American Philosophical Quarterly
Volume 13, Number 3, July 1976

IV. AN IMPORTANT NECESSARY DIFFERENCE BETWEEN PEOPLE AND MINDLESS MACHINES

GEORGE SCHLESINGER

I

I SHOULD like to explore the question whether or not there is a basic difference between men and mindless machines in that a certain kind of freedom is enjoyed by the former and not by the latter. By human freedom I do not mean just that a human being can sometimes do exactly what he wants in contrast to a mindless machine, of which it can never be said that it does what it wants, since in order to want anything it would need to have a mind. This would add absolutely nothing to saying that mindless systems are mindless. The freedom of men I want to discuss here is to be understood in the sense, that some human acts are in principle unpredictable. This is to be contrasted to the unfreedom of machines with respect to whom I shall claim that there does not exist a comparable unpredictability.

First, to the essential unpredictability of some human acts: Some philosophers have pointed out (e.g., Michael Scriven) that a man who is determined to act counterpredictively cannot have his actions predicted and also have the prediction communicated to him. The communicated prediction is assured not to come true by virtue of the fact that the counterpredictively motivated person, having learnt how he is supposed to act, will decide to act differently. This kind of unpredictability, however, does not seem to be connected with the humanity of man. A machine programmed to act counterpredictively would also act in a way which would make communicated predictions impossible. In order to show that men are different (which by the way was not the aim of these philosophers) it would have to be proven that machines cannot counterpredictively be programmed.

I shall be discussing another instance of an unpredictable human act that I have shown to exist.[1] Let me begin by briefly repeating the essential features of the proof that in a certain kind of situation a human act is in principle unpredictable. Suppose an agent is confronted with two boxes, the first one containing $1000 and the second containing either $M or nothing. The agent is allowed either of two choices: he may take box II only or take both boxes. There is a person who is described as a perfect predictor on the basis of an impeccable record of having successfully predicted 24 hours in advance in hundreds of thousands of cases of similar games what the choices of a vast variety of people in all sorts of states of mind, would be. The agent knows this and he also knows that if this person predicted 24 hours ago that the agent would choose box II only, then he put $M in box II, and if he foresaw that the agent would choose to take both boxes then he put nothing in box II. The problem as to what is more advantageous for the agent to do is known as Newcomb's problem of choice. What I have shown amounts to saying that this story leads to a contradiction and hence by reductio it must be an impossible story. The element which is by far the most likely candidate to be the impossible part of the story is the claim that there is an infallible predictor. Hence there can be no infallible predictor of all human choices.

The contradiction referred to is with respect to the question whether it is more advantageous for the agent to take box II only than taking both boxes. On the one hand there is a straightforward argument for saying yes. Everybody in the past who took box II only, found $M in it while the thousands of people who took both boxes ended up with a thousand dollars only, since they found box II empty. Hence—given that the predictor is infallible—there are only two possibilities for the agent: to take box II only and gain $M or take both boxes and be satisfied with $1000. The greater advantage lies obviously in taking box II only.

On the other hand there is a decisive argument for saying that it is less advantageous to take box II only. Before presenting this argument let me

[1] "The Unpredictability of Free Choices," *British Journal for the Philosophy of Science*, vol. 25 (1974), pp. 209–222.

point out that there are several versions of a very tempting but faulty argument for taking both boxes. One goes like this:

Let m = there is \$M in box II
t = the agent takes both boxes
b = the agent is better off than he would be if he chose otherwise.

We have (i) $m \supset (t \supset b)$
(ii) $\sim m \supset (t \supset b)$
(iii) $m \vee \sim m$

Hence $t \supset b$, that is, if the agent takes both boxes he is better off than he would be if he chose otherwise. Q. E. D.

