

Kenny Jones
Tutorial Paper # 2
10/7/2014

Examining the Sane Deep Self View

In *Sanity and the Metaphysics of Responsibility*, Susan Wolf attempts to build a system of moral responsibility off a compatibilist theory she describes as the deep-self view. To the deep-self view she adds the condition of sanity: not only must the agent identify with his own desires in his deepest self, but he also must be able to distinguish between right and wrong actions. Her sane deep-self view, as she calls it, rests upon the disputable existence of objective values, but even given these meta-ethical absolutes, her theory presents numerous problems. The main issue with the sane deep-self view is that it gives new unfamiliar definitions to the terms responsibility and freedom. Moreover, it falls short in this re-defining, because under this theory the only responsible actions are those that are made in the right way and Wolf does not make it clear why this kind of responsibility would even be relevant.

The deep-self view that Wolf expands on is a mixture of the theories of Frankfurt and Watson. The fundamental principle behind the deep-self view is that responsible agency “involves something more than intentional agency” (375). An agent cannot just be in control of his will, but must, in some sense, have a way of modifying or affirming this will and the resulting actions that come from it. This ability for internal revision defines the deep-self in the deep-self view. What differs between Frankfurt and Watson are the ways in which they define what a deep-self is: Frankfurt focuses on the relationship in between first order desires and second order volitions while Watson is concerned with distinguishing between “ ‘mere’ desires... and desires that are values” (374). However, both claim that their deep-self theory is compatible with determinism, and that, even in a situation in which an agent is completely causally determined, that agents can still have free will and hold responsibility for their actions.

Wolf identifies with much of the deep-self view. However, unlike Frankfurt and Watson, she does not believe that the deep-self view can be compatible with determinism. Her argument here is simple: that if determinism is true, then an agent's deepest self must be causally determined, and thus the agent cannot be responsible for their deepest self. To illustrate this point, she comes up with the example of JoJo (379), who commits atrocities as a dictator because he learned directly from his corrupt father. Under the deep-self view, JoJo would be responsible for his actions, because his deepest self approves of his actions, but Wolf makes the argument that because JoJo could not have developed into anything else that JoJo cannot be held responsible for his actions.

However, the example of JoJo leads one to consider what distinguishes his case from everyone else's: surely our deepest selves are also determined to the extent that JoJo's is. Wolf answers this objection by explaining her sane deep-self view. To the deep-self view she adds the qualifier that an agent must be sane in order to be held responsible. She defines sane with two conditions: (1) that an agent is aware of his actions, and (2) that the agent is aware that his actions are in a given situation right or wrong (381). From her paper *The Reason View* we can see that this concept is also defined as an agent who is able to identify with "the True and the Good" (78). When an agent can be said to have known what the correct action would be and to then perform that action, that agent has acted in a responsible manner. Simply put, the agent must do "the right thing for the right reasons (81). With this principle defined, it is clear to see why Wolf believes there is a difference between JoJo and us. JoJo cannot identify with the right action, he has a skewed view of good, and therefore he cannot be responsible for his actions even if he identifies with his deepest self. Responsible agents on the other hand, have the ability to

distinguish between right actions and wrong actions, and therefore can be responsible for actions that they commit.

For Wolf, what makes an agent responsible is their ability to view the world in a sane manner combined with being able to “correct” (385) themselves. In the deep-self view, a person can self-revise themselves to become the person that they want to become. Wolf argues that this is only relevant when the person that they want to become is in an ethical sense the person that they should become, the person that it is right to become. Therefore, when a sane deep-self agent self-revises, if it acts with the True and the Good in mind it actually self-corrects, and in this way has the ability to identify itself and its actions with the right, and so assume responsibility.

This concept of self-correction is how Wolf attempts to get around the problem of self-creation. Wolf believes that each of our deepest selves are given to us without our consent (378,380), but that the concept of self-creation is “logically impossible” (380). How she protects the sane deep-self view from determinists is by drawing from a source of morals that resides outside of the determined agent. For her, an agent can only be responsible when they are determined to have to capability to identify with the True and the Good. My main issue with this is that her theory defines freedom and responsibility in a new way, and she briefly notes this: in the sane deep-self view an agent cannot be “metaphysically responsible for [themselves] ... but rather morally responsible for [themselves]” (385). However I disagree that this definition of responsibility is relevant in any way, and that she offers not justification of separating moral responsibility from metaphysical freedom. Her theory says that agents can be “moral responsible”, but it seems to follow that these agents are either determined in a metaphysical sense to be free (in relation to their moral responsibility), or that some agents hold moral

responsibility even though no agents have a truly free-will, which does not seem to match our intuitive definitions of responsibility and freedom.

