Theology on the Web.org.uk

Making Biblical Scholarship Accessible

If you find it of help to you and would like to support the ministry of Theology on the Web, please consider using the links below:

https://www.buymeacoffee.com/theology

https://patreon.com/theologyontheweb

PayPal   https://paypal.me/robbradshaw

A table of contents for *Journal of the Transactions of the Victoria Institute* can be found here:

https://biblicalstudies.org.uk/articles_jtvi-01.php

JOURNAL OF

THE TRANSACTIONS

OF

The Victoria Institute

OR

Philosophical Society of Great Britain

VOL. LXXXV

1953

912TH ORDINARY GENERAL MEETING

HELD IN THE CAXTON HALL, WESTMINSTER, S.W. 1,

ON MONDAY, 17TH NOVEMBER, 1952.

Professor C. A. COULSON, F.R.S., in the Chair.

FROM MECHANISM TO MIND

By DONALD M. MACKAY, B.Sc., Ph.D.

SYNOPSIS

This paper discusses an aspect of the classical problem of relating mental and physical descriptions of human thought-processes which has acquired new prominence from the development of mechanisms with mind-like behaviour. A factual account of possibilities inherent in mechanisms now known leads to the conclusion (elaborated elsewhere) that any test for " mentality " in terms of the activity of an artificial organism can in principle be met.

The suggestion is not that mentality is thereby guaranteed to such organisms, but that some traditional ways of posing the problem are inadequate and based on wrong assumptions. It is suggested that these developments are in no way inimical to the Christian doctrine of Man, but rather illuminate it by suggesting a possible synthesis between complementary ways of describing his powers.

---

1. A new twist to the classical debate.
2. Towards the " vitalization " of artefacts.
3. The escape from determinacy.
4. The personality of an artefact.
5. Implications.
6. Conclusion.

1. A new twist to the classical debate

1.1. Debate as to the possibility of explaining mental phenomena on a mechanistic basis is as old as the Greeks. Between their subjectively known decisions and the appropriate bodily outcome, men observed a regular relationship of dependence. Between certain physical events in the external world, termed causes, and others, termed their effects, they also observed a relationship of dependence. What more natural than that both relationships should receive the same name of " causality? " The impact of a rolling stone on a stationary one is termed the " cause " of the movement of the other. The decision to move my finger must naturally then be termed the " cause " of the motion observed.

Physical science progressed. Physical " causes " proved to be reducible to a small and apparently exclusive number, in the sense that chains and interlocking patterns of a few causal relationships, interpreted as the manifestation of certain " forces," seemed likely to represent adequately all observed sequences of physical events. Physiological science progressed. Causal links between bodily movements and events in the nervous system were discovered in increasing numbers. Any event simple enough to be chosen for study seemed to have a causal physical antecedent.

And, of course, the question arose, where do my decisions fit into the causal chain? Is any room going to be left for the Mind as controller of these events, if the network of physical cause-and-effect should prove to be complete? Further, if I accept the undoubtable dependence of voluntary movement on my decisions (and call the dependence " causal "), what analogue of physical " force " can be postulated as the link between the two? In short, how can Mind control Matter?

1.2.   The problem, as thus formulated, was sharpened by further and complementary knowledge. It had always been known that physical violence could derange mental activity, and that the taking of drugs could distort the experience and character of the subject. Gradually, however, it became clearer that not only adventitious but fundamental features of personality and mental life were linked with biochemical, electrical and other features of bodily structure and activity. Here was evidence of a significant dependence in the reverse direction. Not only was there a problem of accounting for the action of autonomous Mind on servile Matter, but also one of explaining an apparently comparable action of material agencies on the very springs of mental activity. It began to be whispered, indeed, that Mind might be after all a " mere epiphenomenon " of the motions of Matter:—that Man might be but a " mere automaton," driven by " blind forces "—and so, of course, in no way responsible for his actions. But to the logic of this conclusion we return.

1.3.   Naturally concomitant with these developments were speculations on converse lines. If the human brain and nervous system were in some sense a physical mechanism—or even if it were not—might it not be possible in principle to construct an artificial mechanism or " artefact " which should behave as if it had a mind? For a long time the question had scarcely an academic interest, for technology could hardly point the way to equip an " artificial man " with human powers of locomotion and action, let alone of thought and dialogue. Even when the age of the machine came to render trivial the problems of motor activity, it was easy to ridicule the mental limitations of any foreseeable artefact—chiefly in respect of its inability to modify its responses or carry out any trains of reasoning comparable with those of human minds. " Machine " indeed came to be synonymous for a servile mechanism, capable perhaps of executing more quickly or more powerfully the purposes of its designer, but (more or less by definition) without any power of forming or adopting purposes of its own.

It is to avoid begging the question in this way that I have introduced the neutral term " artefact " (in the sense of artificial construct) for the class of mechanism that we shall here consider.

1.4. The nineteenth century saw the growth of two independent developments that eventually revolutionized the prospects of synthesizing mind-like behaviour in an artefact. Both had their seeds in earlier work. The first was the development of self-adjusting control systems, typified by James Watt's famous steam-governor. The second was the development of symbolic logic, in which George Boole played a classical part, making it possible for arithmetical calculating machines symbolically to carry out trains of logical reasoning. The advent of electronics multiplied the complexity and speed of devices embodying these developments, without introducing any essentially new principle. Indeed in the 1830's Charles Babbage designed an " analytical engine " which in principle had all the powers of modern electronic computors, and brought upon him a spirited debate with those who saw in its imitation of human faculties a threat to the dignity of man.