The three premisses of the foregoing proof must be accepted for the following reasons: Given that once the predictor put \$M in box II, nothing will cause it to disappear and it remains there until the agent opens the box, it follows that if m is true, then by taking both boxes the agent gains \$1,001,000 while by taking box II only he gets less, \$M only. It may of course be pointed out that, given that the predictor is infallible, it is certain that if the agent takes both boxes then there just is no \$M in box II. This however only further ensures that (i) is true! For what we are saying now that if t is true then m must be false; but the falsity of m renders the truth-functional $m \supset (t \supset b)$ true. Now (ii) is of course true since, if box II is empty, then if I take both boxes I get at least \$1000, otherwise I would get nothing. And (iii) is simply true by the law of the excluded middle. But the conclusion logically follows from these three premisses.

Yet it would be wrong to conclude from this that in order to maximize his gain the agent should take both boxes. To see this clearly we shall consider briefly another situation which obtains in what we shall call Game 2, while to the previous game we shall refer as Game 1. In Game 2 the agent has the same two choices as in Game 1, and while once more box I contains \$1000, box II is definitely empty. Instead of a predictor there is an observer who, if he observes that the agent has taken box II only, gives the agent \$M and if he observes that he has taken both boxes he gives him nothing. The fundamental difference between Games 1 and 2 is that with respect to the latter there is absolutely no doubt that there is nothing impossible in the story. This can be verified empirically: unlike perfect predictors, there is no shortage of people competent to observe whether an agent has taken one or two boxes, thus any two people (with lowered stakes if necessary) can play Game 2. Thus it is obvious that the description of Game 2 does not lead to any contradiction; and to the question, is it advantageous for the agent to take box II only, the answer is definitely, yes. But here too an argument, parallel to the one we had before for saying that the greater advantage lies in taking both boxes, exists. For let m' = the agent will receive \$M from the observer. Then, even now, before the agent has made his choice m' is either true or false. Thus we have:

(i) $m' \supset (t \supset b)$
(ii) $\sim m' \supset (t \supset b)$
(iii) $m' \vee \sim m'$

Hence $t \supset b$, that is, if the agent takes both boxes he is better off. Q. E. D. There is no doubt however here that to conclude that the agent would maximize his gain by taking both boxes is wrong. Why?

Let us denote the situation in which the agent receives \$M from the observer by 'ϕ' and the situation in which he gets nothing by 'ψ'. We shall call "advantageous ϕ" that ϕ in which the agent gains the \$1000 which is in box I and "disadvantageous ϕ" that ϕ in which he fails to gain the \$1000 contained in box I. Similarly we shall have "advantageous ψ" and "disadvantageous ψ." The foregoing constructive-dilemma-argument only shows that by taking both boxes the agent will always place himself in an "advantageous" situation (which in fact turns out always to be advantageous ψ) but it so happens that disadvantageous ϕ is preferable to advantageous ψ! It is clear therefore that in order to maximize his gain the agent should bring about disadvantageous ϕ rather than advantageous ψ, that is, go for the million dollars rather than for the thousand dollars. Exactly for the same reason it is obvious that in Game 1 too the argument for saying that the greatest advantage lies in taking both boxes is invalid.

Another interesting but also wrong argument is advanced by Robert Nozick. His argument is essentially that the player cannot now through his choice change what the predictor (whom he calls "the Being") has done:

> The dominance argument. The Being has already made his prediction and has either put the \$1 million in the second box or has not. The money is either sitting in the second box or it is not. The situation whichever it is, is fixed and determined. If the Being put the million in the second box you will get \$1,001,000 if you take both boxes and only \$1 million

> if you take only the second box. If the Being did not put the money in the second box you will get $1000 if you take both boxes and no money if you take only the second box. In either case you will do better by $1000 if you take what is in both boxes rather than only what is in the second box.[2]

But of course while it is true that the situation is now fixed, *how* it has been fixed to begin with may very well be dependent on what the player does now. By taking box II only the player can ensure that there are $M in box II from the very beginning. The fact that the player's present choice may determine whether or not the predictor put money in box II in the first place does in no way imply that the player by his present actions can change the past to have been different from what it has been. It only implies that he may change the past to have been different from what it would have been if he would now act differently from the way he is actually going to act. Once more Game 2 may be used as an illustration to see that the argument is wrong. If it is true that the observer is going to give the player $M then it is impossible for the player to change this since the future cannot be changed to be different from the way it is going to be. But of course by taking both boxes the player makes sure that it has never been true in the first place that the observer is going to give him at the end of the game a million dollars.

