

# I. ACTIONS, PREDICTIONS, AND BOOKS OF LIFE

ALVIN I. GOLDMAN

ARE actions determined? Since it is difficult to tell "directly" whether or not actions are governed by universal laws, some philosophers resort to the following "indirect" argument:

If actions are determined, then it is possible in principle to predict them (with certainty).

It is not possible in principle for actions to be predicted (with certainty).

Therefore, actions are not determined.

A defender of this argument I shall call an "anti-predictionist"; his position will be called "anti-predictionism." In this paper I shall try to rebut anti-predictionism.

Both premisses of the anti-predictionist argument will come under attack here. The first premiss, affirming that determinism entails predictability, is often accepted without adequate scrutiny. Some writers not only assume that determinism entails predictability but even *define* determinism as the thesis that every event is predictable in principle.<sup>1</sup> I believe, however, that it is essential to distinguish between determinism and predictability. We must first notice that there are various kinds or senses of "possibility" which may be involved in the "possibility of prediction." Moreover, it can be shown that in many of these senses, determinism does *not* entail the possibility of prediction. Many anti-predictionists have failed to notice this, however. Therefore, upon discovering some unpredictability in the arena of human action, they have wrongly concluded that actions must be undetermined. This error will be avoided only if we carefully distinguish between determinism and predictability. Hence, an important aim of this paper will be to differentiate various senses of "possibility of prediction" and to ascertain how they are related to determinism.

Let us assume now that we can find some suitable sense of "possibility of prediction" which is closely related to, if not entailed by, determinism. The second premiss of the anti-predictionist argument

asserts that, in such a sense, it is impossible for actions to be predicted. Various arguments have been offered in support of this premiss. One that I shall consider concerns the possibility of writing a complete description of an agent's life—including his voluntary actions—even before he is born. According to anti-predictionism, if actions were determined, it would be possible to write such books. Indeed, it would be possible for such a "book of life" to be written even if the agent were to read its prediction of a given action before he is to perform that action. It seems clear to the anti-predictionist, however, that such books of life are impossible. Predictions of my actions cannot be made with certainty; for when I read these predictions, I can easily choose to falsify them. So argues the anti-predictionist. But it is far from clear that he is right. I think, on the contrary, that it may well be possible (in a suitable sense) for books of life to be written. And thus it seems to me that the anti-predictionist is unable to establish the truth of his second premiss.

In general, anti-predictionists support their second premiss by contrasting the predictability of human behavior with that of physical events. It is alleged that special difficulties of a purely conceptual sort arise for the prediction of action and that these difficulties are unparalleled in the realm of merely physical phenomena. I shall claim, however, that there are no essential differences between actions and physical events with respect to the problem of prediction. More precisely, I shall claim that *conceptual* reflection on the nature of human behavior (as opposed to investigation by the special sciences) does not reveal any peculiar immunity to prediction.

It must be emphasized that I offer no proof of the thesis that actions *are* determined; I merely wish to show that the anti-predictionist's arguments fail to prove that they are *not* determined. It is conceivable, of course, that actions are not determined. And if actions are not determined, then I would admit that they are not perfectly

<sup>1</sup> Karl Popper, for example, defines "determined" as "predictable in accordance with the methods of science," in "Indeterminism in Quantum Mechanics and in Classical Physics," *The British Journal for the Philosophy of Science*, vol. I (1950-51), see p. 120.

predictable (in any sense at all). What I contend, however, is that the arguments of philosophers, based on familiar, common-sense features of human action and human choice, do not prove that actions are undetermined or unpredictable. The basic features of human action are quite compatible with the contention that actions are determined and susceptible of prediction. In other words, my aim here is not to establish the *truth*, but merely the *tenability*, of the thesis that actions are determined.

## II

Let us begin with some definitions. I shall define determinism as the view that every event and state of affairs is determined in every detail. An event is determined (in a given detail) if and only if it is deducible from some set of antecedent conditions and laws of nature. A law of nature is, roughly, any true non-analytic universal statement of unlimited scope which supports counterfactual conditionals.<sup>2</sup> Both "low-level" empirical connections, like all metals expand when heated, and "theoretical" connections, like  $F = ma$ , are included. Antecedent conditions can be either events, like moving at 10 m.p.h., or states of affairs, like having a specific gravity of 1.7. (Throughout I shall be concerned both with events and with states of affairs, but for brevity I shall often omit reference to states of affairs.) Negations of events, like a ball's *not* moving at 10 m.p.h., are also included. Antecedent conditions may be directly observable phenomena, but they need not be. Theoretical, hypothetical, and dispositional states—like being brittle or being intelligent—can serve as antecedent conditions.

