Michael Lucana
PHIL 715
Seminar in Philosophical Writing
Dr. Shelley Wilcox
Originally Written: Fall 2012
Edited and Updated: January 2022

Knowledge & Moral Responsibility

In this paper I will defend the "real-self" account of moral responsibility from two major structural objections, one which is called the "whim problem," and another which is called the "source problem." I plan to do the following:

(1) Address the debate between the two competing views of determinism and free will as they have been articulated by Van Inwagen in order to set up the dilemma that "real-self" accounts of moral responsibility attempt to resolve.

(2) Give an account of Frankfurt's "authenticity" version of moral responsibility as well as Watson's "authoritative" version, both of which can be considered "real-self" accounts of moral responsibility.

(3) Establish that both the "authenticity" and "authoritative" accounts of moral responsibility have two major underlying structural problems, one being the "whim" problem, and another being the "source" problem.

(4) Attempt to resolve the "whim" problem first, by situating it within Frankfurt's conception of "wanton-ness" and denoting the fundamentally epistemological character of this problem.

(5) Situate the "source" problem within a similar epistemological concern.

(6) Finally, I will consider a methodological question to my solution and the possible implications of situating accountants of moral responsibility in epistemological concerns.

(1) Determinism & Free Will

Determinism is the view that a state of affairs in the world at t1 is determined by the state of affairs in the world at t0, viz. a causal law L. If we take free will to be an agent's ability to do otherwise, then it quickly becomes clear that we cannot hold this conception of free will in conjunction with determinism. For, if an agent has the ability to do otherwise at t1, then this

means that the state of affairs in the world at t1 is not determined by the state of affairs in the world at t0. (Van Inwagen 1975)

Say we reject this view of free will in favor of determinism. If so, then we can't be morally responsible for our acts since they come into being within the framework of a causally determined system. If an agent commits some act then, we don't judge that agent to be responsible if they couldn't have done otherwise when committing that act. But this is exactly what determinism, as spelled out above, entails at a first glance, non-responsibility for our acts.

I have given a brief description of the dilemma that we are faced with if we accept determinism in favor of the normally construed conception of free will as an agent's ability to do otherwise, which is, the possibility of losing all sense of moral responsibility for our actions. Next, I will present the real-self account of moral responsibility, with emphasis on Frankfurt's "authenticity" version.

(2) "Real-Self" Accounts of Moral Responsibility

"Real-self" accounts come in two main flavors. (Rasmussen 2003) Those of the "authentic" variety and those of the "authoritative" type. I will mainly focus on Frankfurt's "authentic" account and finally make some notes on Watson's "authoritative" account.

If we think of free will, and by extension moral responsibility, as being dependant on an agent's ability to do otherwise, then, having rejected it in favor of determinism previously, we are inevitably reduced to treating both as metaphysical "miracles," or at worst, "mysteries," in a vague theological sense. (Van Inwagen 2000) However, there is a "counterfactual" objection, that has previously been proposed by Frankfurt (1969) which surmounts this problem and shifts our understanding of moral responsibility towards a view of it as being founded on a concept of "identification." (Frankfurt 1971) The following is my interpretation of it:

> Say, Smith has made it up in his mind to kill Bob at t1. However, at t0, unknown to Smith, some scientist had installed a chip in Smith's brain so that, just in case he didn't carry out the deed (killing Bob) at t1, the chip would activate and make it so that Smith killed Bob nonetheless. As it happens, Smith kills Bob at t1 and the chip does not need to activate. So, even if Smith had chosen to not kill Bob at t1, the chip installed at t0 would have made it so that Smith still would have killed Bob at t1. In other words, Smith would not have been able to do otherwise but kill Bob at t1.

If determinism is true, then Smith isn't morally responsible for killing Bob at t1, because he couldn't have done otherwise, yet *we still think he is responsible in some way*. It can't be on account that he could have done otherwise than he did because either he killed Bob based on his own deliberation, or the chip would have induced him to. As Frankfurt posits, we still could say that Smith is morally responsible because he "wanted to want" to kill Bob, or to put it another way, Smith identified with his desire to kill Bob. But what does this mean?

First, it will suffice to make a distinction between 1st order and 2nd order desires. A 1st order desire can be formulated like this: S desires that S does X, with S being a moral agent, say Smith, and X being any such act that S has a desire to do, such as killing Bob. So an interpretation of this form can be:

> Smith desires that "Smith kills Bob," or, put another way, Smith has an occurrent desire to kill Bob.

