III.—ON THE LOGICAL INDETERMINACY OF A FREE CHOICE

By D. M. MacKay

1. *Introduction*

Nobody knows for certain the extent to which the much-discussed "iron chain of cause and effect" operates in the mechanism of the brain. Those of us whose task is to gather experimental evidence on the question find discussion frequently obscured by emotional prejudice based on the idea that a completely causal explanation, in physical terms, of brain-function would make what Phillip Frank[1] calls "the firm subjective conviction that we have a free will" an illusion.

I will not here discuss the attempts which have been made to invoke Heisenberg's Principle of Indeterminacy as a way of escape. I have argued elsewhere[2] that although physical indeterminacy might indeed give cerebral activity the spontaneity and originality which is associated with one aspect of "free-will", we would be misguided and immoral to regard such indeterminacy as essential to the 'freedom' about which we have this 'firm subjective conviction'.

Nor do I wish to debate directly with those (of whom Professor Frank appears to be one) who despair of the genuineness of such freedom and seek to find room for morality without it. I want rather to attack the alleged ground of their despair, without which the whole pressure behind such arguments disappears. I believe it rests on a presupposition which is widely accepted, but which may nevertheless be questioned for reasons which have nothing to do with physical indeterminacy.

The usual form of argument is as follows: if my physical brain-processes were wholly physically determined, and if my decisions could be inferred uniquely from my brain-processes, then a fully-informed observer of my brain-processes could know the outcome of my choices with certainty before I made them, and my impression of freedom in making these choices would therefore be an illusion, due to mere ignorance of the true state of affairs.

I believe that a close scrutiny of the way in which this argument breaks down may throw some useful light on the whole problem of relating 'mental' and 'physical' talk about human beings.

Our discussion of it falls into three stages. First, we shall take note of some peculiar restrictions on the predictive certainty

attainable in principle by the observer of an agent. In the light of these, we must next question whether ' the true state of affairs ' can validly be identified with the view of the observer, and whether the agent who does not share that view can validly be described as ' ignorant '. The answer will lead us finally to recognise a curious kind of ' relativity principle ' governing talk about the acts of agents, which—if I am right—shows the stock argument above to be fallacious. All this, be it noted, will involve us in no necessity to deny (or affirm) the physically-determinate character of brain processes.

2. *The need for secrecy*

At first sight it may seem inevitable commonsense that if everyone else in the world could be rightly certain that my brain would go through a particular sequence, I should be worse than deluded if I held a different view. If I were to refuse to accept publicly verified evidence that I have red hair, or that I have no back teeth, the commonsense conclusion would indeed seem to be justified. Why then should statements about my brain be in a different category ?

The first clue to a possible answer comes when we reflect that whereas my hair or my teeth are substantially unaffected if someone tells me what he believes about them, my brain, as the organ of understanding, is in a fundamentally different situation.[2,3] Anyone who wished to make a reliable and complete prediction of my brain-activity might in fact have to take great pains to *prevent* my coming to know of it, or even coming under the influence of any relevant factors determined by the conclusion he reaches. The reason is not primarily psychological but logical. His prediction, to be successful, must allow for any relevant effect its formulation and communication will have on my brain ; but these effects could not all in general be calculated unless the prediction itself were already known, so that in general the exact calculation can never be completed. This is in fact a similar logical situation to that treated by Popper[4] in a penetrating analysis of the limitations of computing machines, and although the present argument does not depend on the validity of Popper's thesis, it must be admitted that for at least an important class of cerebral states, no one who intended to *tell* me his prediction of my cerebral activity could remain logically certain of its success. On the contrary, I could quite properly, and on excellent logical grounds, defy anyone to tell me with certainty beforehand the outcome of most of my choices, even if the physical processes

in my brain were wholly determinate in the sense of classical physics and fully accessible to his observation, provided only that the information-receiving system of my brain was causally linked in the right way with my choice-mechanism.*

3. *Conditional certainty*

As an answer to the disputed step in the stock argument, however, this is hardly satisfactory. It does indeed show that my brain could be physically determinate in operation, and yet unpredictable in principle to members of a community whose predictions are accessible to me or are determinants of factors affecting my decisions.