Notice that my definition of determinism is in terms of a formal relationship, i.e., the relationship of deducibility holding between events and sets of laws and antecedent conditions. In particular, this definition makes no explicit reference to the ability of anyone to predict these events, and thereby leaves open the question of the connection between determinism and predictability.

If determinism is true, human actions are determined. But determinism alone does not tell us what laws or kinds of laws take human actions as their dependent variables. I shall assume, however,

that these laws would include ones with psychological states like desires, beliefs, intentions, etc., as their independent variables. This presupposes—correctly, I think—that statements connecting actions with, for example, wants and beliefs, are not purely analytic.<sup>3</sup> Rather, their logical status would correspond to quasi-analytic, quasi-empirical generalizations like many theoretical statements of science. If determinism is true, wants, beliefs, intentions, etc., are themselves determined by prior events of various sorts. The determinants of these mental states are quite diverse, however, so I shall make no attempt to delineate them.

In ordinary language, not all determining factors of an event are called its "causes." A body's having a certain mass may be a determining antecedent condition of that body's moving at a certain velocity after being struck by another object, but its having that mass would not be called a "cause" of its velocity. Similarly, although a person's having a certain intention or desire would not ordinarily be termed a "cause" of his action, it may be an antecedent condition of the relevant sort. Since determinism is often connected with what philosophers call "causal necessity," I shall use the technical term "*causally necessitate*" to apply to antecedent conditions which, together with laws of nature, determine a given event. Thus, I shall say that desires and beliefs (together with other conditions) "*causally necessitate*" a given action, even though ordinary language would not condone such an expression.

In our discussion of predictability we need a sense of "prediction" distinct from mere lucky guesses or pre-cognition. We must be concerned with predictions made on the basis of laws and antecedent conditions. I shall call a prediction a "*scientific prediction*" if and only if it is made by deducing the predicted event from laws and antecedent conditions. A scientific predictor may learn of the laws and antecedent conditions in any number of ways. (On my definition, most predictions made by actual scientists are not "scientific" predictions, for real scientists seldom, if ever, *deduce* what will occur from laws and prior conditions. Nevertheless, scientific prediction as defined here may be regarded as an ideal of prediction to which scientists can aspire.)

<sup>2</sup> There are, of course, numerous problems associated with the concept of a law of nature. But a detailed discussion of these problems would go beyond the scope of this paper.

<sup>3</sup> For a defense of this view, see William P. Alston, "Wants, Actions, and Causal Explanations" in H. N. Castañeda, (ed), *Minds, Intentionality, and Perception* (Detroit, 1967) and R. Brandt and J. Kim, "Wants as Explanations of Actions," *The Journal of Philosophy*, vol. 60 (1963), pp. 425-435.

As  
differ  
tion."  
of pos  
made  
four s  
possib  
compo.

An  
self-co  
only i  
circle  
go fe  
sibility  
two c  
only i  
logica  
(i.e.,  
of the  
juncti  
event.

12 o'  
12 o'

An  
not in  
an ex  
there  
sistent  
exami  
only c  
but a  
kinds  
weigh  
but it  
being  
consti  
it imp

*Cau*  
bility  
events  
causal  
allow  
to occ  
of ev  
possib  
jointly  
a cau.

<sup>4</sup> Th  
on the  
laws.

For  
distinct  
be caus

As indicated above, it is important to identify different senses of the phrase "possibility of prediction." I shall now distinguish four relevant species of possibility, though further distinctions will be made later within some of these categories. The four species are: (1) *logical possibility*, (2) *logical compossibility*, (3) *physical possibility*, and (4) *causal compossibility*.

An event is *logically possible* if and only if it is not self-contradictory, and *logically impossible* if and only if it is self-contradictory. Drawing a square circle is a logically impossible event, while jumping 90 feet is a logically possible event. *Logical compossibility* is defined for two or more events. A set of two or more events is logically compossible if and only if the conjunction of the members of the set is logically consistent. A set is logically impossible (i.e., not logically compossible) if and only if each of the events is logically possible but their conjunction is logically inconsistent. Thus, the two events, (a) *x*'s being a pumpkin from 11 o'clock to 12 o'clock, and (b) *x*'s turning into a pumpkin at 12 o'clock, are logically impossible.

An event is *physically possible* if and only if it is not inconsistent with any law or laws of nature; an event is *physically impossible* if and only if there are laws of nature with which it is inconsistent. Traveling faster than the speed of light, for example, is physically impossible. I shall speak not only of events being physically impossible *in general*, but also as being physically impossible *for* certain kinds of entities. Thus, the act of lifting a ten-ton weight is not, in general, physically impossible; but it is physically impossible *for* (normal) human beings to lift ten-ton weights. Given the physical constitution of human beings, laws of nature make it impossible for them to lift such weights.