But it was the advent of high-speed computors, using thousands of electronic valves, and capable of solving in seconds problems on which men spend months, that in the last two decades brought sudden popular attention to our question. Regrettably dubbed " electronic brains," these devices acquired a reputation for mental power that seemed to put the human brain itself in the shade. The inevitable reaction has followed. It is already no longer fashionable to suggest that such computors provide a good model of the brain, nor to take seriously the analogies between their disorders and mental disease. But the question has at last arisen in realistic terms: how far could we go if we *wanted* to make, not a computor, but an artefact with characteristics that in a human being we should regard as evidence of mentality? What are the differences between present-day computors and human brains, and could they be eliminated— in principle—if we wished to do so? The answer is largely a matter of fact, and it is chiefly towards clarifying some of the facts that this paper is directed. For good or ill, the classical debate has taken a new twist. Factual developments make it no longer derisory to ask: could an artificial mechanism be said to have a mind?

2. Towards the " vitalization " of artefacts

The author has elsewhere (*Brit. J. for Phil. of Sci.* **2**, 105 (1951); *Proc. Arist. Soc. Suppt.* 1952, pp. 61–86 and references therein) discussed the technicalities of securing mind-like behaviour in artefacts, but a brief explanation of some of the principles on which present possibilities rest may help to place these in perspective.

2.1. What is perhaps the basic principle is illustrated by such familiar devices as the thermostat. An electric heater warms a room. When the thermometer rises to some preset level, the mercury pushes open a switch that cuts off the heat. When the room cools a little below that temperature, the mercury falls and closes the switch—and so on. The system behaves as if it were *trying to resist* changes in temperature. If the preset level is raised, the heater at once comes on until the room settles down at the new temperature. The system's basic " goal " is the matching of the level of the mercury to the preset level, wherever that may be. Any discrepancy between the two levels occasions activity (heating or cooling) calculated to *reduce* the discrepancy.

The activity of the heater is controlled by signals " fed back," as we say, from its field of action.   Such a system is called a " goal-seeking " system because these so-called " feedback " signals drive it to minimize the interval between its present state and the preset state or " goal." against any opposing influences (within limits).

The " feedback principle " so illustrated can be applied in any situation in which a mechanism is required to act as if it had a purpose.   It need only be provided with appropriate receptors of the necessary information as to its separation from its goal—i.e. as to the success of its activity— and means of calculating from this information the next step to try in order to reduce that separation.   If the output of the calculator is used to steer the mechanism, it will then automatically pursue its goal to the limit of its powers.   Examples now realized are the various self-guided missiles that can detect and pursue targets in spite of all evasive action.

2.2.   In mechanisms such as these the various distances and speeds entering into the calculations are represented in the calculating device by electrical or other physical quantities.   The representation of features of the field of action by internal configurations of the mechanism in this way is a very general principle, which can readily be extended to the field of abstract ideas.

In one possible method every fact to be represented is given a code-number, such that each digit in the number is either 1 or 0, representing the answer (yes or no) to one of a set of standard identifying questions— as in the popular game of " twenty questions."   Making deductions from facts coded in this way then amounts to doing arithmetic (in the scale of 2) with the numbers representing them, and standard calculating-machine technique can be used to mechanize processes of reasoning in principle as complex as desired.

A code-system of this kind is ideal for handling exact information of limited variety.   It can enable an artefact in principle to engage in active, responsive and apparently purposive interaction with any field of activity capable of representation in such a code, including dialogue with a human interlocutor on suitable subjects (such as chess, for example: see Shannon. *Phil. Mag.* **41**, 256 (1950)).   But the artefact, despite the flexibility of its responses, is still deterministic in its function.   It may be judged to be so by a simple test: two such identical artefacts supplied with identical information would at all times be found acting in exactly the same way.

3.   The escape from determinacy

3.1.   The reader may have his own views as to the extent to which the above statement would also be true of human beings, but it is at least commonly supposed to be false, and it is certainly not necessarily true of all conceivable artefacts.   There are many ways in which a limited amount of indeterminacy could be introduced into the functioning of even such an artefact as we have discussed, so as to enhance its resemblance to a normal imaginative human being (MacKay, D. M., *The Christian Graduate*, September, 1949; Turing, A. M., *Mind*, **59**, 433 (1950)).

3.2.   There is, however, an opposite approach.   Instead of introducing indeterminacy into the functions of a deterministic artefact, we might

begin at the other end, as it were, and consider the possibility of introducing a measure of co-ordination and purpose into the activity of an artefact initially designed to function more or less randomly (MacKay, *Proc. Arist. Soc. Suppt.* 1952, *loc. cit.*).

As a brief illustration of this new principle, let us imagine a printing machine designed rather like a large typewriter, but printing complete English words instead of letters. For simplicity, let us assume that it has a " vocabulary " of 5,000 words, controlled by 5,000 keys. It is easy to devise some mechanism which would normally punch keys at random, producing a meaningless jumble of words. Suppose, however, that we could *control the probabilities* of its punching different keys, rather in the way that a loaded die or roulette wheel controls the probabilities of different numbers. It would then be possible to increase the frequency with which meaningful sequences of words were produced, by " weighting " the chances of each word according to the words preceding it, so that a word that made sense was more likely to follow than a word that did not.

For example, if the words " eaten my " happened to occur, the probability that some food-word should follow is much higher than the probability of one describing something non-edible. We should therefore arrange that when the mechanism has produced the words " eaten my," it automatically (by reference to stored information) " weights " its vocabulary so as to favour all such following words that could make sense, in proportion to their likelihood of doing so.

3.3. Now the labour of supplying the necessary information in this form to our artefact would be prodigious. The interesting possibility exists, however, of making such a mechanism *acquire the information for itself*. The key principle is in effect one of " natural selection." It will be remembered that by receiving information as to the success of its activity, our earlier artefact was enabled to pursue any preset goal. The " feedback signal," after some automatic calculations, was used to steer the mechanism.