This amounts to saying that in such a case no one in fact could be validly certain that my brain would go through a particular sequence. What it shows is not that I can disregard other people's certainty when they are entitled to it, but that in one important case—namely where they wish to tell me, or cannot help affecting me by, their prediction—they could not achieve such certainty, even if the "iron chain of cause and effect" were known to operate.

We are thus still left with our main question: If (by preventing me from knowing or being influenced by their prediction until after the event) everyone else *could* validly be certain how I shall choose, can I justly claim that my choice is still free? Surely, it might be argued, I am simply showing my, perhaps necessary, ignorance of the true state of affairs; I am under an illusion?

A preliminray skirmish might well take place over the concept of certainty as used here. When I am certain that 13 is a prime number, or that London (England) is on the Thames, I feel I am up against facts that everyone must reckon with. I don't care who knows them. But the 'certainty' of a secret prediction P of an agent A's subjectively free choice is subtly different. It is only a conditional certainty—a certainty whose value varies according to the person contemplating the prediction. In short, despite the public character of the proposition P, truth, reliability, or certainty cannot here be attributed to it in isolation, but only to P-in-relation-to-the-person-entertaining-P. It is *people*, not propositions, who are certain, either rightly or wrongly. We cannot automatically attribute certainty even to propositions which they are certainly right to maintain.

Nor do we escape this qualification by modifying the prediction as 'Provided A does not get to hear of this, he is certain to do X';

*But see section 7 below.

for this new proposition in turn will have a different truth-value for A and for the silent onlookers. At most, A could say of it 'This is out of date'. For A, it has no predictive information-content.

There is a further logical qualification of the 'certainty' of the silent onlookers. If we asked them (quietly, so as not to disturb A) 'How can you be certain that A will not be affected by your calculating P?' their answer might have to refer in part to their *intentions*: 'we *intend* not to tell him or influence him'. Their 'certainty' that A will do X is thus conditional on their own fulfilling certain intentions; and (by the same line of argument) they could never in general have all the up-to-date information about their own brain to predict this with even the qualified certainty which they have about A's action. Can they be *certain*, for example, that a wave of pity for A will not sweep over them and impel them to intervene? Presumably their best course would be to destroy all outgoing lines of communication with A, as the practical equivalent of hardening their hearts. From a moral standpoint this action would appear to give them a significant share in responsibility for A's choice, since they are taking deliberate steps to ensure that it shall be X when they know or fear that it might otherwise be not-X. But it would I think remove one element of uncertainty as to their success; and our purpose will be served equally well if we suppose that the outgoing communication channels are cut by some third agency.

4. *Logical indeterminacy*

Assuming then that all this is done, and the onlookers are powerless spectators of a sequence whose outcome, for them, is 'certain'—what can A validly believe about the choice he has yet to make? If he knows nothing of the onlookers and their predictive knowledge, he may ignorantly suppose that no one can tell how he will choose. In this, *ex hypothesi*, he would be simply mistaken, and if his subjective conviction of freedom amounted to no more than this belief, it would indeed be illusory.

Suppose, however, that A knows all about the onlookers, and has sufficient past evidence (say from predictions secretly written down and shown to him after the events they predicted) to convince him that they have these predictive powers, so that he believes they *can* be secretly certain of the outcome of a choice he has yet to make. Can he still validly believe he is choosing freely, not only in the sense of feeling unrestrained, but also in the sense of

having a genuine option logically open to him ? Or is he merely ignorant of what the onlookers know to be really the case, and so still under an illusion ?

