Kenny Jones
Tutorial Paper # 3
10/20/2014

On the Conceivability of Mechanism

Norman Malcolm in *The Conceivability of Mechanism* presents a comprehensive look at a "neurophysiological theory which is adequate to explain and predict all movements of human bodies" (127). This theory is composed of mechanisms, input-output systems that turn human beings into little more than automatons. He argues that this kind of neurophysiological system, if true, would not allow for purposive explanations of behavior, and further, that in this view of the world, it would be impossible for wants, desires, and language to exist (149). However, it seems that Malcolm may be over-exaggerating the potential effects of a scientifically proven theory of mechanism. Due to unsubstantiated claims that rely on faulty logic, specifically that purposive explanations must work in an a priori manner and cannot be contingent upon the neurophysiological state of the organism, Malcolm overreaches himself in saying that mechanism leaves no room for purposive explanations.

The crux of Malcolm's paper relies upon the notion that it is impossible for there to be a "complete neurophysiological explanation and also a complete purposive explanation of one and the same sequence of movements" (149). The basic principle of a theory of mechanism is as follows: "Whenever an organism of structure *S* is in state *q* it will emit movement *m"* (128). This is contrasted with purposive causation "Whenever an organism *O* has goal *G* and believes that behavior *B* is required to bring about *G, O* will emit *B*" (129). Malcolm states that the main difference between these two laws is that a theory of mechanism embodies contingent laws, while purposive explanations are a priori principles. To illustrate this point, Malcolm asks us to consider a situation where a man wants to retrieve his hat back from top of a roof. The man "wanted to retrieve his hat and believed this required him to climb a ladder… therefore, he

climbed a ladder" (129). He uses this example to show that under a purposive explanation of behavior, and in a situation without countervailing factors, an agent cannot want to do X, or intend to do X, without doing X. Therefore, Malcolm claims, this purposive explanation of behavior is not contingently based on anything, but is rather an a priori principle.

This distinction is crucial, as it allows Malcom to claim that purposive explanations cannot be dependent upon mechanism. He argues that a law could only be "more basic" than another law if "L_1 is dependent on L_2, but L_2 is not dependent on L_1" (131). Thus, because a priori principles are never dependent on anything, a neurophysiological explanation could not be more basic than a purposive explanation. An interesting dilemma then arises, as discussed by Malcolm on page 132, in that because purposive explanations are a priori principles, "they cannot be proved false… it is a conceptual truth" (132). However, he interprets this to mean that in light of the verification of a theory of mechanism, the "only possible outcome… would be to prove that purposive principles have no application in the world" (132).

The major conflict between these two theories lies in the fact that mechanism does not allow a place for desires and intentions. Therefore, if proved true, a theory of mechanism would be, by definition, sufficient to account for all action, and the purposive accounts of behavior would "never be true" (136). The neurophysiological explanation of behavior would "rule out" (142) the possibility that the agents intentions or wants *caused* that action. Malcolm uses this set of conclusions as a diving board into his larger fear: that the truth of mechanism would make us stop ascribing actions to people. That because a person could not "mean" (148) with intention any sound they produce, there would be no language. But before I address these problematic conclusions that Malcolm draws from the application of mechanism, I want to first address his troubling labeling of purposive explanations as both necessarily casual and as a priori principles.

In what follows I will attempt to show how a purposive explanation system can co-exist with the concept of mechanism. This will primarily be done by showing that a purposive system neither has to be directly causal or work in a priori manner in order to give a genuine explanation of a behavior – that a non-casual purposive theory that is contingent upon the neurophysiological state is not precluded by the validity of a theory of mechanism.

My counter to Malcom's argument runs parallel with the remarks made by the epiphenomenalist (134). The epiphemonenalist makes the case that "the neurophysiological condition that contingently causes the behavior … also contingently causes the intention… but that the intention stands in no causal relation to the behavior" (134). As Malcolm points out in addressing this counter argument, a purposive theory of this kind is not defined by a priori principles, but is rather contingent upon the neurophysiological state. This means, according to the definition that Malcolm gives us of a basic law (131), that purposive explanations can be dependent upon the theory of mechanism – that the neurophysiological theory can be more fundamental than a purposive explanation (131).

A purposive theory of this kind could look something like this. Whenever an organism of structure *S* is in neurophysiological state *q* it will have goal *G* and will emit movement *m*. Moreover, as a function of being in *q*, that organism will hold a belief that movement *m* is required to bring about G. Organism *O* of structure *S* was in neurophysiological state q. Therefore, *O* emitted *m*. A theory of this sort would not be a large digression from the definition of mechanism put forth by Malcolm. The main difference lies in the inherent properties of the purposive explanation. The purposive explanation in this account is not directly bringing about movement *m,* it is a by-product of *q*. However, this does not mean that this kind of purposive explanation does not give a meaningful account to why an agent acted in the way that he did.

