1971

On the Conceivability of Mechanism

Martin, Michael

University of Chicago Press

http://hdl.handle.net/2144/3672
Boston University
In a recent paper Norman Malcolm has argued that mechanism is an inconceivable position.\textsuperscript{1} By this Malcolm does not mean that mechanism is logically contradictory. Rather mechanism is a position which cannot be asserted or argued for. For, according to Malcolm, asserting a position and arguing for a position involve purposeful behavior and mechanism excludes all purposeful behavior. Thus mechanism may well be true, but if it is true, then it cannot be asserted or argued for.

The bulk of Malcolm's paper is an attempt to develop the thesis that mechanism excludes all purposeful behavior. This exclusion turns on another of Malcolm's theses, namely that a neurophysiological explanation and a purposeful explanation of the same bodily movements are impossible. If this latter thesis is shown to be false, Malcolm's argument for the inconceivability of mechanism collapses. Malcolm is quite aware of this and in his final paragraph he says:

I must confess that I am not entirely convinced of the correctness of the position I have taken in respect of the crux of this paper—namely, the problem of whether it is possible for there to be both a complete neurophysiological explanation and also a complete purposeful explanation of one and the same sequence of movements. . . . Perhaps the publication of the present paper will be justified if it provides a truly convincing defense of the compatibility of the two forms of explanation ([1], p. 72).

My paper perhaps will not provide a "truly convincing defense" of the compatibility of the two forms of explanation Malcolm mentions. However, it will show that Malcolm's uneasiness about his own argument is quite justified. I will show that Malcolm has given no good reason to suppose that neurophysiological explanations and purposeful explanations are not compatible and thus that his thesis that mechanism is inconceivable is unwarranted. I will also show that in one important sense mechanistic explanations could be more basic than purposive explanations.

\textbf{1. Ambiguities in Malcolm's Definition of 'Mechanism'.} As Malcolm initially introduces the concept of mechanism it refers to "A special application of physical determinism—namely to all organisms with neurological systems, including human beings" ([1], p. 45). This version of mechanism assumes a neurophysiological theory which is adequate to explain and predict all movements of human bodies except those caused by outside factors. The explanation provided by the theory would

\textsuperscript{*} Received September, 1969.

state “sufficient conditions of movements and not merely necessary conditions” ([1], pp. 45-46). Let us call his version of mechanism $M_1$.

It is important to realize that as Malcolm characterizes $M_1$, $M_1$ provides some sufficient neurological condition as well as some necessary neurological condition for any bodily movement that is not caused by outside factors. This characterization is perfectly compatible with: (a) some necessary condition of some bodily movement (which is not caused by outside factors) being non-neurological and (b) some sufficient condition of some bodily movement (which is not caused by outside factors) being non-neurological. These non-neurological necessary or sufficient conditions could be certain psychological states, e.g. purposes or beliefs.

Later on in his paper Malcolm introduces—apparently without realizing it—a stronger notion of mechanism, one that is not entailed by $M_1$.

Let us remember that the postulated neurophysiological theory is comprehensive. . . . It is a closed system in the sense that it does not admit, as antecedent conditions, anything other than neurophysiological states or processes. . . . If the neurophysiological theory were true, then in no cases would desires, intentions, purposes be necessary conditions of any human movement ([1], p. 56).

It is clear that in this passage Malcolm is at least suggesting that neurophysiological theory would provide some sufficient condition and some necessary condition for any bodily movement (not caused by outside factors) and moreover any necessary condition for any bodily movement would be a neurophysiological one. Let us call this version of mechanism $M_2$.

Malcolm does not distinguish $M_2$ from another even stronger version of mechanism suggested in the above passage by the phrase “it does not admit as antecedent conditions anything other than neurophysiological states or processes.” On this version neurophysiological theory provides some sufficient condition and some necessary condition for any bodily movement (not caused by outside factors) and excludes any necessary condition other than a neurophysiological one for any bodily movement and excludes any sufficient condition other than a neurophysiological one for any bodily movement. Let us call this version of mechanism $M_3$.

It is important to realize that although $M_3$ entails $M_2$ and $M_2$ entails $M_1$, $M_1$ does not entail $M_2$, and $M_1$ does not entail $M_3$, and $M_2$ does not entail $M_3$. Once these three progressively stronger varieties of mechanism are clearly distinguished much of the force of Malcolm’s thesis is lost.