It seems to me that he is not merely entitled to believe that his choice is undetermined, but that this is the only view of it that he may validly hold until he has made it. For him, despite all we have just granted, any proposition purporting to specify his choice in advance is simply *logically indeterminate* until he makes the choice. In other words, I wish to deny, as itself fallacious, the view that his subjective conviction of freedom is based on ' ignorance ' of the ' true state of affairs '.

The argument is twofold. First, we must question the identification of the onlookers' view in this case with the ' true state of affairs ' ; and secondly, we must ask whether A's lack of an onlooker's predictive knowledge of his own brain-workings can properly be described as ignorance. I shall argue (1) that the onlookers, despite their numerical superiority, are here disqualified equally with A from claiming to describe the ' true state of affairs ' ; and (2) that as the crucial belief on which the onlookers base their prediction has here no predictive information-content for A, it is a misuse of words to describe A's relationship to it as ' *ignorance* of the true state of affairs '.

5. *The elusiveness of ' the true view of affairs '*

The first question might seem to strike at the root of our whole idea of objectivity. If the true view of affairs is not the view from Olympus, we may ask, what standard of objective reality is left ? I do not propose now to follow this metaphysical sidetrail, fundamental though I believe its goal to be.*

All I wish to do here is to press home the implications of the *necessary secrecy* surrounding the onlookers' knowledge. There is no dispute that *they* are right to believe what they do about A's brain-processes. But even they would insist that *A* would not be right to believe the same, since a precondition of its validity is that A must not be influenced by it. Clearly then the onlookers' view represents a true description of the state of affairs only *for the onlookers*, since if it were universally true, A would be wrong *not* to believe it. From a logical standpoint their numerical superiority is irrelevant ; and A need not be terrified at finding

* Let me only point out that the ' view from Olympus '—the view of the non-participant observer, which I am rejecting as an absolute standard, is something very different from, for example, the ' knowledge ' attributed to God in Christian theology, which in a special sense combines that of spectator and participant.

himself out of step. There are essentially just two parties in the situation, one looking on and one making a choice. The onlookers' view can logically be recommended as valid only to one of the two.

What then of the other party? A's view is necessarily different from the onlookers'—but it is not arbitrarily different. Indeed there is *ex hypothesi* a definite correspondence between the two, so that the validity of A's description could be checked, afterwards, in terms of the onlookers'. But—and this is the point—the check is not for identity between A's and the onlookers' descriptions; on the contrary, in order that A's should be valid, the onlookers' must not in general say the same thing—and conversely.

Thus on the one hand, the idea that either party can give a universally-valid description of the 'true state of affairs' in this case is false; on the other hand, any idea that this proves there is *no* 'true state of affairs' is invalidated on the assumption that the two descriptions stand in a rigid relationship. We might call them two different but related 'linguistic projections' of one and the same state of affairs. It is perhaps not surprising, if tantalising, that no single standpoint, whether of onlooker or agent, appears to allow us to put into words the whole truth about ourselves.

6. *The 'ignorance' of a free agent*

What then of our second question? Can A be said to be 'ignorant' of a fact that the onlookers 'know for certain'? In one sense it is doubtless true that A does not know what they know: in fact they have taken care that he does not. Normally we can translate 'A does not know what B knows' by saying 'A is ignorant of a fact known to B'. But here we are in trouble, for we have already seen that the onlookers' prediction P does not have a unique factual status in isolation. The interesting point which emerges is that what we are tempted to call A's 'ignorance' *would not be remedied by supplying him with the proposition P describing the state of affairs of which we are trying to say he is ignorant*, since P would lose its factual status if A were to entertain it. In short, the onlookers *have no predictive information to give A*, even if they would. A may not realise this, and may even 'wish he knew' what they know; but in respect of predictive information his wish is based on a fallacy—the fallacy of supposing that what he wants to know is a universal fact. The truth would seem to be that at this point there is no gap in his knowledge;

the place of the onlookers' knowledge is already preoccupied for him by the knowledge that the choice awaits his decision. To make room for it, he would have to resign from his role as agent: but then the choice would not be made.