In analogous fashion we could provide our present mechanism with signals indicating the meaningfulness or otherwise of its current output-sequences, and make these signals *control the probabilities of sequences according to their success or failure*. The simplest way to do so would be to sit by it in much the same way as one would with a child for the same purpose; but in principle the mechanism could be designed to extract its own corrective signals if given access to a large enough supply of standard English text suitably coded for its use. The effect would be that meaningful sequences should steadily become more frequent, and meaningless sequences steadily be eliminated. Finally, if the mechanism were designed also to receive and react to information from an external field of activity as in our previous examples, it may be seen—or at least perhaps accepted—that the incoming information could be used to secure for the sequences of the artefact's activity not only meaningfulness but relevance.

3.4. Such suggestions may sound fantastic, and it is necessary to remind ourselves on the one hand of the quite impracticable complexity of any such artefact to be comparable with the human brain (which has some 10^{10} elements), and yet on the other of the simple factual basis on which the suggestions rest. We have considered word-sequences only by way of

example. It would be a much simpler matter to organize the artefact's activity at a sub-verbal level, making its verbal output the consequence of selections among a much smaller number of basic symbols. But the technicalities are not our chief concern.

The point that I believe to be established, though only scantily illustrated here, is that known physical mechanisms can suffice, in a suitably-designed artefact, to enable it to meet in principle any test for mental attributes that we can specify in terms of its internal or external activity.

As I have shown in the papers cited, such an artefact could pursue an active, autonomous, logically disciplined yet imaginative course, exhibiting any features of human personality that we are clever enough to know how to specify. It could make hypotheses, could form and change its own purposes either spontaneously or according to its experience, and could do all these things in responsive intercourse with human beings on human topics.

No barrier of principle—and it is only questions of principle that concern us from the philosophical standpoint—would seem to prevent an artefact from meeting any test of the kind usually suggested to " justify the inference to other minds."

4. The personality of an artefact

4.1. Does such an artefact then have a mind? Is it conscious of what it is doing? Does it feel and not merely simulate emotion? Such are currently popular questions. One might join in ridiculing the suggestion " that a mass of wireless valves could ever fall in love."

But to consider the suggestion in this form is, of course, to commit a vulgar error. In the analogous case of a human being, it is not the mass of nerve-cells inside the skull that has fallen in love. To say so would be a misuse of language. It is the *person* who has fallen in love; and to assert— or even to deny—that the nerve-cells of his brain are in love would be to show ignorance of the proper uses of the terms.

It would therefore be but a perverse distortion of the issue to ask whether an artefact could be angry or affectionate, if by " artefact " we meant " some box of wires and valves." If we are interested in evaluating the true parallel, we must compare the box of wires and valves with the sight that a surgeon sees on an operating table; we can compare only the *personality that it mediates* with the human personality.

We are accustomed to the unconscious abstractive process that can hear a declaration of love in the noisy wobbling of the red-and-pink protoplasm we call a face. We choose to use *personal* language in describing such an encounter, because it makes more sense to do so. It may require much mental discipline to bring ourselves to the corresponding abstractive effort with our artefact. One could perhaps be helped by imagining the artefact as a correspondent, or as decently clothed in some fashion! But it is only when this effort has successfully been made that we are in a position to face the philosophical question. In its original form the problem is quite overlaid by what amounts to the humour of buffoonery.

4.2. Our first question is therefore: could personal language consistently be used to describe encounter with such an artificial personality? I believe

that it could, for the simple reason that any deficiencies in the " personal " features of its activity can be remedied as soon as they can be specified (MacKay, *Proc. Arist. Soc. Suppt., loc. cit.*).

But in the last phrase there is a rub. It is easy enough to specify enough characteristics to make the artificial personality a tolerably intelligent and interesting and even emotionally-motivated interlocutor. To that extent personal language would indeed seem to be not only justified, but the only sensible language to use, just as in the case of a human being. But it is by no means obvious whether now or at any time or even in principle we can understand enough of the depths of human personality to be able to specify adequately *all* the deficiencies remaining unremedied.

4.3.   To the second question therefore as to whether such a personality could ever be fully human, we must return the old Scots answer of " Not proven." The one thing that seems safe to assert is that the barrier, if barrier in principle there be, rests on limitations to our psychological rather than our mechanical knowledge.

In short it is worse than folly to consume energy in searching for " something you'll never be able to reproduce in a machine." To do so is to accept a misconstruction of the real issue, which concerns the extent to which man can understand his own nature well enough to specify the requirements for an artificial human personality.

4.4.   What then of consciousness and mind? I should be prepared to defend the thesis that *as far as we can find words for tests* for these attributes, it is possible for an artefact to meet those tests. But if we were to leap the ditch that is deductively unbridgeable, and say that an artefact that behaves in every way as if it were conscious *is* conscious—what then? Or to put it conversely, what do we think we are denying if we say it is not? We are surely facing a problem quite similar to the classical one of deciding whether any one is conscious but ourselves; the reader who would venture to frame a deductive test for the artefact had better walk warily, lest he deprive himself also of consciousness (in the eyes of all others) by the same stroke.

4.5.   But more seriously—and particularly to our present purpose—it may be asked whether there are any grounds in Christian revelation for pontification here where deduction fails. Bluntly, one might ask whether God's licence to men to grow new personalities places any restriction in principle on the manner in which their necessary bodies are made, or on the material—whether protoplasm or copper or anything else—from which these are constructed.

We have already seen how unlikely this problem is to arise as a practical issue. But it is difficult to see any specifically Christian objection to the possibility. Our suggestion would be that in the face of our patent ignorance, and even doubt as to the meaning of the question, the Christian attitude should be one of " reverent agnosticism "—reverent because personality, even an imitation of our own, is a great mystery; and agnostic because plain honesty thus best describes our position. What would seem a real disservice to the Faith would be to presume to foreclose a possibility that God appears to have left open, and so, as sometimes before, to distract men's minds from the real content of Christian belief.

