perhaps that worry is mis-guided. Let’s take a closer look at the Deep Self View:

**The Deep-Self View:** An individual is morally responsible for some action if and only if she is able to govern that action by her desires and to govern her desires by her deep self.

Unfortunately, there seem to be serious counter-examples to this view:

**JoJo:** JoJo is the son of Jo, an evil sadistic dictator. JoJo’s entire upbringing and education consists of shadowing his father on his daily routine, which is filled with many horrific deeds (e.g., imprisoning, torturing, and murdering citizens on a whim). Having had Jo as his sole role model, JoJo grows up to be an evil sadistic dictator himself, much like his father.

Note also that JoJo desires to torture and execute citizens, and has also reflected upon these desires and concluded that *these are the desires that he wants to have*.

Is JoJo responsible for his actions? [What do you think?] The Deep-Self View entails “yes”. But, Wolf believes that JoJo is NOT responsible (for many of the same reasons that the Situationists cited). She writes,

“As light of JoJo’s heritage and upbringing—both of which he was powerless to control—it is dubious at best that he should be regarded as responsible for what he does. It is unclear whether anyone with a childhood such as his could have developed into anything but the twisted and perverse sort of person that he has become.” (379-380)

Wolf then points out that, it seems that, if JoJo isn’t responsible for his actions, then NONE of us are responsible for our actions. For, ALL of our deepest selves are the products of our genes and upbringing, just as JoJo’s is. (Strawson’s worry)

**The Gap to be Filled:** Wolf concludes that the conditions cited by the Deep Self View may be necessary for moral responsibility, but they are not sufficient for it (i.e., the “if and only if” should be replaced with a simple “only if”). What is needed, she thinks, is something that we can tack onto the DSV so that we can restore the “if and only if”. 
We've been very concerned with how to gain ultimate POWER or CONTROL in order to fill in the gap. But, Wolf believes that that project is mistaken:

“But not all the things necessary for freedom and responsibility must be types of power and control. We may need simply to be a certain way, even though it is not within our power to determine whether we are that way or not.” (380)

2. The Sane Deep-Self View: Wolf proposes to fill in that gap with a condition of sanity.

The Sane Deep-Self View: A is morally responsible for some action if and only if (a) A is able to govern that action by her desires and to govern her desires by her deep self, and (b) A’s deep self is sane.

This proposal explains why JoJo is not responsible for his actions. Simply put, JoJo is insane. To be “sane” in Wolf’s sense is to know what one is doing AND to have beliefs that correctly correspond to the way the world is. JoJo’s beliefs do not match up correctly with the world in this way. He is mistaken about the nature of right and wrong.

Same goes for slave-owners (e.g., Thomas Jefferson), Nazi war criminals (e.g., Adolf Eichmann), or even your racist or chauvinist grandpa, etc. Their deep selves are disconnected from reality. They fail to see their own mistake. In effect, these people are no better off than a hypnotized or brainwashed person—except their “brainwashing” was a lifetime of being immersed in a culture that impressed mistaken values upon them. Perhaps it is a stretch to call them insane in the ordinary sense, but Wolf says,

“Like JoJo, they are, at the deepest level, unable cognitively and normatively to recognize and appreciate the world for what it is. In our sense of the term, their deepest selves are not fully sane.” (383) “If we think that the agents could not help but be mistaken about their values, we do not blame them for the actions those values inspired.” (382) In other words, “if we believe they are unable to recognize that their values are mistaken, we do not hold them responsible for the actions that flow from these values, and if we believe their ability to recognize their normative errors is impaired, we hold them less than fully responsible for the relevant actions.” (382-383n)

[So, perhaps Wolf would NOT give slave-owners, Nazis, and racist grandpas a pass by claiming that they are not morally responsible for their actions? Consider Jefferson. Surely he (unlike JoJo) COULD HAVE formed correct beliefs about reality. In fact, it seems likely that he did IN FACT believe that slavery was wrong and so DID have correct beliefs about reality. Therefore, he IS morally responsible for his actions. What do you think?]
If JoJo isn't responsible for his actions merely because he is not ultimately in control of his deepest self, then we are no better off. For, we aren't ultimately in control of our deepest selves either.

What separates us from JoJo is not that we have more control, but rather that we are more sane. We do have a better understanding of the nature of right and wrong, so we are responsible for our actions.

**3. Objections**: Some criticisms:

1. **Sanity is a Matter of Luck**: Still, Wolf uses the word “lucky” several times. We are “lucky” that we do have an accurate understanding of right and wrong. In short, is it true that whether or not one is “sane” is a matter of luck (i.e., it is outside of our control)? If so, it is unclear why we would be responsible for our actions even if we are “sane” in Wolf’s sense.

   **Reply**: Frankfurt wanted the mere ability to revise ourselves. But, that isn’t enough, because JoJo has this but does not seem responsible. Chisholm wanted the ability to create ourselves. But, this seems impossible; and not really even desirable. We all have to start with some initial conditions that are just handed to us.

   What is wanted, is something in between: The ability to correct or improve ourselves. Not merely revise, but revise in light of an accurate understanding of right and wrong. In that way, we do create our deeper selves—but not out of nothing. Rather, out of something (namely, whatever genes and upbringing we are handed at birth). This may not get us responsibility in the ultimate, metaphysical sense—but, it seems to be enough to get us responsibility in the moral sense.

2. **Are We Sane?** How do we know that we are sane? We might be mistaken about morality, just like slave-owners or Nazis were. (For example, some have said that, someday, future generations will look back on our unquestioning slaughter and consumption of animals as appalling.)

   **Reply**: Wolf isn’t claiming that we have everything right. Obviously, lots of moral topics are still very controversial. But, surely we can say that our moral conclusions are closer to the truth than those of, say, the Nazis. We must simply do the best with what we have. We probably WILL discover that we are still wrong about a bunch of things.

3. **Is Anyone Ever Morally Responsible for Wrongdoing?** Imagine anyone who acts wrongly. If they do not perceive their act as wrong, then they are apparently not “sane”, on Wolf’s view, because it seems like “the most horrendous, stomach-turning crimes could be committed only by an insane person.” (387)
So, is anyone EVER responsible for horrendous wrongdoings? It seems like very few people (if any) would be (namely, only those who do wrong, and correctly BELIEVE that they’re doing wrong).

Reply: Sanity requires the ABILITY to correct or improve one’s self. Most people have this ability, even if they are mistaken in the moment about whether they are acting wrongly.

Rebuttal: What sense of “ability” is Wolf using? If determinism is true, and I do not in fact improve myself, in what sense is it true that I nevertheless had the “ability” to improve myself?