7. *Criteria of logical indeterminacy*

Obviously not all of the onlookers' knowledge is of the peculiarly evaporative character we have been discussing. All of us constantly gain useful information about ourselves from others which we would find it hard to obtain unaided. We must then try to make explicit the conditions under which propositions are not merely unknown by A but *logically indeterminate* to A.

The general condition is easily established. We have been working out the assumption that activities such as formulating or believing a statement require corresponding brain-processes to take place. Consider now the class of statements that describe the subject's own brain-processes or activities causally linked therewith. Among these, there will be a subclass which cannot be formulated or cannot be believed by the subject, without brain-activity that interferes with the brain-process they describe. It is unnecessary to assume that *all* such interference nullifies the statement.[4,5] One of the best ways of making an angry man more angry may be to tell him he is getting angry! What we can always say is that in such a case the formulating or the believing of the statement becomes one of the factors determining its truth or falsehood, so that it cannot claim unique factual status in isolation. But where the brain-activity in formulating a statement is by itself sufficient to make it false and so out of date, an up-to-date formulation becomes impossible to the subject.

Similarly, where the brain-activity in believing a statement is by itself sufficient to make the statement false (though disbelief might not suffice to make it true), the statement for the subject is *logically incredible* in the strict sense of the term.

My suggestion, then, is that our 'firm subjective conviction of freedom' is not primarily a belief about the unpredictability of our brain-processes but is the entirely justifiable corollary of these peculiar logical facts. For us as agents, any purported prediction of our normal choices as 'certain' is strictly *incredible*, and the key evidence for it *unformulable*. It is not that the evidence is unknown to us; in the nature of the case, no evidence-for-us at that point exists. To us, our choice is logically indeterminate, until we make it. For us, choosing is not something to be observed or predicted, but to be done.

I thus find myself in a curious blend of agreement and disagreement with Professor Ryle, who draws attention to some of the same peculiarities in his analysis of self-prediction in *The Concept of Mind*. On the one hand (p. 197) he agrees that 'the fact that ... my future is systematically elusive to me has ... no tendency to prove that my career is in principle unpredictable to prophets other than myself'. Yet he elsewhere (pp. 77-80) uses analogies of games of chess and billiards, and of the writing of prose, where the point is repeatedly made that from the initial conditions and the rules, the subsequent course cannot be uniquely determined. It is difficult not to construe this passage as a denial of the predictive sufficiency (for the onlooker) of an onlooker's mechanical data—a denial which I would argue to be philosophically needless. My chief disagreement—or perhaps difference of emphasis—however, comes where Ryle speaks of the self-predictor as 'having to overlook at least one of the data' relevant to his prediction (p. 197). It is the presupposition underlying this way of expressing the case that I have been chiefly concerned to question.

As will now be clear, this is the presupposition that the acts of an agent can be characterised, like the workings of any physical object, by a set of data expressible uniquely in propositions universally valid in and mandatory on the linguistic community, so that it always makes sense to speak of the agent in particular as 'overlooking' some or 'ignorant' of others. This goes with a more basic notion that might be called the presupposition of *transferability*: that if A agrees that B is right in believing P, A logically commits himself to P and to all consequences deducible from it. Despite its obvious validity in most contexts, we have seen that this can break down where P is an assertion about an agent as viewed by an observer. If it were logically (as distinct from physically) possible to exclude the agent from membership of the linguistic community, then the presupposition could stand, as it does in face of an inanimate agency. If, however, we are concerned to ask what a *member* of our linguistic community may validly believe (*e.g.*) about the actions he would subjectively term 'free', I see no escape from the necessity to make room for two complementary stories, one validly believable only from the standpoint of detached onlookers, which the agent would be wrong to try to believe until after his action,* the other validly