2. The Argument for the Inconceivability of Mechanism. A neurophysiological explanation, according to Malcolm, would have the following form:

(1) Whenever an organism $O$ of structure $S$ is in neurophysiological state $q$, $O$ will emit movement $m$, provided there are no countervailing factors.
(2) Organism $O$ of structure $S$ was in neurophysiological state $q$.
(3) There were no countervailing factors.
(4) ∴ Organism $O$ emitted movement $m$.

Premise (1) in the explanatory argument is, according to Malcolm, a contingent proposition that could be refuted by empirical evidence. (1) is to be contrasted with

---

MICHAEL MARTIN
some premises in purposive explanation. Purposive explanations have the following form:

\[(1')\] Whenever an organism has goal \(G\) and believes that movement \(m\) is required to bring about \(G\), it will emit \(m\) provided there are no countervailing factors.

\[(2')\] Organism \(O\) had \(G\) and believed \(m\) was required.

\[(3')\] There were no countervailing factors.

\[(4')\] Organism \(O\) emitted \(m\).

Now \((1')\), according to Malcolm, is not a contingent truth but an *a priori* principle; it is true because of the meaning of the terms 'goal' and 'believe.' Malcolm argues that because of the *a priori* nature of \((1')\), \((1)\) cannot be more basic than \((1')\). According to Malcolm \((1)\) would be more basic than \((1')\) only if \((1')\) was dependent on \((1)\). But since \((1')\) is *a priori* true, it cannot be dependent on any contingent statement like \((1)\).

Malcolm points out, however:

Someone might suppose that although purposive explanations cannot be dependent on nonpurposive explanations, they would be refuted by the verification of a comprehensive neurophysiological theory of behavior. I think this view is correct: but it is necessary to understand what it *cannot* mean. It cannot mean that the principles (the universal premises) of purposive explanations would be proved false. . . . Since the verification of a neurophysiological theory could never disprove any purposive principles, the only possible outcome of such verification, logically speaking, would be to prove that the purposive principles have no application to the world ([1], p. 51).

But to say that \((1')\) has no application to the world would be to say that either \((2')\) or \((3')\) is false for any organism \(O\). Malcolm rejects the suggestion that \((3')\) would be false in every case and concludes that if mechanism is true, \((2')\) is false for any organism \(O\). But to say that \((2')\) is false for any organism \(O\) is just to say that organisms do not ever have goals or beliefs about these goals. Thus mechanism excludes an organism asserting mechanism or arguing for mechanism since asserting and arguing presuppose goals and beliefs.

The crucial question is why Malcolm supposes that the confirmation of mechanism would make alleged *a priori* principles like \((1')\) not applicable to the world and thus make \((2')\) false for any organism \(O\). It would seem that the only reason for supposing this to be true is that Malcolm is talking about mechanism \(M3\). Neither mechanism \(M1\) nor mechanism \(M2\) would exclude some non-neurological sufficient condition for some bodily movement (not caused by outside factors) and thus the confirmation of \(M1\) or of \(M2\) would not entail that organisms do not have purposes and beliefs that were sufficient conditions for their behavior. In short, on \(M1\) and \(M2\), \((1')\) could apply to the world and thus \(M1\) and \(M2\) could be asserted or argued for.

The only passage I can find in which Malcolm tries to answer the sort of objection just raised is the following:

It might be thought that there could be two different systems of causal explanations of human movements, a purposive system and a neurophysiological system. The antecedent conditions in the one system would be the desires and intentions of human beings; in the
other they would be the neurophysiological states and processes of those same human beings. Each system would provide adequate causal explanations of the same movements.

Generally speaking, it is possible for there to be a plurality of simultaneous sufficient causal conditions of an event. But if we bear in mind the comprehensive aspect of the neurophysiological theory—that is, the fact that it provides sufficient causal conditions for all movements—we shall see that desires and intentions could not be causes of movements. It has often been noted that to say \( B \) causes \( C \) does not mean merely that whenever \( B \) occurs, \( C \) occurs. Causation also has subjunctive and counterfactual implications: if \( B \) were to occur, \( C \) would occur; and if \( B \) has not occurred \( C \) would not have occurred. But the neurophysiological theory would provide sufficient causal conditions for every human movement, and so there would be no cases at all in which a movement would not have occurred if the person had not this desire or intention. Since the counterfactual would be false in all cases, desires and intentions would not be causes of human movements. They would not ever be sufficient causal conditions nor would they ever be necessary causal conditions ([1], pp. 56–57).