5. Implications

5.1.   We began by considering without comment the view that the relationship between my decisions and my bodily movements was one of causality.   We saw that this view implied, but did not suggest the nature of, some mechanism of interaction between an entity termed my mind and my body.   We saw how physical causation has gradually spread through the picture, steadily diminishing the area on which " mind " might be said to lay causal hands.   We saw that, in this language, mind itself seemed subject to the action of physical causes.

And then from the opposite direction we have followed a new twist in the story.   It has appeared that those features of behaviour which we most commonly attribute to the " causal action of Mind " can be quite well reproduced by a mechanism functioning throughout according to ordinary physical principles.

5.2.   Squeezed out in one direction, never admitted in the other, it seems as if Mind might soon find no place in our view of Man.   But of course it is not so.   What we are being forced to realize is, I suggest, rather that " Mind " is a word which belongs to a different logical vocabulary altogether from words describing physical causes, in the sense that words of an algebra problem belong to a different logical vocabulary from words describing the ink that delineates it.   " What is there " can be described completely in terms of algebra or in terms of ink, but the two descriptions do not mix.   In the same way our suggestion is that the " mentalist " or personal description of a human activity does not rival but is complementary to a description in physical terms.   It is not the descriptions which are exclusive, but the logical backgrounds in which the respective terms are defined.

5.3.   What then is our alternative to the classical account?   Between my decision and my *responsibility* it would seem proper to posit a causal relationship.   Between the physical events in my brain concomitant with my decision, and the appropriate bodily motions, it may be proper to posit a causal relationship.   But to attempt to use an identical relationship of causality as a link between my decisions and their physical expression appears to be an error.   If we must call the link " causal," we should logically use some distinguishing adjective to prevent our habit from leading to nonsense.

For what we call our " decision " may from the physical observer's standpoint be an abstraction from a whole sequence or pattern of events whose causal linkages, even if not complete and unbroken, may extend backwards and forwards considerably in the time of the observer.   Not even temporal priority could therefore be guaranteed to what we wish to term the mental " cause " of our action, and it seems not unlikely that in the physical picture the room available for a causal antecedent would often be almost completely occupied by well-knit physical events.

5.4.   But why should we wish at all to use this language of pseudo-physical causality?   Perhaps the commonest reason is the belief that unless I can call my decisions the (pseudo-physical) causes of my actions, I am not responsible for those actions.   We cannot here discuss this view adequately;   but I believe that it is fallacious.   If I find my body jerking

in activity *against* my will, then I may fairly disclaim responsibility. But the reason is not that there was a physical cause of my action and therefore no mental cause, but that if I am asked " Was this of your will? " I know directly what is meant and can answer " no." If I choose deliberately to take some action, my answer to the same question is " yes " and I cannot evade responsibility, for the physical description of what went on in my brain, however causal, is but an account in a complementary language of the very process of deliberate choice that is apparently in question. In short, our suggestion is that responsibility is to be judged not by the question: " had the act a physical cause? " but rather: " was the act the outcome of a decision? " The language of the actor, rather than the complementary language of the observer, is the group in terms of which the calculus of responsibility is framed. And in the last analysis it is neither acts nor consequences that Christianity declares to be the first objects of moral appraisal, but attitudes, in the most fundamental sense of the term.

5.5. At the same time we may note that current physiology in any case gives little encouragement to the view that the physical course of a human brain should be predictable—even in principle—over any appreciable length of time. And we have seen that an artefact could show an enhanced resemblance to a human being in the domain of originality and choice if it incorporated a measure of indeterminacy in its mode of operation. The significance of this indeterminacy is yet another of the problems to which these developments direct attention, but which we cannot now discuss.

5.6. It may seem shocking to some to be invited to modify a thought-model so traditionally wedded to Christian apologetic. We have perhaps been accustomed to think of Mind as a kind of " stuff " inhabiting the body and exerting occult forces on its movements; and to suggest that an artificial organism could show the behaviour we have always interpreted as evidence of these forces may seem heretical.

But is that currently " traditional " view—or habit of speech—in fact Biblical? It would seem that for the Hebrews at least a debate in these terms could scarcely have been formulated, for their view of Nature entertained no such concept as " mere matter obeying mechanical laws." The main Biblical distinction would seem to be between " Spirit " on the one hand, and " mind-body " or " organism " on the other. *Spiritual* life is declared to be something not automatically present in a human being, but having to be received in repentance as the gift of God; it is eternal, and not limited to the spatio-temporal confines of the human organism.

The concepts of *mental* life on the other hand find no mention apart from a body of some sort. The doctrine of the resurrection of the body indeed lends weight to our suggestion that Biblically mind and body constitute two aspects of a concrete unity. This is not to say that the perishing of the body is the end of the personality it mediated; it need not be even an interruption. Even in the case of an artefact a complete knowledge of its momentary state before destruction could enable its personality to be reproduced in a new mechanism, not necessarily built of the same materials.

Nor do we imply that " spiritual life " and " mental life " are two varieties of the same thing. But here the water is deep, and speculation

finds few landmarks in revelation.   It is evident that no linguistic distinc-
tion that we might wish to draw has any parallel in common usage, even in
translations of the Bible, where " spirit," " soul " and " mind " are often
interchanged.   But conceptually the distinction seems clear and necessary,
and might perhaps be followed up with profit by those more competent to
do so.

5.7.   Underlying our whole approach has been the conviction that
either to assert or to deny that mind is " nothing but " a by-product of
mechanism, is to lend countenance to a false formulation of the problem.
The phrase " nothing but " begs the question here as in other debates,
and typifies what one might call " reductionist " thinking.