One reality of the sane deep-self view is that a person who has the capability to act in a morally responsible way was determined to both commit that action and to have the upbringing that allowed them to make that action. Conversely, a person who does *not* have the ability to act in a morally responsible way was causally determined to never get the opportunity to act responsibly. This definition of responsibility does not seem to be based on free will but rather efficacy. This view of responsibility, as Wolf points out, is a moral one, supported only by “widespread intersubjective agreement and the considerable success we have in getting around the world and satisfying our needs” (386). Practically, on an everyday level, this type of responsibility seems like it might be useful in determining punishment or culpability. It tries to answer the question of who do we find to be responsible and why. However, this definition of responsibility offers no answer to larger questions, such as, who *should* we find responsible. Moreover, Wolf’s definition of moral responsibility does not really address the issue of free-will. If it does, it identifies itself with the deep-self definitions of free will – where free will can be found within the deep self of the agent. However, Wolf is not satisfied that this claim can hold up to thesis of determinism, and when she justifies her views, she makes no mention of how the sanity of the agent relates to his metaphysical freedom, only his moral responsibility. So even though Wolf gives us a definition of responsibility that we can use on a practical level, it is not the kind of responsibility that one searches for when looking for answers about the nature of free will. But even within Wolf’s definition of moral responsibility I find inconsistencies with the thesis of determinism and of how we practically judge people to be responsible today.

In *Freedom within Reason* Wolf gives examples of how actions can be labeled with moral praise and moral blame. Curiously, for many of these conditions she adds the qualifier that the agent *could have* done something else, or *had the ability* to choose another thing. The reason view, for most situations, requires “that the responsible agent have the ability to do otherwise” (69). These qualifiers seem to be directly at odds with the thesis of determinism. It seems impossible to say, when all events are causally determined, that an agent committed an act for a specific reason, but at the same time had the capability to do *something else*. I argue that it is a useless distinction to say that an agent could have done something else: an agent, in terms of causality, can only have ever done one thing, the thing that they do. In a deterministic scope, these qualifiers are meaningless. When we then analyze Wolf’s theories with this fact in mind, we can see some problems with this definition of responsibility. Take for example Wolf’s conception of moral blameworthiness. An agent can only be blameworthy when an agent “is able to do the right thing for the right reasons” (87), but for some reason, by somehow defying causality, does not commit this action. Wolf’s justification for this are unclear and shaky. She talks of how if persons with a similar history all fail to develop a particular desirable trait, “we tend to conclude that their history renders them incapable of developing it” (86) but that “if... some persons develop badly and no apparently relevant differences are found in the hereditaries and environments of the two groups” (86) that we might say that these individuals were responsible for their actions. To me, this sounds like a very basic and misguided application of the theory of determinism. In fact, looking at this situation with the full theory of determinism in mind, it is clear that there is in fact, no distinction between the two groups of people in her example, every single one of the persons was causally determined to do exactly what they did, but rather a distinction in our understanding of their causality, which can become very complex.

But a lack of understanding in causality does not indicate moral blameworthiness, only that we do not have the ability to gain perfect knowledge of every situation. Following this line of argument therefore, we can say that because no agent every has the ability to do something else, no agent is ever to blame for their actions; that the only responsible actions are praiseworthy actions, and that an agent can only be responsible if he does the right thing.

So, upon looking at this new kind of responsibility further, we must ask the question of what use does it have? Wolf's responsibility, in light of the thesis of determinism, only appears in agents who do right things for the right reasons. Every good action is a responsible one, but that's it. There certainly seems to be no use for responsibility defined in this way and it runs counter to our everyday sense of responsibility. We typically think of responsibility as coming from inside of an agent, but this definition of responsibility turns this notion on its head by arguing that an agent's responsibility comes from outside of him; if a self aligns with the True and the Good, then that self can be responsible. Moreover, with this definition we can no longer use responsibility to justify punishment. Either we punish people who we do not believe are responsible for their actions, or there is no punishment for any action committed, since the only crimes committed are those that are committed in ignorance. Wolf may be right that the deep-self view may need something more to strengthen it against the arguments of determinism, but her addition of sanity to the deep-self theory only seems to obfuscate the relationship between responsibility, free will, and determinism, rather than clarifying anything.

Works Cited:

Wolf, Susan R. "Sanity and the Metaphysics of Responsibility." *Free Will*. Oxford: Oxford UP, 1982. N. pag. Print.

Wolf, Susan R. "The Reason View." *Freedom within Reason*. New York: Oxford UP, 1990. N. pag. Print.