Robert Durfee

David Balcarras

24.00 – Topic III

April 10, 2018

**Circular Theory-Theory**

Sarah Paul argues for her theory-theory of self-knowledge in the paper “John Doe and Richard Roe.” This theory rejects the premises of the Cartesian and inner-sense theories of self-knowledge in favor of a modified Rylean argument. With this argument, Paul seems to come to a satisfactory, economical conclusion as to why we sometimes fail to correctly know our own mental state yet seem to have more insight to our own attitudes and emotions than others’. In this paper, I will outline Paul’s theory-theory and how she’s justified in rejecting the Cartesian and inner-sense premises. However, I will continue by showing how the theory-theory results in a circular method for ascribing mental states to oneself where one’s interpretation of evidence is filtered through one’s current mental state.

Paul begins her argument by listing the claims of “common-wisdom”:

[No Evidence]: When a subject ascribes a mental state *M* to herself, it is not on the basis

of evidence or grounds for thinking that she is in *M*.

[Infallible]: If a subject sincerely believes herself to be in some mental state *M*, then she is

in *M*.

[First-Personal]: The method a subject normally uses to ascribe mental states to herself

cannot be used in the same way to ascribe mental states to other people.

Although at first glance all these claims seem obviously true, Paul introduces a scenario, proposed by Ludwig Wittgenstein, which brings this into question. Imagine you, and everyone around you, has been given a box. Everyone is instructed to refer to the contents of the box as a “beetle,” but the contents may or may not actually be a beetle. Since you cannot observe the contents another’s box, you have no way of knowing what “beetle” means to anyone but yourself. If you think of the boxes as individuals and the “beetle” to be an individual’s mental state, it stands that you could never know someone’s attitudes or emotions (Rosen 242). Yet there are several examples where others’ emotions are clearly knowable. Such as knowing an individual who just stubbed their toe is in pain.

Thus, Paul proposes to reject the claims of “common-wisdom,” starting with [Infallible] using the following common situation. Recall a time when a parent or sibling has denied the fact that they are angry when it is clear to you that they are as they exhibit noticeable physical indications (Rosen 242). This brings to question whether someone always knows their own mental state infallibly. As such, Paul points to the inner-sense theory which tries to fix this situation by weakening [Infallible] with:

[Privileged]: Beliefs about one’s own mental states are more likely to be true and amount

to knowledge than beliefs about others’ mental states.

However, Paul notes that an implicit requirement of the inner-sense theory is the existence of an “introspective mechanism” which is not known to exist (Rosen 243). Thus, the claim is not economical. In addition, it seems impossible that a rational being could exist without an inner-sense as they would be “self-blind,” yet able to identify emotions in others (Rosen 244). Thus, Paul is hesitant to accept this theory.

The reason for the development of the inner-sense theory, according to Paul, is the desire to uphold [No Evidence] and [First-Personal]. Therefore, Paul then tries dropping both, resulting in a clean slate. She is left with a self-knowledge theory where we learn about ourselves in a similar way in which we learn about others but with an additional twist preventing it from completely agreeing with Ryleanism (Rosen 244).

However, before I present Paul’s theory for self-knowledge, since it rests heavily on the construct of a theory, it is important to review the argument in the origin of theories context. As a result, I will use Paul’s argument with additional premises about theories and evidence interpretation and show the result is self-referentia. Here is an outline of my argument:

P1: [Paul’s Theory-Theory] All knowledge of our own attitudes and emotions is the

product of theorizing about ourselves, usually on the basis of evidence.

P2: Theories are developed through interpretation of evidence.

P3: Our own attitudes and emotions affect our interpretation of evidence.

C1: All knowledge of our own attitudes and emotions is a product of evidence subject to

our attitudes and emotions.

Let’s consider the premises in turn. The first premise is simply a restatement of Paul’s theory-theory. Paul proposes that we know about our own attitudes and emotions by theorizing from the presence of evidence. It is clear how this would work to learn about others’ attitudes and emotions as seeing someone grasp their toe after stubbing it saying, “Ow!” would be evidence to support the theory that they are in pain. This can also be extended to oneself, but with additional sources of evidence such as “inner speech” and somatosensory cues (Rosen 245). This is an economical theory as it doesn’t depend on the existence of some unknown mechanism. Paul also finds the theory pleasing as it explains both how we can know about others’ mental state and why we can sometimes mistake our current mental state--through incorrect interpretation or collection of evidence.

The second premise highlights an important fact about theories. When an individual formulates a theory, they first collect evidence. Without any evidence to support a theory, there would be no cause for an individual to conceive of it to begin with. However, two people can look at the exact same evidence and theorize two completely different things. For example, an atheist may observe the sediment layers at the Grand Canyon and cite it as evidence for evolutionism (long-term erosion, old Earth). On the other hand, a theist may observe the same sediment layers and cite it as evidence for creationism (Great Flood, new Earth). This points to the important role of interpretation in theories.

The third premise highlights specific factors that contribute to bias when interpreting evidence. Not only do varying worldviews contribute to different interpretations, as shown in the second premise, but personal attitudes and emotions do as well. For example, suppose a wife’s husband is having an affair. Despite him accidently sending her a text meant for his mistress, she still believes he is being faithful. Her emotions and attitudes towards her husband cause her to interpret the seemingly obvious evidence in such a way that her conclusion is opposite from any impartial viewpoint. This clearly shows that attitudes and emotions are very influential on how evidence is interpreted.

When these three premises are combined, our conclusion is peculiar. All our knowledge of our attitudes and emotions is the product evidence subject to our attitudes and emotions. Paul even mentions a case example of this in passing. In a study, subjects were instructed to nod or shake their heads while listening to a speech. Those who were nodding were more likely to agree with the thesis of the speech and those shaking their head was more likely to disagree (Rosen 246). This “misleading evidence,” as Paul puts it, is more than indicative of inducing false conclusions, it also suggests that whatever we are feeling at any given moment will directly influence our knowledge of our attitudes and emotions. Thus, Paul’s theory-theory shows that our evidence for our current attitudes is circularly dependent on our current attitudes.

**Bibliography**

Rosen, Gideon, et al. *The Norton Introduction to Philosophy*. 2nd ed., W.W. Norton, 2018.