There are a number of problems with this passage. First, although it may be true that "\( B \) causes \( C \)" cannot be analyzed in terms of "Whenever \( B \) occurs, \( C \) occurs" and must be analyzed in terms of subjunctive and counterfactual conditionals it does not follow that "\( B \) causes \( C \)" entails "If \( B \) had not occurred, \( C \) would not have occurred." Malcolm is apparently confusing here nomologically sufficient conditions and nomologically necessary conditions: "Jones' taking poison caused Jones' death" entails "If Jones would have taken poison, Jones would have died" but not "If Jones would not have taken poison, Jones would not have died" since Jones might have been shot at the same time as he was poisoned.

Secondly, even if one assumes that causality involves both nomologically necessary and nomologically sufficient conditions, Malcolm's conclusion does not follow. If \( B \) is a necessary and sufficient condition for \( C \), this does not mean that some distinct factor \( D \) could not be both necessary and sufficient for \( C \). Thus the fact that neurophysiological factors are both necessary and sufficient for behavior does not mean that purposes and beliefs could not also be necessary and sufficient for the behavior. As we have seen, this possibility is allowed for in mechanism \( M_1 \).

Moreover, as we have seen, even if purposes and beliefs for behavior are ruled out as necessary conditions (as they are in mechanism \( M_2 \)) they might still be sufficient conditions for behavior. And only mechanism \( M_3 \) seems to rule out this possibility.

We must conclude that Malcolm has not shown that on mechanisms \( M_1 \) and \( M_2 \) purposive explanations and neurophysiological explanations cannot be given of the same human behavior (where an explanation provides sufficient conditions) and thus he has not shown that mechanisms \( M_1 \) and \( M_2 \) are inconceivable.

3. The Tenability of Mechanism \( M_3 \). Thus far we have argued that at most Malcolm has shown that \( M_3 \) is inconceivable. But has he shown this? To decide the issue we must distinguish between two different versions of \( M_3 \). On the first version of \( M_3 \) all nomologically non-neurological sufficient conditions of human behavior (that is not caused by outside factors) are excluded. Nomologically sufficient conditions are specified by contingent law-like propositions. On this version of mechanism \( M_3 \) all such contingent non-neurophysiological propositions which purport to explain
human behavior would be supposed to be either false or not applicable to the world, i.e. their antecedents would be false. Let us use ‘$M_{3a}$’ to denote this version of mechanism $M_3$.

On the second version of $M_3$, all sufficient conditions (nomological and logical) are excluded. By a logically sufficient condition I mean this:

$$A \text{ is a logically sufficient condition for } B \text{ if and only if } A \text{ entails } B.$$  

According to Malcolm

(a) An organism $O$ has goal $G$ and believes that movement $m$ is required to bring about $G$ and there are no countervailing factors entails

(b) Organism $O$ emits movement $m$.

Thus the goals and certain beliefs of an organism under certain conditions provide a logically sufficient condition for the organism emitting a certain movement. Indeed, it is because of this logically sufficient condition that (1') is supposed by Malcolm to be an a priori truth.

Now on this version of $M_3$ it is supposed that any logically sufficient condition for (b) that is not in neurophysiological terms is false. Put in a different way: any a priori true hypothetical proposition (not in neurophysiological terms) in which the consequent specifies that an organism emitted movement $m$ would not apply to the world, i.e. the antecedent of the hypothetical would be false. Let us use ‘$M_{3b}$’ to denote this version of mechanism $M_3$.

Now clearly Malcolm has not shown that $M_{3a}$ is inconceivable since all contingent non-neurophysiological law-like statements that purport to explain human behavior may be false. This would be perfectly compatible with the existence of purposes and beliefs and with allegedly a priori principles like (1') applying to the world.

At best Malcolm’s argument has shown that $M_{3b}$ is inconceivable. But what a strange position this is. Who has ever held it? For to advocate $M_{3b}$ would be to advocate that any proposition (even a proposition which is completely innocuous to mechanists) that entails (b) would be false. Consider, for example:

(c) Organism $O$ emitted movement $m'$.

