Brit. J. Phil. Sci. 40 (1989), 404–408 *Printed in Great Britain*

DISCUSSION

Epiphenomenalism and Machines: A Discussion of Van Rooijen's Critique of Popper

1 EPIPHENOMENALISM AND PHILOSOPHY

Some claims of Popper (Popper and Eccles [1977]) about the humans, the machines, interactionism and epiphenomenalism are discussed by Van Rooijen [1987]. I would like to discuss some aspects of Van Rooijen's critique. If the objective approach to science is applied to the study of behaviour, then even if epiphenomenal qualities exist, they cannot be scientifically proven. As Van Rooijen said, they belong to the subjective sphere and the subjective sphere is not observable to anybody else except to the subject to whom it belongs. So, one can subjectively observe only one's own subjective sphere. Thus, epiphenomenal qualities, as subjective qualities, are not accessible to objective scientific research and to no more than one person. Here, some classical questions can be stated. How can Van Rooijen conclude that other people also have epiphenomenal qualities only on the assumption that *he* has them and that other people are *physically* similar to him? That they are physically similar to him can be proved by scientific, objective, method because the physical structures of human beings can be studied by more than one person and can be observed by exterior senses, but epiphenomenality cannot be observed by more than one person and is not accessible to the exterior senses of others, so it is not accessible to scientific methods and cannot be scientifically proven. From that other people are physically similar to Van Rooijen, it does not follow that they are also epiphenomenally similar to him, *i.e.* that if they are physically similar to him, they are also similar in non-physical respects. Why should it be that if the objective, physical sphere of beings are similar then the subjective spheres must be also similar? Maybe some beings that are physically similar to each other have a subjective sphere and the others of the same species do not. Scientifically, this claim has the same

value as the claim that all people have epiphenomenal qualities. They are the same because neither can be scientifically proven since science cannot observe these subjective epiphenomenal qualities. Objective science can always find only physical or physiological properties which are, according to Van Rooijen, accessible to exterior senses and to observation of more than one observer. Interactionism as a theory has that advantage over epiphenomenalism, that it has the property of falsifiability. According to interactionism, mental non-physical events, if they exist, can causally influence the physical world. So, we must find cases where we have a complete physical or physiological description of events in the brain, nervous system and the behaviour triggered, but where these descriptions are not enough to explain that behaviour which is triggered. If we were not able to derive the triggered behaviour from these physical or physiological events only, then we would have to add also a mental non-physical component as a cause that completes the causal conditions which brought about the observed behaviour. However, if we can explain the events in the brain, nervous system and the behaviour with just physical and physiological explanations, then postulating an epiphenomenal subjective sphere is an unnecessary metaphysical addition.

2 MACHINES

Van Rooijen also offers a defence of the thesis that epiphenomenalism does not imply that men are machines. What in fact is a machine is unclear from Van Rooijen's article. In his own words (Van Rooijen [1987] p. 90):

> The physical similarity between me and a configuration of the matter that we call a machine is very superficial compared with all the similarities between me and other human beings (and the higher animals). Therefore there are not as many reasons to assume that machines have a psychological dimension, as there are reasons to make this assumption in relation with higher organisms. This seems to be a good reason to keep the distinction between men (and the higher animals) and machines.

But it does not seem so to everybody. I think that we can make several objections to the paragraph cited. First, what is the criterion that makes some configurations of matter a machine? We can also ask what is the criterion for differentiation and distinguishing configurations of matter which are machines from those configurations which are not? Van Roojen does not offer any criteria. Similarity between things of the same type of the same configuration of matter cannot be such a criterion. It presupposes that we already know a criterion for distinguishing configurations of the matter which are machines from those which are not, and then, for every other thing which is similar enough to the thing for which we apply the criterion, one can conclude that, because of that similarity, it is or is not a machine.

Let's assume that TV-set is a machine. But the configuration of the matter of the TV-set is so different from the configuration of the matter of the steam-engine of a steam ship, or in other words, the similarity between the steam-engine and TV-set is very small, if it exists at all. Then, according to Van Rooijen's theory, the steam-engine would not count as a machine.