There is however a different valid argument for taking both boxes in Game 1, an argument which cannot be applied to Game 2. Suppose a friend, who is sufficiently intelligent and a perfect well-wisher of the agent is allowed to look at the contents of box II. If he sees that the box is empty he will advise him to take both boxes so as to get at least a thousand dollars. If he sees that there is money in box II he will still advise him to take both boxes since he knows that by doing so the agent will not make the money diappear from the box, something that has never happened before (except in this case, of course if the friend believes in the infallibility of the predictor he will sadly note that his good advice is surely going to be disregarded). Now it does not matter that in fact there is no such friend at hand, since it is certain that if there was such a friend he would advise the agent—no matter what—to take both boxes. It logically follows that it is best for him to take both boxes. It is self-contradictory to assert that it may not be in my best interest to follow the advice of a sufficiently well informed and intelligent, perfect well wisher. Since it is known for sure what the advice of such a well-wisher would be it follows analytically that the agent's best interest is served by taking both boxes. It is quite useless to try to argue as some people have tried that probabilities are relative to one's state of knowledge, and the agent's state of knowledge is different to that of his friend; hence, while relative to the friend's information it is a certainty that it is best to take both boxes, this may not be so relative to the inferior state of knowledge of the agent. It is entirely irrelevant for the agent to know on the basis of exactly what information his friend advises him to take both boxes; it is sufficient for him to know for sure that his friend believes it is better for him to take both boxes, in order to know with absolute certainty that indeed it is better for him to take both boxes.

Thus we have a deductive argument for saying that it is more advantageous for the agent to take both boxes; hence, a contradiction. Therefore, we must withdraw our assumption that an infallible predictor is possible. From this it follows that not all the acts of a free agent are predictable.

In the paper referred to, I have also shown that not only can there be no perfect predictor but even a weak predictor is impossible. A weak predictor is defined as one who predicts that the agent will take both boxes with probability p if he takes box II only, but with probability $p+\epsilon$ when he indeed takes both boxes, where ϵ is a finite number.

II

Suppose an attempt is made to resist our conclusion that Game 1 illustrates a case in which the acts of a free agent are unpredictable in the following way: The friend who advises to take both boxes even if there is money in box II is giving an advice which the agent cannot follow. On the supposition that the predictor is infallible, when there is money is box II, the agent just cannot take both boxes. Therefore we should continue to maintain that the predictor is infallible, in which case we have an argument for taking box II only. There is no valid argument leading to a contrary conclusion, since the argument from the perfect well-wisher does not hold: it is not analytically true that it is in the agent's best interest to follow an advice which he definitely is incapable of following!

But the agent must ask himself: is it in his best interest to take box II only? The answer to this

[2] *The Scientific American*, vol. 230 (1974), p. 102.

question must be, no, since the perfect well-wisher says so, or would say so. Hence the agent must try and take both boxes. Should he indeed find that he is just incapable of doing so, then there is immediate proof that we are not confronted here with an infallible predictor of free choices, since the way in which he ensures that his prediction that the agent will take box II comes true is by forcing him, if necessary against his will, to do so. On the other hand, if he feels that he is capable of taking both boxes, he should certainly do so, for it is in his best interest to do so, given that this is what the perfect well-wisher advises. But if taking both boxes is indeed what leads to the maximization of gain, then the predictor cannot be infallible.