Reductionism is properly attacked not by disputing the *exhaustiveness*
of a given reduction—say to mechanical terms—but by challenging the
implicit and undefended assumption of *exclusiveness*.   The real question
is not whether mind is an abstraction from the workings of a mechanism,
but whether that fact if true affects the responsibilities of the personality
so mediated.

6.   Conclusion

The foregoing inadequate discussion has had one limited objective.   It
is not contended that artefacts constructed along these lines *must* in
principle be admitted to have " mentality."

Our suggestion is merely that the contrary is not proven, and that any
attempt to " maintain the dignity of man " by searching for limits to the
powers of artefacts is misguided and foredoomed.   This is no prophecy,
but a deduction from the demonstrable fact that to specify exactly a
behavioural test amounts in principle to specifying a mechanism that can
meet it.

We have left open the question whether we could ever enunciate an
adequate test for mentality in the full human sense.   Indeed our plea
would be for more open-mindedness in facing an issue on which it is
difficult to conceive of the kind of evidence that would be adequate.   The
view here offered is that these developments only illuminate and in no way
challenge the Christian doctrine of Man.

<div align="center">DISCUSSION</div>

The CHAIRMAN (Professor C. A. Coulson, F.R.S.) said: I believe that there is a
profound change taking place in the whole field of Christian apologetics. To some
extent this is being forced upon us by the astonishing discoveries of modern science,
and by its evident power of building up a coherent picture of the universe in which
we live.   But, whether that be its origin or not, it seems to me almost wholly good.
This change is well illustrated in the approach which we now make to the central
problem of the nature of mind, so ably discussed in the paper by Dr. MacKay.   There
was a time once when Leibniz could say that everything that went on in the mind
of a man was as mechanical as what went on inside a watch.   And such a view,
magnificently supported by the physiological researches of Sherrington and Adrian,
and no less by the corresponding advances in pure physics, biophysics and bio-
chemistry, seemed likely to make God into a hypothesis for which there was no real
use.   So long as Christians were willing to accept a " region of science " and a " region
of religion ", parcelling out the country of the mind into departments under indepen-
dent authority, there was no hope for religion.   Every new scientific discovery en-
larges one department at the expense of the other, until religion, deprived of any

solid basis, becomes a sentimental nostalgia. Descartes had to locate the soul in the pineal gland since there was nowhere else for it. Even that habitat could only be granted on sufferance, until such time as the anatomist had found its " real " function.

There is—and never was—any hope that way. Advance, by which I mean the recognition of the right relationship between science and religion, could only come by an enlargement of our concepts. Such an enlargement would show that what appeared not to fit was in fact part of a bigger pattern. It has always been like this in science. In arithmetic, for example, immense difficulties appeared inescapable until the realm of positive integers $(0, 1, 2, \ldots)$, was supplemented first by rational numbers $(\frac{3}{4}, \frac{7}{5}$, etc.), then by irrational numbers $(\sqrt{2}, \pi \ldots)$ and finally by complex numbers $(x+iy$, where $i = \sqrt{-1})$. Each successive enlargement of the concept of number has been like a release, leading to a deeper understanding of the real meaning of a number than would have been possible before. Not infrequently the enlargement leads to a reconciliation of apparent opposites, as in the now famous situation in which an electron has to be thought of either as a particle or a set of waves according to circumstances.

I believe that this is the situation which we are now reaching in our thinking about the nature and science of mind. I am reminded (if I may use an illustration from mountaineering) of the different descriptions that a climber might give of any selected mountain. If this mountain were Ben Nevis, and the climber was standing on the North looking at the Ben, he would report that it was a rugged mountain with rock buttresses that required some skill to surmount. If he stood looking from the South, he would report that it was a gently rolling smooth surface, with grass almost to the top. Other points of view, such as from the loch side near Fort William, would yield yet other descriptions. But no one would say that the divergence of description mattered; all were partial views, and all cohered in the single concept of the mountain.

This is only an analogy, but it should remind us that if we ask questions about the nature of mind, framed in biophysical terms, we shall be bound to expect answers dealing with the almost innumerable collection of nerves and nerve endings that comprise the brain: if we ask questions in biochemical terms the answers are bound to be in terms of phosphorus uptake: questions in terms of the idea of beauty, or poetry, will necessarily receive answers within that same context (otherwise they are not answers). Questions in the language of a " spectator " of the mind may receive entirely different answers from apparently related questions in the language of an " actor " description. There is no conflict, nor can there be. Each view of the mountain, each description of the mind, is coherent and consistent in itself. It is exclusive—for in this matter it is dangerous to mix our drinks, and a superposition of two photographs of Ben Nevis from the North and from the South, would only lead to confusion and muddle—but it is not exhaustive, as MacKay so properly points out. The only really dangerous people on the mountain are those who, having familiarity with one way of ascending to the summit, claim to know all about the mountain.

This is familiar enough—now—in physics, where we call it the Principle of Complementarity. If we are prepared to carry it into our discussion about mind and matter, many of our difficulties melt away, and we are ready for an intellectual awakening. One illustration will suffice to show what I mean. It is taken from the scientific autobiography of Max Planck, one of the pioneers of twentieth-century physics. When Caesar crossed the Rubicon and burnt his boats, he felt himself free to cross or not to cross: and indeed he must have wrestled mentally some time before he felt he could make up his mind about this momentous decision. For him, as actor in the play, his will was free. But for the historian, writing his account of Caesar's military life and triumphs, just the opposite must be true. We count him a great historian just insofar as he is able to show us how inevitable the decision to cross the river had to be. For the historian, and for us as spectators of the event, it was inevitable: Caesar's will was not free.