Now let us define movement $m'$ as a movement immediately preceding movement $m$. Then (c) entails (b) and thus (c) is a logically sufficient condition for (b). Therefore:

$$(1') \text{ Whenever an organism emits movement } m', \text{ then the organism emits movement } m.$$  

is a conceptual a priori truth analogous to (1') and by an argument analogous to Malcolm’s (1') could never apply to the world. Therefore (c) would be false. This would mean that no movement of an organism is preceded by any other movement.

It is far from obvious that any mechanist has ever embraced $M_{3b}$. But even if
some mechanists have embraced \( M_{3b} \) with its absurd implications they would not need to. Weaker and less objectionable versions are open to them, e.g. \( M_1, M_2, \) and \( M_{3a}. \)

Now it may be objected that (c), unlike (a), does not specify a cause of the movement of the organism. This suggests still another version of \( M_3. \) On this version neurophysiological theory excludes all logically sufficient conditions for human behavior which could specify causes of the behavior. Let us call this version of \( M_{3b}. \)

Consider however:

(d) Organism \( O \) receives external stimulus \( s. \)

External stimulus \( s \) can be defined as a stimulus immediately preceding organism \( O \) emitting movement \( m. \)

Now (d) entails (b) and it is not unusual to speak of a certain external stimulus that precedes behavior as a cause of the behavior. For example one might say “Jones’ walking into the room caused Smith to fidget” or “The fumes from the fire caused Jones to cough.” Now:

\( (1') \) Whenever an organism receives stimulus \( s \) then the organism emits movement \( m. \)

would be an a priori truth.

By an argument similar to Malcolm’s \( (1") \) would not apply to the world. Thus (d) would be false. This would mean that organisms never receive stimuli immediately before they emit movements. As far as I know no mechanist has ever embraced a position with such absurd implications, and as we have seen there is no need for a mechanist to do so.

Instead of embracing \( M_{3b1} \) a mechanist would surely argue that if stimulus \( s \) is the cause of behavior \( B \) a neurophysiological explanation could be given of the causal relation; he would maintain that there is a neurological connection between the stimulus \( s \) and behavior \( B \) if \( s \) is the cause of \( B. \) It should be noted that neither the fact that (d) entails (b) nor the fact that \( (1') \) is an a priori truth prevents him from doing this.

This consideration also applies to purposes and beliefs. The fact that (a) entails (b) and the fact that \( (1') \) is an a priori truth does not prevent neurophysiology from specifying a neurophysiological connection between purposes and beliefs and the behavior which results from these purposes and beliefs.

4. Mechanism and the Meaning of Terms. So far we have seen that a mechanist need not hold a position that entails that \( (1') \) does not apply to the world. Moreover, we have seen that one mechanistic position that seems to be implicit in Malcolm’s paper would have such absurd consequences that no sane mechanist would embrace it.

Let us suppose, however, that some particular version of \( M_{3b1} \) was formulated which would entail that \( (1') \) does not apply to the world but which would not have the kind of absurd implications specified above. Let us call this version of mechanism \( M_{3b1}, M^{*}_{3b1}. \) Would Malcolm be correct? Would mechanism \( M^{*}_{3b1} \) be in-
conceivable in the sense that if \( M^*_b \) were true, it could not be asserted or argued for?

Let us grant Malcolm’s thesis that \( (1') \) is true a priori. As Malcolm correctly notes the a priori nature of \( (1') \) could not be due to the form of the statement. Rather if \( (1') \) is true a priori, then it is because of the present meaning of the terms in it, in particular the meaning of terms like ‘goal’ and ‘belief.’

Now although Malcolm is correct that \( M^*_b \) cannot be asserted or argued for given the present meaning of ‘goal’ and ‘purpose,’ it does not follow that \( M^*_bl \) could not be asserted given some other meaning of ‘goal’ and ‘purpose.’ If ‘goal’ and ‘belief’ meant something different in \( (1') \) from what they mean in \( (1') \), then \( (1') \) might be found to be a false contingent statement. If it were, \( (2') \) need not be false and organisms could have goals and beliefs. Hence, \( M^*_bl \) could be asserted.

Indeed, it is not implausible to suppose that the confirmation of a comprehensive neurophysiological theory would bring about changes in the ordinary meaning of psychological terms like ‘purpose’ and ‘belief.’ These changes might make it possible to assert \( M^*_bl \) in the ordinary sense of assert, i.e. the sense of ‘assert’ that prevailed at that time in which the theory had gained great prominence.