But 1s order desires are a dime a dozen. Smith's desire to kill Bob can be one of many 1st order desires, a good deal of which Smith probably never acts upon. So these 1st order desires aren't necessary for moral responsibility. After all, if one subscribes to the view that any and all desires that happen to cross our minds on any given day and time imply moral responsibility as if we had actually acted on those desires, then there would not be a single person who would not be guilty of at least one thing or another. It is really in the 2nd class of desires that reflection and the possibility of responsibility come into play.

A 2nd order desire can be formulated like this: S desires that "S desires that S does X", or to give an example, Smith desires that he have the desire to kill Bob. But there are two kinds of 2nd order desires, one in which an agent simply wants a desire (Smith wants to want to kill Bob), and another, much stronger version, in which an agent identifies with that desire (Smith's will is congruent with his wanting to kill Bob). The latter we can call a 2nd order volition in order to make it distinct from the other kind of 2nd order desire.

It is within this 2nd order conception of identification, or congruence, that responsibility can be situated in. Say, in a case of a 2nd order desire, Smith has a strong desire to want to know what it is like to want to kill someone, say Bob. His desire isn't necessarily to kill Bob, only to want to know what it feels like, in a distanced way, perhaps purely for theoretical reasons. Butin

a stronger version, that of a 2nd order volition, Smith's will identifies with the desire to kill Bob, that is, Smith cannot want to do anything but kill Bob. If there is any deliberation involved here that could even be said to be related to a could-have-done-otherwise principle, it would merely involve Smith's deliberating, after the fact, whether he ever really identified with the desire or not. But it is important to remember, once this "identification condition" were met, Smith would not be able to do otherwise than kill Bob.

I have given a quick summary of the "identification" principle that Frankfurt places responsibility in. It bypasses the could-have-done-otherwise principle while still admitting that an agent, so long as they identify with their acts, is morally responsible for those acts they've identified with.

Rasmussen considers Frankfurt's account of moral responsibility to be of the "authentic" variety in that an agent is identifying with some internal (i.e. authentic) desire irrespective of any idealistic objections. To put it another way, the desire does not have to be rational, however, the agent's deliberative process in which they come to identify with that desire can be, and, as I'll discuss later, it is a condition in all "real-self" accounts of responsibility that an agent must be minimally rational in this deliberative sense also.

The essential structure of Watson's account (1975) is still very much of the same form as Frankfurt's, except that, in Watson's account, Smith only holds himself responsible for killing Bob if the killing of Bob is a desire that runs counter to an external conception of right and wrong that Smith identifies with. Or, to explain it another way, Watson's "authoritative" account of responsibility requires that an agent dis-identifies with an act, that is, the act does not accord with their idealistic (i.e. authoritative) notion of the rightness of that act.

So we have looked at both "authentic" and "authoritative" versions of the "real-self" account of responsibility and observed that an agent is only responsible for their actions if they either identify with that act (in Frankfurt) or dis-identify with it (in Watson).

(3) Structural Objections to "Real-Self" Accounts

Up until now, I have not treated "real-self" accounts of either "authenticity" or "authoritativeness" in any critical sense. I have simply attempted to outline what their essential structures demand for an agent to be held morally responsible. I will now present the "whim"

objection and the "source" problem, both of which seriously undermine any "real-self" accounts of moral responsibility that we can come up with.

The "whim" objection consists of the following: An agent performs an act but they neither identify with it nor dis-identify with it, yet we still would want to hold them morally responsible for it. We can look at a case such that:

> Smith kills Bob but when he did it, he neither identified with, that is, admitted that the act aligned with his will, nor did he dis-identify with the act, that is, repudiated it as disanalogous with his ideal of right and wrong. But we still found Smith to be morally responsible for the act.

The "source" problem can be divided into a weak and strong form. I will begin with the weak form:

> Say, Smith kills Bob and he identifies completely with the act. But, unknown to Smith, he has a brain tumor that has made it so that he still has the power to reason and deliberate about his own acts and the accordance of his will with them, but at the same time, his will has been altered so that at say, t0 when he did not have the brain tumor, he would not identify his will with the desire to kill Bob as he now does at t1.

The strong version of the "source" problem is of this form:

> Say, Smith kills Bob and identifies completely with the act. But, again unknown to Smith, a sophisticated brain surgeon has at t0 manipulated Smith so that at t1 although Smith, completely using his reason and deliberation, has decided that the killing of Bob is in accordance with his will, is actually acting upon a simulated desire that is in actuality foreign to his own motivations had the brain surgeon not intervened.