Now both of these views are correct. The point is that we must not mix the actor and spectator descriptions. We must choose our language according to our intended mode of discourse, or type of discussion. Shakespeare's poetic outburst, " What a piece of work is man! . . ." is as much out of place in a treatise of anatomy as an enumeration of the function of each of our many bones would be in an exegesis of the words of Genesis which describe the way in which Eve was manufactured out of one of

Adam's ribs. We are on a new plane of thought now, and, perhaps for the first time for hundreds of years, we can begin to see the manner in which, behind all our varied descriptions of the nature of the brain as mechanism, as biochemical reaction, as seat of nervous control of the body, or as one of the manifestations of mind, where concepts such as truth, beauty and goodness begin to take on meaning, there is one reality. Just as the separate pictures of the mountain all cohered in the idea of Ben Nevis, so our separate pictures of man and his universe cohere in the concept of God. The truth has made us free—gloriously free, within a wider context and pattern than ever we could have dreamed.

I believe that Dr. MacKay's paper is most valuable, because it tells us of this new situation, and sets us along lines of thinking and conversation which must inevitably turn out to be utterly rewarding. A Christianity, free from the wasteful necessity to defend its little strip of the mind's territory, can appear more brave, more convincing, more fulfilling than it ever could have been for earlier generations.

Dr. O. R. BARCLAY said: It seems to me that the artefacts which Dr. MacKay describes are really logical machines, i.e. they are theoretically capable of carrying out any logical process. When they are described as showing non-logical properties such as free will, these properties have to be defined negatively for the artefacts, e.g. as degrees of non-logical behaviour. What is shown is not free will but freedom from logical constraint, which is one negative aspect of free will.

This has two consequences. On the one hand it defines the capabilities and limitations of the artefact. On the other hand, because it is possible to say something (albeit only negative) about all the functions of human personality in terms of this artefact, there is a misleading impression that all the functions of human personality can be described adequately in such terms. There is no activity of the mind of which this artefact is completely incapable, because every mental function has a logical (or non-logical) content. What the artefact can do, however, is so small a fraction of many mental processes as to be unrecognizable as the same activity. Looked at from the point of view of logic the description may be complete, but a description of some human activities in terms of logic only is so incomplete, and often so largely negative, as to be actually misleading. The impression that these artefacts can show homologues (not merely logical analogues) of free will, etc., is therefore seriously misleading. Yet it is given plausibility by the fact that they can show true equivalents of one aspect of every mental activity.

The Rev. U. E. SIMON said: Dr. MacKay's empirical approach is not out of harmony with the epistemology of Leibniz who, in his monads, makes provision for all levels of interpenetration and response to stimulation. Similarly I feel the subject-object relationship hinted at in the paper would be clarified if it adopted the Kantian *Critique of Pure Reason* or indicated its point of departure from the position there given. In other words, I am concerned to show that a line of continuity with historical philosophy exists, and might be exploited with beneficial results.

Mr. GORDON E. BARNES said: It is often assumed that the basis of the psychological study of personality is an analogical argument—we find in our own personality that certain behaviour is a manifestation of certain inward experiences, and by analogy we conclude that the same behaviour in others is an accompaniment of the same inward experiences.

But I suggest that this view is a rationalization of something which we apprehend on other, alogical, grounds. If analogy were the sole basis of psychology, there would be no grounds for a psychological investigation of the insane, since by regarding a person as insane we are implying, amongst other things, that he has an abnormal relation between behaviour and subjective experience. Furthermore, a child seems to know whether its parent is angry, pleased, or fearful, etc., long before it appreciates argument by analogy. The basis of the " I-Thou " relationship seems to be, not a logical one, but an intuitive one.

Now I quite agree with Dr. MacKay that it is probably impossible to frame a " deductive test " that would distinguish between his hypothetical artefact and a human being, but I wonder whether his artefact would pass the more fundamental test of whether we, knowing it to be constructed of valves and wires, would believe intuitively that it mediates a personality.

On a logical basis, analogy would lead us to impute both mental and material aspects to the artefact as to other people. Now it would be theoretically possible to construct a whole series of machines with every grade of behaviour complexity from Dr. MacKay's hypothetical artefact down to a simple thermostat or governor, or even a cork floating on water and oscillating about a mean level. To be consistent then it seems that one would either have to adopt a panpsychism or else be prepared to say at what point in the series we should find a qualitative difference in behaviour that would justify the use of mental concepts in describing the more complex machines.

I think the same sort of argument holds if we start not with a complex artefact but with a complex organism. We are faced with the alternative of adopting a panpsychism or of explaining the qualitative difference between a living organism such as man and an inanimate object credited with no mental attributes. It was the latter alternative which led ultimately to the Cartesian dualism.

I should like to ask Dr. MacKay what his views are on this problem.

Mr. C. D. CURLING said: Dr. MacKay's paper is to be welcomed if only because it may help us to see more easily that the aim of philosophy is indeed just that enlargement of thought to which Professor Coulson referred. This was always a belief of A. N. Whitehead and his work on the theory of the abstractive process may soon receive more attention in the light of the developments reviewed in this paper.

Any attack on reductionism is in accord with this belief, but I am not clear that we know enough of the properties of exclusive logical backgrounds to do more than point out the paradoxes into which the reductionist is driven. What kind of theory of truth do we need that will admit of several descriptions with different logical backgrounds? Are these to be held equally true?

A scientist knows well enough when to speak of particles and when of waves. Is this possible elsewhere? Is the criterion " it makes more sense to do so " sufficient?

I accept Dr. MacKay's conclusions, but the grounds for decision between complementary descriptions do seem to need further exploration; if we knew more we might see a little better the kind of background which leads a person to take up a fundamental attitude which orders his thinking about all descriptions of, for example, mind and matter.