Before I come to discuss what happens when the player is a computer, and show that the unpredictability in principle we have discovered here is intimately connected with the mindedness of humans and does not arise in the case of machines, I should like briefly to consider an interesting point made by Professor Abner Shimony in correspondence. Shimony raises the query that perhaps one could claim that there is no inductive basis for the agent to argue that it is better for him to take box II only, since he does not have any solid evidence that the predictor is infallible. For what after all is the alleged inductive evidence for the infallibility of the predictor ? The answer is the fact that the predictor has not predicted wrongly the choice of thousands of people placed in the same situation as the present agent. But who could have been in the same situation as the present agent ? Not the first person whose action the predictor predicted, because the first person did not have any data yet about the success of the predictor. Not the second person, because even though he had data about the success of the predictor regarding the first person, that first person was not in a position to make an inductive inference—and therefore the inductive inference of the second person is essentially different from the present agent's inductive inference. This argument can now be reiterated, with the conclusion that no one before the present agent was in a position to make an inference about the success of the predictor concerning persons making decisions in situations which were really like the situation he is in now.

First let us see what Shimony's point amounts to. He agrees that the argument from the perfect well-wisher is decisive and therefore he agrees that if the agent wants to maximize his gain he should choose taking both boxes. Also as to the way the agent should regard the predictor, he agrees with us : he should not assign to him any special ability to predict what his choice is going to be. What he questions however, is whether I am entitled to infer from this story that a perfect predictor does in principle not exist. May be all we can conclude is that in principle there just cannot exist any *evidence* that we are dealing with a predictor who is perfectly capable to predict what the choice of the next agent is going to be. This conclusion, while interesting in itself, is considerably less fargoing than the conclusion I wished to draw concerning the nature of human choice.

I believe that several replies can be made to this. True enough, the state of mind of the present agent is not exactly similar to the state of mind of any other person who has played Game 1 before. But few would insist that in a valid inductive reasoning absolutely perfect similarity must exist between the members of the sample class and the instance about which an inference is to be drawn. Such an insistence would invalidate practically all inductive reasoning.

Another reply is this : suppose our agent is agent number n and knows it, but quite a few previous players thought by mistake that they were agent number n to play Game 1. Under such conditions the present agent *is* in exactly the same situation as other players were in the past and would have inductive evidence that the predictor is infallible.

Finally, it is by no means the case that the only inductive evidence for the proficiency of the predictor we could have is from the cases of past players who´were in the same situation as the present player. The predictor may claim that he is a precognitor who does not predict the future in a Laplacian manner on the basis of his knowledge of the present conditions and the laws of nature, but by directly perceiving events yet to occur like others perceive events occurring in the present. There may not be any good evidence—that anyone is equipped with such kind of vision to any degree, but it is certainly conceivable that there be inductive evidence—unconnected with Game 1 —that many people, including our predictor, have a vivid perception of all future events.

We are forced therefore to the conclusion that the reason why we could not accept any inductive evidence for the reliability of our predictor does not lie in the intrinsic weakness of any such possible evidence but in the existence of the decisive deductive argument against the possibility of such a predictor.

III

Now let us see what happens when Game 1 is played with various types of machines. Let us have once more a predictor who undertakes to put $M into box II in case he predicts that the player will take box II only, and to put nothing in box II otherwise. The player is a machine of type α which is programmed to take always box II only. There are clearly no difficulties for our predictor here, he can observe his rules and successfully predict that α will take box II only. No contradiction arises here. To the question would α be better off if it took both boxes, the answer is, yes. If it took both boxes it would get $1,001,000, but of course it cannot take both boxes since it is an α-machine. There is no contrary conclusion based on inductive evidence, since no α-machine ever takes both boxes and ends up with $1000, as no α-machine ever takes both boxes. There is no empirical basis for saying that the counter-factual "If α had taken both boxes it would have found the second box empty" is true. After all the only reason why our predictor is infallible with respect to the choices of an α-machine may be that α-machines can be relied upon always to take box II only. There are no grounds to expect that if the impossible happened and an α-machine took both boxes the predictor would foresee this and leave the second box empty.