Both of these versions of the source problem serve to show that an agent can be considered to be in the correct state of mind to deliberate on an act's accordance with their will, and they could perform the act of identification without a problem, but we would not find them morally responsible for their actions due to the fact that they are not the source of the fundamental desire that they act upon. And if this is the case also, then identification is not necessary for ascribing responsibility to an agent.

(4) The "Whim" Problem and its Relation to Identification

When acting on a whim, an agent neither identifies nor dis-identifies with their desire. As Rasmussen argues, using an "authoritative" context, this is essentially the same as an agent being "psychologically unable to so govern [their] will because [they] consider the issue too trivial to

engage [their] values." (Rasmussen 2003) Another way to say this is that they do not consider it worth deliberating whether or not a particular desire to act coheres with their will. This places an act like the murder of a human being on the same moral level as that of absentmindedly swatting at a gnat that hovers by one's face.

The first way to dismiss cases of whimsical behavior is due to the rate at which they are likely to occur within the psychological life of an agent. Take Smith killing Bob. Let's say Smith kills Bob on a purely whimsical desire, and doesn't consider it worth his time to deliberate whether or not that act was in accordance with his real desires, he just did it. But Smith has deliberated whether an act coheres with his will in cases prior and will do so afterwards, just in this case, he hasn't. One single case of an agent being inconsistent does not completely dismantle a theory of established behavior. So Smith is not let off the hook in this case. We might need to consider other reasons why Smith did not consider the desire to kill Bob worth deliberating about, but there is no reason why this should weaken the theory of responsibility attribution we have already formulated.

The second way to resolve this issue of whim is to consider if a particular agent always acts on whimsical desires. There is actually such a closely related concept of an agent that always acts this way as theorized by Frankfurt in his case of a "wanton." To Frankfurt, when a person acts, the desire by which [they] are moved is either the will [they] want or a will [they] want to be without. When a wanton acts, it is neither." A wanton has this lack of concern because they lack the capacity for reflection on the congruence of their acts with their wills. They simply do things in an indifferent manner, with no act ever being worth engaging critically in the way required by "real-self" theory.

But this seems to be very similar to the description of someone who acts on a whimsical desire, someone who according to Frankfurt fails to identify or dis-identify with an act because they are "psychologically unable to." But if an agent is psychologically unable to consider if their acts coherence to their will then, they probably should not be considered to be sufficiently competent enough to be responsible for their actions.

This brings us up to this concept of "identification" which to me seems to be the underlying area of weakness in all "real-self" accounts of responsibility. "Real-self" accounts of

responsibility are unclear what is required in order for an agent to identify with a given 1st order desire aside from the condition that they be a rational being and be able to deliberate if, at least not upon what course of action to take, but on whether or not they identify with the action they have taken. This is similar to Rasmussen's condition that any agent who's acts we are considering should be "well-behaved," or, to put it another way, mentally healthy.

To make a judgment like identification of one's will with their desire requires, in my view at least, that an agent's ability to identify with a particular desire is conditional on their being in a strong epistemic position to judge their relation to that desire as being the case. But if this is so, then the problems we are considering are not really structural but rather epistemological in nature.

(5) Knowledge and its Relation to the Source Problem

A typical hypothetical skeptical scenario has this form:

> Smith knows that he is sitting in front of the fireplace in his house. He knows this because he is mentally healthy and he has strong perceptual data to back up his belief. However, unknown to Smith, an external manipulator, say, a malevolent demon, has made it so that Smith is actually in a simulacrum that simulates every single aspect of Smith's house down to the tiniest detail. So, even though Smith may be in good mental health, and he has strong perceptual data to back up his belief that he is sitting in front of the fireplace in his house, we cannot judge that he actually knows it.

This is an example of a strong skeptical scenario that I've just outlined. We can admit that Smith, no matter how strong of an epistemic position he is in, will always believe, but not actually know, that he is in front of his fireplace. There are weaker forms of a skeptical scenario such as this:

> Smith goes to visit Bob. When he gets to Bob's house, Bob's family tells Smith that Bob is sick. Smith doesn't get a chance to see Bob, but takes the family's word that Bob is sick and Smith is in good mental health. Smith doesn't have any reason to doubt what he's been told either. Bob's family are nice folks and have always been friendly and honest with Smith. However, in this instant, Bob is not really sick, he just doesn't want to see Smith. So, even though Smith may be in good mental health and he took the word of Bob's family as good evidence, we can't judge that he really knows that Bob is sick.