Dr. J. T. AITKEN wrote: In theory, a machine can be constructed to do anything, but the project usually fails on such practical details as accommodation and power. The value of " brain-like " machines lies in their ability to foster testable hypotheses. I agree with Dr. MacKay that much trouble has resulted from the false comparison of artefact and " mind " instead of " some of the activities of the brain ". If mind is defined in terms of brain activity *only*, then I am happier.

When a moral choice has to be made, then Christians would be expected to react differently from Pagans because Christians *have been given* a bias which is not naturally present. Similarly even Pagans would react differently from animals. A machine which " learns from experience " and has the capacity to store suitable information is not likely to " commit suicide " and will thus choose the less lethal of the alternatives or the most advantageous and pleasant to itself, so to speak.

I am not quite clear about paragraph 5.6 (2) and would suggest the following hypothesis.

All animal life, including man, has a body with or without a co-ordinating nervous system. In the more complicated and specialized forms (I resist the temptation to say " higher animals ") there is found a nervous system which not only reacts to the incoming stimuli but is capable of initiating action and creating sensory pictures. Now man differs from other forms of animal life, I *believe* (because of the revelation in Genesis and other parts of Scripture), in that man possesses also spirit (something with at least potentialities for eternal existence). After the Fall, man has lived off-side, and biassed by sin. Regenerate man is body *plus* spirit *plus* Holy Spirit. The presence of the Holy Spirit makes the regenerate man on-side and corrects the bias. (The latter is not completely corrected in this life because of the effects of sin on the body—the " pollution of sin " of the Reformers.)

Machines (artefacts) can be conceived of which will perform without the bias of sin all the actions of man's body and therefore perform them better. The artefact may even be trained to make moral judgements of a sort. But unless *God* in His wisdom *gives* that artefact spirit, it can never compete with man in " glorifying God and enjoying Him for ever."

Dr. R. J. C. HARRIS wrote: I have read Donald MacKay's paper " From Mechanism to Mind " with considerable profit, but there is one question which I would like him to answer.

He suggests (4.1) that the brain " mediates the personality " and further states (4.2) that " it is by no means obvious whether now or at any time or even in principle we can understand enough of the depths of human personality to be able to specify adequately *all* the deficiencies remaining unremedied ". What does Dr. MacKay intend by " *in principle* "? Does he, for example, mean that the personality is only *partially* " mediated by the brain "? What, in fact, does he mean by personality?

We would all agree (5.6) that in Scripture the perishing of the body (physical organism) is not held to be the perishing of the personality, and that we shall, e.g. recognize each other on the other side of the grave, but surely it is an extension of an analogy (an over-extension) to suggest that the " personality " of an artefact could similarly be reproduced in *different materials*. For does Dr. MacKay mean " non-material " material? I suggest that by " reproduced in a new mechanism " he really means constructed in the *same materials* to a different pattern—i.e. the same valves, relays, etc., arranged in a different way.

Dr. MacKay suggests that by mental discipline we can bring ourselves to use " personal language " in describing our encounter with an artefact where this is appropriate. Perhaps he would like to go on and suggest what " language " one artefact could use to describe " its " encounter with another. What would constitute " individuality " in an artefact?

I feel, in view of the strictures of 4.1, that I ought to apologize in advance for what may appear to be the commission of not one but several " vulgar errors ".

Dr. H. MARTYN CUNDY wrote: I have read this paper with considerable interest, but I feel rather inadequately equipped to comment on it, since I am no philosopher. My immediate reaction to all philosophical language is to translate it into simple words! It seems to me that what Dr. MacKay has shown is that if we decide in advance what we want a machine to do to resemble a human personality, then we can imagine a machine which will do it. (Is there any difference between " construct in principle " and " imagine "?) But this is surely self-evident. We need not do anything very elaborate. We could simply record all the sense-impressions received during the life of an actual person, and reproduce them. All these supposed " goal-seeking " machines can only seek a pre-chosen goal, or possibly, if they incorporate a random element, one of a number of possible pre-chosen goals. It would be simpler, and no different in principle, to make the machine do what someone has already done. Man still makes the machine, determines its laws of behaviour, and is himself the originator of whatever " order " or " purpose " it displays. My only comment is —" so what? " Surely man is a very wonderful person to be able to do all this, but it does not affect one way or other the question whether the wonder of man is the same kind of marvel as the wonder of the machine he makes. In short, I consider the question of what a man-made machine can do totally irrelevant to the Christian doctrine of man.

I consider the real point at issue to be the point brought out by C. S. Lewis in his book on *Miracles*. The thing which distinguishes Man from animals or other creatures (spiritual issues apart) is his Reason. This seems to be intimately bound up with his self-consciousness. I infer the self-consciousness of other men from observing their rational behaviour by the same kind of inference which is habitually used in scientific method. But if it could be shown that the apparent rational behaviour of other men was due to irrational (mechanical) causes, I should at once reject it as valid Reason; and if mechanical causes could be found for everything that appeared to present evidence for their Reason, I think I should be logically compelled to reject the inference that they were self-conscious. For if my Reason could be shown to be the product of irrational causes, then the whole validity of my thought is undermined.

It appears to me that " cogitat, ergo est " is neither more nor less reasonable than " cogito, ergo sum ". Lewis (op. cit., p. 29) quotes Haldane, who put the argument in this form: " If my mental processes are determined wholly by the motions of atoms in my brain, I have no reason to suppose that my beliefs are true . . . and hence I have no reason for supposing my brain to be composed of atoms."

If Reason therefore is not extra-physical, non-mechanical, super-natural, or what-ever way you like to put it, no thought is valid, no science can lay claim to truth, no human being can be held responsible for anything. This is a possible philosophy, but the best answer to it is that nobody has ever found it possible to live consistently with it. The alternative is that Reason is self-authenticating, and is not a product of a mechanism of any kind. I know that I can reason, therefore I am not a mere machine. I infer that because you are in all other points the same kind of object as myself, that you also can reason. It is a short step from here to say that all human beings share this Reason, and a longer one to say that this Reason inheres in God. But this is one of the places where we are forced to consider Christian revelation, for this is one of the things (if Christians are right) in which man is made in the image of God.