The goals, desires and intentions of an agent and the neurophysiological factors that make up his structure can be viewed as different sides of the same coin. In one sense, the neurophysiological factors underlie and comprise the existence of these desires and intentions. However, we can also view the relationship in reverse. If we look at these beliefs as interpretable representations of what is happening at a deeper level, it is not a meaningless distinction to say that an agent made a movement because he had a desire X when his movement was necessitated completely by a mechanism, if by saying he wanted X, we mean that the mechanism both caused this desire X in him and his eventual motion. This theory contingently links the intention not only to the underlying neurophysiological state, but also to the movement necessitated by that structure.

For instance, take Malcolm's example of a man who is climbing a ladder in order to retrieve his hat from a roof. The input into this man system was neurophysiological – maybe the wind swept his hat off, the light reflected off the hat into his eye, and his mind received this input through a series of synaptic discharges. From here, this neurophysiological input will result in two observable phenomena: 1) the man will develop a want, based completely off of this neurophysiological state, to retrieve his hat, and 2) that the man will, in the absence of countervailing factors, retrieve his hat. So, it is true in this case that the theory of mechanism gives a complete explanation for the man's action. But it is also true that we can say that the purposive account gives an equally complete explanation of his action. This is because both accounts describe the same system but on different levels. The mechanistic level is concerned with the interactions of billions of synapses and neurons, the movement of particles, and of how each particle will affect the next particle. On the other hand, the purposive account is an abstracted view of this process, but importantly, it conveys the information to us in a

fundamentally understandable way. So when we say that the man desired his hat, we implicitly mean that his action was a direct result of mechanism, but we also mean that the state *q* which caused his movement *m* is directly correlated to his goal *G,* and that, as humans, we are able to understand and relate much more to goal *G* than we are to some neurophysiological state *q*. A purposive account of this nature allows us to interpret underlying neurophysiological states of other agents and ourselves in a way that we can comprehend and work with from the human perspective.

Malcolm's response to the epiphenomenalist's argument seems slightly puzzling to me. His general counter argument is that "it would be conceivable that the neurophysiological condition that always causes ladder-climbing movements should also always cause the intention to *not* climb up a ladder" (134). This idea is not only almost entirely unsubstantiated, it would also seem to go against most, if not all, of our everyday experiences. For example, when I find myself with a desire to climb a ladder, and this is what I most desire, most of the time I then climb that ladder. While not logically impossible, I do not often have urges to do certain actions and then find myself doing the exact opposite of that action for no apparent reason. But his argument does lead us into one of the problems of this kind of purposive explanation, that while holistically it may be a perfect describer and predictor of movement, agents can often misinterpret the effective desire or intention of another agent or themselves. This can be explained in terms of higher-order desires. Say for example, in a simplistic case, an agent is in a neurophysiological state in which he has two first order desires, X and Y. The agent also might have a second-order desire Z of a want to X. If X becomes the agent's effective desire, through underlying neurophysiological states that relate to Y, then the agent might be confused as to why X happened over Y and not vice versa, if he cannot also discern that he had desire Z. If *all*

desires, goals and beliefs are known, then a purposive explanation can perfectly predict behavior, however, in some and maybe all situations, it is impossible for an agent to know what all of his higher-order desires are, and thus, he might be surprised by some of his actions and choices.

Another valid point against this type of purposive explanation is that in some situations, a movement *m* will emit from state *q*, but a goal *G* will not result from state *q.* An agent can act in a way that is unintentional, i.e. reflexes like the jerking of one's knee or, in the case of the kleptomaniac, the propensity to steal. However, while there are a few cases of such sort that need to be addressed in order to complete this theory, it still stands that for most human actions a purposive explanation can co-exist with a mechanistic explanation. Indeed, I have shown how Malcolm's claim that the truth of a mechanism would necessarily take away fundamental parts of the human experience, such as language, to be false. It is not that there are no purposive explanations, rather it is that these purposive explanations are both linked to the eventual movement of the organism but also contingent upon the organism's neurophysiological state. Therefore, I have turned Malcolm's worries away from the question of whether or not these purposive explanations can co-exist with mechanism, to whether or not we can deal with this kind of system in terms of responsibility and free will. However, while this is an argument to be carried out in the compatibilist realm, I do not find it hard to imagine that a system of free will and responsibility, such as the one set up by Frankfurt, would be incompatible with the view of mechanism and the resulting purposive explanations I have tried to develop in this paper. Nevertheless, it is now clear that the mere proof of a deterministic neurophysiological theory does not necessitate our abandonment of purposive explanations or the principle and ideas that we find fundamental to the human experience.

Work Cited:

Malcolm, Norman. “The Conceivability of Mechanism”