Belief has the following sort of general structure to it: There is a relationship that an epistemic agent has to a certain belief, namely that he believes it. That belief is founded one evidence that the epistemic agent takes to be sufficient for formulating a given belief. Both of

these skeptical scenarios bring in the following condition: There is an external agent that undermines this relationship by presenting evidence to that belief that the epistemic agent will take to be reliable. But both of the cases that we considered in the "source" problem fit this structure also. There was a relationship that held between the moral agent and the act, they identified with it, that is, they believed it cohered with their will. In other words, they had no plausible reason to believe that they were not the source of the decision to identify with the act to be performed. There was an external agent that undermined this relationship by altering some aspect of the agent's psychology so that the agent believed that they were the source of making the decision to identify with that act to be performed.

The " nbsource" problem then, both in its weak and strong form relies on a skeptical scenario obtaining that an agent is not aware of that causes the agent to misidentify the object of belief, in one case, that they are sitting in front of the fire, in another, that a particular act coheres with their will.

However, in the weak version of the skeptical scenario presented above, Smith is always in a position to strengthen his epistemic position. It is always possible that he could investigate further and find out that Bob's family was lying to him. Similarly, in the weak version of the "source" problem, the cause of Smith's mis-identification is a brain tumor that he can have diagnosed by a doctor and can have treated for in some way. Or to put it another way, in the weak version of the "source" problem, Smith can always be made aware of the condition that undermines his ability to identify clearly with an act.

The strong version of the skeptical scenario cannot be resolved so easily. That is why it is so intriguing in the history of philosophy. There are two ways that the strength of it, and the strong version of the source problem, can be mitigated however.

First, as DeRose has posited, we can treat strong skeptical scenarios as being so far beyond the scope of normal everyday scenarios that while we can admit they are possible in some sort of highly restricted sense, they can't be practically treated as the actual case in virtue of their hypothetical construction. (DeRose 1995)

Second, we can treat skeptical scenarios as exercises that allow us to see which beliefs are undeniable and beyond doubt, much as Descartes implemented them. For Descartes, the

skeptical scenario, as exemplified in the evil demon hypothesis, was an exercise for him to remove all doubtful beliefs and begin as if from scratch, so that he would have a stronger basis by which he could allow himself to affirm beliefs. Similarly, though we might not be able to resolve the strong "source" objection, we can consider it instead, as an exercise for a moral agent to determine how strongly a certain act conforms to their will. Does it do so even if it were the case that a brain surgeon had implanted the volition in their mind? If it does so, if an agent can answer that it would, despite a strong skeptical scenario such as that, then this would make it an undeniable affirmation that the act cohered with their will.

(6) A Methodological Question

The final question that I will consider is one of a methodological nature. Can we bring in epistemological considerations into the question concerning moral responsibility?

I have only brought up epistemological considerations in order to better understand the major structural objections for "real-self" accounts of responsibility within their correct contexts which are epistemological and not necessarily related to the structure of the theories in question. This being the case, it could be that any "real-self" theory, in order to withstand objections like those generated by the "source" problem, needs a more robust articulation of what "identification" is and how it is constituted in a well behaved moral agent. However, I maintain that "identification," and an agent's relation to it, is structurally epistemological.

While repositioning our vantage point to the "source" problem so that it is construed as epistemological in nature we have not resolved the "source" problem, we have simply identified why it is such a strong objection and what we have also done is show that the book is not closed on the problem. In the field of epistemology there is still strong debate as to how the skepticism is to be resolved, if it can be at all, with various solutions constantly being formulated to the present day. When and if the skeptical problem is ever solved, it is my position that it will have a direct impact on the "source" problem.

Works Cited

Gary Watson "Free Agency." Journal of Philosophy Vol. 72, No. 8, 1975 pp.205-220

Harry Frankfurt “Alternate Possibilities and Moral Responsibility” The Journal of Philosophy Vol. 66, No. 23 Dec. 4, 1969 pp.829-839

Harry Frankfurt “Freedom of the Will and the Concept of a Person” The Journal of Philosophy Vol. 68, No. 1 Jan. 14, 1971 pp.-20

Keith DeRose “Solving the Skeptical Problem” The Philosophical Review Vol. 104, No. 1 Jan, 1995 pp.1-52

Kasper Lippert-Rasmussen “Identification and Responsibility” Ethical Theory and Moral Practice, Vol. 6, No. 4 Dec. 2003 pp.349-376

Peter van Inwagen “The Incompatibility of Free Will and Determinism” Philosophical Studies 27 1975 pp.185-199

Peter van Inwagen “Free Will Remains a Mystery.” Philosophical Perspectives 14: 2000 pp. 1-19