I am not prepared to give here the necessary and sufficient conditions for what is a machine, but we say that some device is a machine on the notion of how it functions, how it operates, how it changes its states, how it interchanges matter and data with environment, what program it operates and so on. The configuration of the matter of which the device is composed rarely plays a role in determining whether it is a machine or not. If we have an electronic device and a mechanic device (so, it means that the configuration of the matter of which they are composed is totally dissimilar) which both show the time and do only that, shall we call them both clocks, and hence machines? Yes, we shall call them both clocks, but on the notion that they instantiate the same programs, that their operational aims are defined in the same way, that they instantiate the same functions and so on. The configuration of the matter would be totally unimportant. There is not only one configuration of matter that can be called a machine, and that everything that is dissimilar to it is not a machine. From the dissimilarity of the TV-set and the steam-engine we can conclude that, let's say, the steam-engine is not a TV-set, but not that it is not a machine. So, the steam-engine, refrigerator, TV-set, formula one V-8 engine can all be called machines despite the fact that they differ in configuration of matter very much. If what I have said is true, then why should we use the notion of configuration of matter in deciding whether or not humans are machines? We do not have much reason to do it!

In fact, it seems that Van Rooijen a priori and tacitly assumes that *he* is not a machine. But it is given without any convincing argument. Then it is extrapolated to every other human being who is similar to him in configuration of matter, and hence every human being is not a machine. This cannot be a valid argument. The answer is given before the question is put.

Van Rooijen would be in an even worse position if successful functionalistic computational programs can be given for describing and explaining human psychology and behaviour. (But what follows can raise difficulties for Popper also.) In general, functionalistic computational programs do not depend on how they will be realized. The same program can be realized by various different physical devices. This means that very different physical devices can be built for realizing the same programe, and that physical devices could have very different configurations of matter. They do not have to be similar at all. Moreover, functionalistic programs do not even depend on *physical* realizability; in fact, they do not depend on any particular realization. They can be realized by non-physical things also.

If certain functionalistic programs were to be successful in describing

human psychology and mind, it would follow that that program can be realized by various other devices including physical ones, such as computers. But we call computers machines. If the same program can be performed by computers and human beings, why shouldn't we say that human beings are machines also? It is said that programs can be realized by non-physical things also. So even if the human mind *is not* physical, it can nevertheless be a machine.

If we assume that the physical world is causally closed, then if we want the human mind not to be determined, it must be non-physical. But even if the human mind is not physical, it does not follow automatically that it is not determined. There can also be cases of determinism and determination in non-physical worlds and spheres. Why cannot possible worlds exist in which non-physical events, things and entities are also subordinated to lawful and causal relations, but, of course, not to physical lawful and causal relations? If something is not physical it does not follow immediately that in such kind of a world things and events are not determined. There exists a possibility of totally determined non-physical worlds if there exists a possibility of non-determined non-physical worlds. If the human mind is not physical, this is not enough to establish the non-determination of mind. It must be shown that the human mind is a non-physical entity which does not belong to a deterministic non-physical world, or part of the world if world is mixed, and that it is not, at least fully, determined by the physical. In fact, I introduced a possibility of mixed worlds which consists of physical *and* non-physical entities. Our world could possibly be of this kind. But we must distinguish various kinds of non-physical things and entities. If abstract entities like mathematical entities, *e.g.* numbers, have independent existence, their existence is non-physical, but it is not the same kind as the non-physical existence of human mind. In connection with causal influence, there can be various causal bounds in mixed worlds:

(a) Physical can causally influence the non-physical *but not* vice versa.
(b) Physical can causally influence the non-physical *and* vice versa.
(c) Non-physical can causally influence the physical *and not* vice versa.

We can also vary the degree of causal influence and determination. Interactionism suits well in (b) kind of worlds. But if it is true that the degree of causal determination can vary, then we can have a totally determined interactionistic world. For example, the physical event f_1 causally and inevitably leads to mental event (which is non-physical) m_1; m_1 causally leads to physical event f_2; f_2 to f_3; f_3 to f_4; f_4 to m_2; m_2 to m_3 and so on. Every event can causally follow from previous events no matter if they are physical or non-physical. (For interactionist, I think, there remains the question of what to do with conservation laws, but at least, the logical possibility of such a world exists.) If interactionist wants non-deterministic interactionism, he or she

must show that at least some mental non-physical events do not causally and inevitably follow from previous mental or physical events.

DAVOR PEĆNJAK
Moše Pijade 61
41000 Zagreb
Yugoslavia

REFERENCES

POPPER, K. R. and ECCLES, J. C. [1977]: *The Self and Its Brain*. Springer International.
ROOIJEN, J. VAN [1987]: 'Interactionism and Evolution: A Critique of Popper', *The British Journal for the Philosophy of Science*, **38**, pp. 87–92.