Now let the player be a β-machine which is programmed always to take both boxes. The predictor can once more keep to his rules, predict with perfect confidence that β will pick both boxes and hence leave box II empty. Would it be better off if it picked box II only? No, if it picked box II only it would end up with nothing, but of course it cannot pick box II only. No contrary claim could be advanced on the basis of the predictor's principle always to reward those who are content with taking box II alone by a million dollars. There is no evidence that if what is thought to be ruled out in practice happened and a β-machine took box II only, then the predictor would anticipate this and put a million dollars in it. The very basis for the competence of the predictor with respect to the choices of the β-machine is that the β-machine can be counted on never to take box II only. Thus to the question "Would a β-machine be better off if it took box II only" there are no contradictory answers—one based on the perfect well-wisher and one based on induction—and therefore we need not deny that the choices of a β-machine are predictable.

Now the following question may seem to arise: Having free will means only having the ability to act in accordance with one's will; a free agent is not constrained to act contrary to his will but does exactly what he wants. I certainly do not wish to define a free agent as somebody who is necessarily undetermined or unprogrammed to will what he is going to will (otherwise the unpredictability of the choices of a free willed agent would be established by fiat, whereas I claim it interestingly follows from the arguments of this paper). Thus it is not illegitimate to claim that the human agent presently facing the two choices of Game 1 is determined (or programmed) to will either to choose box II only or to will to choose to take both boxes. But then in essence, our agent is either like an α-machine or like a β-machine and we have just shown that in either case it can be maintained without fear of contradiction that the predictor is infallible.

The answer to this is that while it may be maintained that the choice of the agent is predetermined and while it is undoubtedly true that he must either choose box II only or both boxes, it does not follow that the player must be either in α-state and then the same argument which we applied to α-machines apply to him or else in β-state in which case the argument applied to β-machines apply to him. For suppose he ends up taking both boxes in which case we will be inclined to say that he was in β-state all along. Let us ask whether he would have been better off if he had taken box II only? We can obviously not use the same reply here we gave to this question when it was asked about the β-machine since the human player *could have* if he wanted, taken box II only and if he thought this would benefit him more he probably *would have* taken box II only. Thus in his case there is empirical evidence for saying that everybody who was in a state like him and ultimately decided to take box II only found a million dollars in it. Thus in the case of the human player, unlike in the case of α and β-machines, we do get involved in a contradiction if we accept the available inductive evidence for the predictor's ability to foretell his choices.

IV

This brings us at once to the question: What about a γ-machine which is a machine programmed to make the choice which will maximize its gain? How will such a machine act? It seems that—assuming it is an intelligent machine—it should reason that taking both boxes will result in the highest amount of money to be gained and thus take both boxes. The predictor can of course know that this is how the γ-machine will argue and predict that it will take both boxes and leave the second box empty. But does a contradiction not arise here as well ? Is there no argument also for saying that the machine would gain more by picking box II only ? The answer is that there does not seem to be any valid case for saying that it would gain more by taking box II only, in spite of the fact that every player who takes box II only gains a million dollars. None of the players who is a γ-machine who takes box II only gains a million dollars, since no γ-machine takes box II only. There is no empirical basis for maintaining the truth of the counter-factual 'If γ would have taken box II only it would have found a million dollars in it' for there is no basis to assume that if entirely unexpectedly a γ-machine took box II only the predictor would still be able to predict the machine's choice. In fact the competence of the predictor with respect to the choices of a γ-machine is guaranteed only by the fact that a γ-machine can be depended on always to take both boxes.

From what has just been said it may appear that one could advance an argument to the effect that maintaining that a perfect predictor with respect to a human agent may exist does not lead to a contradiction either. For suppose, as we have already supposed, that our human agent wants to maximize his gain; then he is in essence, as we said, a γ-machine. Then because of the decisive argument from the perfect well-wisher he must choose to take both boxes. There is no argument, so it might be claimed, which leads to the contrary conclusion, that he would be better off by taking box II only. True enough, every player who takes box II only, gains a million dollars while every player who takes both boxes gets only a thousand dollars, but nobody who is a player and is in a γ-state and who takes box II only gains a million dollars, since nobody who is in γ-state takes box II only.