<center>AUTHOR'S REPLY</center>

Professor Coulson's lucid discussion of the notion of complementarity goes far, I think, towards answering some of the points raised by other speakers. There are just two comments I might make. First, I fear that I owe an apology to him, and doubtless to other readers, for not making clear my use of the terms " exclusive " and " exhaustive ". Observer- and actor-descriptions are indeed " exclusive " in the sence used by Professor Coulson; but the truth of one does not exclude the other, so that they are not mutually " exclusive " in the sense I intended. An observer-description does not indeed exhaust all that can be said, and so is not " exhaustive " in Professor Coulson's sense; but it does exhaust all that is to be observed, and so is " exhaustive " in the sense I intended. Take the ink away from the page, and nothing is left. The ink-description is " exhaustive ". But the algebra-description is equally " exhaustive " in its own language. And the two are not in any way mutually exclusive.

I was glad that in his analogy of Ben Nevis Professor Coulson represented the Christian's goal, the knowledge of God, by the conception of the whole mountain, rather than by just one of the complementary views. The illustration is easily and often misunderstood to imply that the " Christian view " is just one aspect, on the same footing as others such as the scientific or aesthetic.

I agree with Dr. Barclay that one cannot be compelled by observable evidence to attribute " freewill " to an artefact. But neither is this true of another human being. Any observable evidence can in principle be provided in both cases: there is no restriction to purely logical aspects of thinking. Creation of hypotheses, spon-taneous innovation and the like can all be shown. But I would draw attention once again to the caveats of paras. 4.2, 4.3 and 6 in my paper.

Mr. Barnes's dilemma is illusory. To be able to distinguish beardedness from beardlessness it is not necessary to be able to say at what point in an increasingly hirsute series of chins " a qualitative difference " enters. I certainly agree that we do not in practice deduce the experience of others by logical argument: we rather " resonate " with them in dialogue. In fact we would demand evidence before disbelieving in the reality of their experience. But Mr. Barnes does not show why evidence of their bodily composition should be conclusive or even relevant to such an issue.

Mr. Curling lays his finger on the real problem in combating reductionism, which is the development of rigorous ways of distinguishing complementary from contra-dictory statements. I entirely agree with him; but the understanding of the very nature of this problem in both camps must, it seems, proceed gradatim.

I agree with what I think Dr. Aitken means when he says " Man possesses also spirit ", but the sentence is easily misunderstood. To " possess spirit " is not I think the same kind of possessing as to " possess a body " or to " possess a watch ". The verb " possess " means something different in the two cases. " I possess a body and I possess a spirit " may be a valid statement. " I possess both a body and a spirit "

may quite easily be subtly misleading. As a simple example, " I am in a towering rage and I am in pyjamas " may be a valid statement. " I am in a towering rage and pyjamas " is a misuse of language, inviting the question: " Well then, are the pyjamas inside the towering rage or is the rage inside the pyjamas, if you are in both of them? " There is a serious need, if it is not a duty, for Christians to analyse their use of language on many topics in these terms, without any consequent obligation to take the Gadarene plunge of some contemporary language analysts.

In reply to Dr. Harris, I would say that the act of analysis of a personality, whether one's own or someone else's, must introduce a " perturbation of the system observed " which I suspect may be irreducible in principle in the same sense as quantum " uncertainty " in atomic physics. By " personality " I mean roughly that to which reference is made when we use the word " he ", as opposed to " his body ".

By " reproduced in a new mechanism " I did *not* mean " . . . in the same materials ". On the contrary, just as an algebra problem or a message could be the same whether it were written in ink or chalk, so I would suggest that a personality could be the same whether it were mediated in copper or protoplasm—or indeed in any other physical (or non-physical) structure.

Dr. Cundy's contribution shows that I have not made clear what is meant by " making an artefact behave like a human being ". It is quite inadequate merely to reproduce recorded behaviour, because behaviour includes *dialogue*, and Dr. Cundy would be disappointed if the artefact's response to a question from him were a recorded reply to quite a different question in the past history of the artefact's prototype. No, the problem of securing spontaneous, originative, purposive and reactive behaviour in an artefact is, I believe, soluble in principle, but it is not trivial. The irrelevance of the achievement to Christian doctrine, as to which I agree, does not arise from any necessary inferiority to human capabilities demonstrable in the characteristics of such an artefact.

I agree that it would be difficult to hold Reason to be a " product " of mechanical causes. But it is a fallacy, though a common one, to suppose that a mechanical account of cerebral function *could* have any such consequence. An algebra problem is not the " product " of the chalk that delineates it. The conclusions of even a deterministic computor are not the " product " of the electrical causes of their appearance. To show that they " follow " we must talk, not electrical language, but the language of mathematics or logic. To be prepared to reject the validity of other men's Reason if mechanical causes were found for their behaviour is thus, I submit, totally irrational, and is indeed to share in the real error of the older-fashioned behaviourist. Reason is non-mechanical, not because of Lewis's or any *argument,* but because either to assert or to deny that it is mechanical doesn't make sense. It is as senseless as to ask whether algebra is chalky, or a mathematical conclusion electrically true.

May I end by reinforcing Professor Coulson's plea in rather a general and quite modest form: that whenever we meet two different accounts of what is claimed to be the same thing, we at least consider, critically but habitually, the possibility that the statements are in different complementary " languages " and *may* both be valid. It is not always easy to be sure. We must not admit contradictory nonsense under the aegis of Complementarity. But a wide field of new understanding awaits intelligent exploration in this spirit.