It is also important to realize that if changes in the meaning of certain psychological terms were to occur due to the wide acceptance of \( M^*_bl \), this would not necessarily bring about any changes in the denotation or reference of these terms. Change in meaning or sense is compatible with fixed reference. Thus although ‘purpose’ and ‘belief’ might come to have different senses than they do now, they might still refer to some class of mental states, dispositions and the like. Thus one might assert \( M^*_bl \) and in this activity be doing something that has always been called ‘asserting,’ i.e. the denotation of ‘asserting’ would be the same as it is now; the only thing that would have changed would be the meaning or sense of ‘asserting.’

Thus \( M^*_bl \) is conceivable given certain changes in the meaning of our psychological language which are compatible with invariance in the referent of such language.

5. The Primacy of Neurophysiology. So far I have argued that Malcolm has not shown that several different versions of mechanism are inconceivable. Moreover, I have argued that there is implicit in Malcolm’s paper another sense of ‘mechanism’ that is inconceivable but that this sense of mechanism has such absurd consequences that no sane mechanist could embrace it. Finally, I have argued that even if a definition of mechanism could be constructed which does not have these absurd consequences this purified version would still be conceivable given changes in the meaning of some of our psychological terms that are compatible with fixed references for these terms.

I have throughout defended the compatibility of purposive explanations and neurophysiological explanations. One question that remains is in what sense a mechanist could maintain that although purposive explanations and neuro-

---

physiological explanations are compatible neurophysiological explanations are 
more basic than purposive explanations.

It is clear that a mechanist could maintain that mechanistic explanations are 
more basic if the meaning of certain psychological terms changed. If the meaning 
of ‘goal’ and ‘belief’ changed, (1’) could be a contingently true statement. The truth 
of (1’) might be dependent on some neurophysiological state. For instance an 
organism which has a goal $G$ and which has the belief that movement, $m$, is 
necessary to achieve $G$ will not emit movement $m$ unless the organism has struc-
ture $S$ and is in neurophysiological state $q$. In this sense neurophysiological explana-
tions are more basic. This is in fact the sense of ‘more basic’ specified by Malcolm.

Now Malcolm’s sense of ‘more basic’ would not apply if (1’) is an $a$ priori truth 
for if (1’) is an $a$ priori truth, the truth of (1’) could not be dependent on some neuro-
physiological state. However, perhaps some other meaning of ‘more basic’ can be 
given which would apply. I believe a different and relevant sense of ‘more basic’ can 
be constructed.

First of all it is important to note that nothing Malcolm says rules out the pos-
sibility that goals and beliefs have neurophysiological sufficient conditions. Thus a 
law of the following form might well hold:

(5) Whenever an organism with structure $S$ is in neurophysiological state $q$
the organism has goal $G$ and believes that movement $m$ is necessary for 
achieving $G$ provided there are no countervailing factors.

(5) Combined with (2) and (3) would enable us to deduce

(6) Organism $O$ has goal $G$ and believes that movement $m$ is necessary for 
achieving $G$.

Thus neurophysiology could explain all the movement which purposive explana-
tions could and more. First, neurophysiological theory could explain the causes of 
purposive behavior that are postulated in purposive explanations, i.e. purposes and 
beliefs. Purposive explanations presumably could not explain this. Secondly, 
nurophysiological theory could explain nonpurposive behavior. By hypothesis 
purposive explanations could not explain this sort of behavior.

But in one important sense a theory $T_1$ is more basic than $T_2$ if $T_1$ can explain 
everything which $T_2$ can explain and more, and $T_1$ can explain the causes postulated 
by $T_2$ in explaining what $T_2$ can explain whereas $T_2$ cannot explain these causes. In 
this sense of ‘more basic’ a mechanist might argue that neurophysiology is more 
basic than the commonsense principles or “theories” used in purposive explana-
tions. It is important to realize this sense of ‘more basic’ is compatible with the 
allegedly $a$ priori nature of the principles of purposive explanations.

Now whether neurophysiological theory is (or will become) more basic in this 
sense I do not know. But it is clear that Malcolm has said nothing to show that it 
is not more basic or will not become so.

REFERENCE

pp. 45–72.