The cases of a human-player and the computer-player are however fundamentally different. In the case of the human being it makes full sense to say that, though he wants to do an act which achieves a given aim he actually performs a different act. Of a non-minded machine one can of course not say this. Consequently the many thousands of people who in past games chose to take box II only, may well include many who were in γ-state, that is they were desirous to maximize their gain, yet took box II only. Therefore the present agent who wants to maximize his gain does have good grounds for arguing that if he were to take box II only he would be better off than if he took both boxes, since in the past people who were in a γ-state like himself and took box II only ended up with a million dollars, while those who took both boxes got only a thousand.

Thus, man's possession of mental properties by virtue of which he may *want* to perform an act designated by logic as the best act, but *actually* performs a different act, is the crucial factor in rendering the predictor impotent. In the case of the machine which has no wants, there is no contradiction in maintaining that the predictor is infallible and the best thing is to take both boxes, since no other argument exists for saying that it would be better off taking box II only. In the case of the human agent there is such an argument. Hence, a contradiction, and we must conclude that a reliable predictor with respect to him cannot exist.

An attempt could conceivably be made to defend the claim that an argument for saying that the player would be better off by taking box II only, lacks in the case of the human agent too. All those people in the past who have wanted to maximize their gain yet took box II only, did apparently not appreciate the argument from the advice of the perfect well-wisher which logically implies that in order to achieve their goal they should take both boxes. The present agent however realizes this and thus he cannot argue that if he took box II he would be better off on the basis that everyone with a state of mind like his who took box II only found a million dollars in it. The reason is that nobody who was in a similar state of mind as the present agent—that is, that he was both desirous to maximize his gain and was aware of the deductive argument for taking both boxes—has ever in the past taken box II and found a million dollars in it, since no such person has chosen to take box II only.

This attempt to claim that in the case of the human player there is no empirical evidence either that he would be better off if he took box II only consists essentially in dividing γ into γ_1 and γ_2. An agent is in γ_1-state if he wishes to maximize his gain but does not realize that logic requires that in order to do so he must take both boxes. An agent is in γ_2-state if he is desirous to maximize his gain and realizes that logic requires that in order to do so he has to take both boxes. What is being claimed then is that nobody in the past who was in γ_2-state took box II only and hence there is no inductive evidence for saying that anybody who like the present player is in γ_2-state and took box II only he would end up with a million dollars.

But the erroneous assumption underlying such an attempt is that a human agent who is aware of the deductive argument for taking both boxes will not ever choose to take box II only. It is easily imaginable that an agent strongly desirous to gain the maximum amount of money and aware of the argument that a perfect well-wisher who could observe the contents of box II would certainly advise him to take both boxes, yet decides to take box II only. He might end up doing so for any number of reasons. For example he may decide (fallaciously) that this argument does not after all lead to the conclusion that he would be better off taking both boxes, or that logic is not a good guide in practical matters. Most importantly, he may decide to take box II only without feeling obliged to give a rational defense—even to himself—for his choice. It is simply not true then that anybody in γ_2-state necessarily takes both boxes. Therefore, the sample class of the present agent may contain past players who were desirous to maximize their gain and were aware of the deductive argument for taking both boxes but have decided not to pay attention to it and took box II only, ending up with a million dollars. Thus, on the assumption that the predictor is reliable, the present agent has a solid case for saying that it is in his interest to disregard the argument from the perfect well-wisher and decide to take box II only. He is entitled to maintain the generalization: everybody, who like himself, wanting to maximize his gain and being aware of the argument for taking both boxes, who decides to ignore that argument and takes box II only gains a million dollars, while those who do pay attention to that argument and go for both boxes end up with a thousand dollars only. But then there is the argument leading to the contrary conclusion that he be better off taking both boxes. Hence a contradiction, and we arrive at the conclusion that a reliable predictor does not exist.