*In retrospect, of course, the agent can join the onlookers (e.g. in witnessing a moving film of his own brain processes) and share in their 'outside' view of his physical past as 'determined'. Past and future have an asymmetric logic for an agent.

believable from the standpoint of the agent,—and by the onlookers too insofar as they permit themselves the exercise of sympathetic imagination. The two are not mere *translations* of one another, since it is *what is asserted* by the onlookers, and not just its conventional *form*, that is logically unacceptable to the agent; yet (*ex hypothesi*) there is a definite *transformation rule* by which (in retrospect at least) the accuracy of one may be checked in terms of the other.

8. *Conclusion*

Implicit in this treatment of a well-worn topic there have been one or two ideas which it might be well to make explicit in conclusion. First, we have found it expedient to shift our emphasis still further than some modern philosophers from discussion of the truth of certain *propositions* in the abstract, to discussion of the validity of the activities of *formulating and believing* these propositions. By doing so we have been able to express in non-contradictory terms what would otherwise have had to be expressed in paradox. It seems possible that this shift of emphasis might be rewarding (and far from destructive) in some other metaphysical and theological contexts. Second, and not unrelated, there has emerged the idea of *transformation rules* according to which it may be essential for A's belief to differ from B's in order that both may be valid. This denial of simple transferability constitutes a kind of philosophical Principle of Relativity, very different from that exaltation of the arbitrary which goes by the name of 'moral relativism'. It resembles rather Einstein's physical principle in its insistence (i) that only one rigorously prescribable belief is valid for A if B's belief is also valid, but (ii) that the validity and meaningfulness of a belief may depend in a definite and rigorous way upon who entertains it. It differs, however, in giving no guarantee that A can even formulate from his standpoint the belief that would be valid for B (until it is out of date) and in making no assumption that their situations must be symmetrical.

Despite its radical consequences this principle has no sanction for the idea of 'two kinds of truth'; and of course it leaves the bulk of public human discourse unaffected. But it does suggest that—and why—the traditional method of *comparing notes* in order to 'arrive at the truth' must break down in certain special cases, leaving the truth in such cases incapable of unique and universally valid expression.

This brings me to a final point. To discover that nothing

physically odd *need* accompany a free action, though we have no way as yet of knowing whether it does, undoubtedly disposes of a potential 'mystery'; but it is far from showing, as Ryle (*loc. cit.* p. 198) avers, that 'there is nothing mysterious' about our self-consciousness. On the contrary, I suggest that it only helps the really mysterious aspect of selfhood to be more squarely faced: for to see ourselves as part of an objective situation which yet must elude unique description is, I think, to be aware of a genuine mystery.

REFERENCES

1. FRANK, P. 'Present Role of Science,' *Atti del* XII *Cong. Int. di Filosofia* (1958) p. 5.
2. MACKAY, D. M. 'Brain and Will', *The Listener* (9 and 16 May 1957).
3. MACKAY, D. M. 'Information Theory and Human Information Systems', *Impact of Science on Society*, **8**, (1957), 86-101.
4. POPPER, K. R. 'Indeterminism in Classical and Quantum Physics', *Brit. J. Phil. Sci.* **1**, (1950), 117-133 and 173-195.
5. WORMELL, C. P. 'On the Paradoxes of Self-Reference', MIND, lxvii (1958), 267-271.

Summary:

It is sometimes suggested that physical determinacy in the brain would prove our subjective conviction of freedom to be illusory. Even though it is granted that an observer's prediction could not be shared with the agent and remain valid, it is held that the agent would be merely ignorant of the true state of affairs.

The present paper contends (*a*) that the observer's view in this case cannot claim to represent uniquely the true state of affairs, (*b*) that despite linguistic appearances there is no predictive information in the observer's view of which the agent may properly be said to be ignorant.

The suggestion emerges that in this domain conventional ideas of arriving at the truth by comparing notes require to be radically revised.

University of London