V

Finally we must ask, what about a δ-machine which is programmed to randomize its choice through a strictly indeterministic physical process? The immediate answer seems to be that indeed if the player is a δ-machine then its choices are unpredictable, but that this does not affect our thesis concerning the unique freedom enjoyed by human beings. There remains after all a very significant difference between a human being and a mindless machine. In the case of the former we found that it was not necessary to postulate explicitly that it makes unpredictable choices; this conclusion imposed itself upon us and it follows from the fact that a human being is capable of having such mental properties as being desirous to achieve a certain aim and as being aware of a given argument as to how to secure this aim and then deciding to act in a way not conducive to achieving that aim.

But we can go even further than this. Let our predictor be a person to whom physical indeterminacy is no obstacle for he is capable of directly perceiving future events. In the case of a human player we know that even such a person cannot reliably predict what the choice is going to be because of the argument we have already repeated several times in this paper. But there is no reason why he should not be able to predict the choices of any machine, including a δ-machine. In maintaining this we are not led to any contradiction. For suppose the randomizer causes δ to choose box II only. The predictor foresees this and puts \$M in the box. Suppose we ask whether it would gain more by taking both boxes. The answer from the perfect well-wisher is, yes. But it cannot therefore proceed and take both boxes since it is not a γ-machine and has to make the choice forced upon it by the randomizer. But is there no contradictory answer based on inductive evidence, namely that every δ-machine which took both boxes found the second box empty? The answer is, no, there isn't. No δ-machine, the random process of which had the same outcome as our present δ-machine, ever took both boxes.[3]

[3] Suppose it is asked, what about a γ/δ machine i.e. a machine which part of the time makes the choice that maximizes its gain and part of the time makes random choices? After all when such a machine is in γ-state and chooses both boxes we can ask what would have happened if it took box II only and expect an answer since the same machine could have chosen box II only. But of course it would have to be in a different state to take box II only i.e. in a δ-state and the randomizer determining that its choice be box II only, in which case however the predictor would have put money in box II. Thus had the machine been in a state in which it was determined to take box II only it would have ended up with \$M but even in that case it would have been better off if it took both boxes and gain \$M + \$1000 except of course that it cannot do so not being in γ-state and its choice being determined by the randomizer. Thus by saying that all the choices of a γ/δ machine are predictable we are not driven to give contradictory answers to any questions.

Thus all the relevant types of machines α, β, γ, and δ are deprived of the kind of freedom enjoyed by a human being. Until someone comes along and describes a machine for which he can show that its choices are in principle immune to predictions made by any method we are entitled to maintain that there is an important difference between people and mindless machines concerning predictability in principle. I do not believe I need to expatiate upon the significance of this conclusion, should it be indeed correct. It is also clear that the conclusion is not only important in itself but it opens up all sorts of fascinating avenues for further research. For example there is a perplexing problem of how to tell apart a mindless machine which obeyed all the stimulus-response laws humans obeyed from a mind possessor. Some philosophers have been driven to the position of maintaining that since the two behave exactly alike it is meaningless to maintain that there is any difference between them. Others find this entirely unsatisfactory and hold that though there is absolutely no way of distinguishing between the two there is a fundamental ontological difference as well as a moral one concerning the way we ought to treat them. But we may have provided a way out of this difficulty. For if a predictor should have consistent success in predicting the choices of a certain type of player, would this not amount to strong empirical evidence that the player did not possess a mind? Thus empirical evidence might in principle be available showing that a mindless machine which stimulated human behaviour most perfectly was nevertheless a mindless machine. Alternatively, if a predictor who has proven himself an accomplished precognitor in a great variety of fields yet turned out to be entirely unsuccessful in predicting correctly the choices of a certain kind of a player, should this not be taken as firm empirical indication that the player possessed a mind? But these topics will have to be treated elsewhere.

The University of North Carolina

Received February 7, 1975