*Causal compossibility* differs from physical possibility in attending to groups of events rather than events taken singly. Roughly, a set of events is causally compossible just in case laws of nature allow each of them to occur singly and allow them to occur as a group. More precisely, consider a set of events  $\{e_1, \dots, e_n\}$  each of which is logically possible and physically possible, and which are jointly logically compossible. Then  $\{e_1, \dots, e_n\}$  is a causally compossible set if and only if there is

no set of laws of nature such that the conjunction of these laws with  $e_1, \dots, e_n$  is logically inconsistent.<sup>4</sup> I shall say that the set as a whole is causally compossible or that each member is causally compossible "with" or "relative to" the other members.

A set of events  $\{e_1, \dots, e_n\}$  is causally *incompossible* (i.e., not causally compossible) if and only if there are some laws of nature  $L_1, \dots, L_k$  such that the conjunction of  $L_1, \dots, L_k$  with  $e_1, \dots, e_n$  is logically inconsistent. Assuming, as we do, that  $e_1, \dots, e_n$  satisfy the other three species of possibility, the set  $\{e_1, \dots, e_n\}$  will be causally incompossible if and only if the negation of (at least) one member of the set is entailed by the conjunction of the other members of the set conjoined with  $L_1, \dots, L_k$ . Thus, if the negation of a given member of the set is causally necessitated by the other members of the set, then the set is causally incompossible.

### III

The most interesting questions concerning the prediction of action are best handled in terms of the notion of causal compossibility. The reflexivity of predictions—the fact that a prediction often has an effect which bears on its own truth—can be understood properly with the use of this notion. But the question of the causal compossibility of predictions of action cannot arise unless the other three species of possibility are satisfied. Our definition of causal compossibility makes a set causally compossible only if its members are logically possible, physically possible, and (jointly) logically compossible. For example, if it is physically impossible to make scientific predictions of actions, the question of causal compossibility does not even arise. Therefore, before turning to the questions of reflexivity, including the question of whether "books of life" can be written, we must focus on certain problems connected with the logical compossibility and the physical possibility of predicting actions.

The logical possibility and compossibility of predictions can be discussed together, since the distinction between them is somewhat blurred. This is because a correct prediction is not really a single event, but a pair of events—a prediction and an

<sup>4</sup> The term "event" is here used to designate event *kinds*, not necessarily ones that have been actualized. The term "law," on the other hand, will be used only to designate actual laws, i.e., laws that obtain in the real world, and not merely possible laws.

For the most part, I shall consider events with built-in time references. Sam's jumping rope at 10:35 will be treated as a distinct event from Sam's jumping rope at 10:45. This is very natural in the present context, since a given set of events may be causally impossible with Sam's jumping rope at 10:35 but causally compossible with Sam's jumping rope at 10:45.

event predicted. Two different examples of logical impossibility have been uncovered in connection with the prediction of behavior. I shall discuss these examples briefly and argue that, contrary to what their authors suppose, they do not prove that actions are undetermined and they do not prove that actions have a peculiar immunity to prediction unparalleled by physical phenomena.

The first logical impossibility, as discussed by Maurice Cranston,<sup>5</sup> can be summarized as follows. Suppose that Sam invents the corkscrew at time *t*. In the intended sense of "invent," this means (a) that Sam thinks of the corkscrew at *t*, and (b) that no one ever thought of the corkscrew before *t*. Cranston argues that no one could have predicted Sam's inventing the corkscrew. In order for him to make this prediction, he would himself have to think of the corkscrew. And had he thought of the corkscrew, it would be false to say that Sam "invented" the corkscrew. Yet, *ex-hypothesi*, Sam *did* invent the corkscrew. Using the terminology of "logical impossibility," we can formulate Cranston's problem by saying that the three events, (a) Sam thinks of the corkscrew at *t*, (b) no one ever thought of the corkscrew before *t*, and (c) someone predicted Sam's inventing the corkscrew, are logically impossible.

The second example poses a problem for predicting not actions, but decisions. However, since the concept of a voluntary action is so closely tied to that of a decision, an unpredictability connected with decisions is very important for us to discuss. Carl Ginet claims that it is impossible ("conceptually" impossible) for anyone to predict his own decisions.<sup>6</sup> The argument begins by defining "deciding to do *A*" as *passing into* a state of knowledge (of a certain kind) that one will do *A*, or try to do *A*.<sup>7</sup> Suppose now that Sam, at *t*, decides to do *A*. Had Sam predicted that he would make this decision—and had this prediction involved *knowledge*—he could not have decided later to do *A*. For if, before *t*, he had known that he would decide to do *A*, he would have known then that he would do *A*, or try to do *A*. But if, before *t*, he had known that he would do *A* (or try to do *A*), then he

could not, at *t*, have *passed into* a state of knowing that he would do *A*. Thus, according to Ginet, Sam could not have predicted that he would make this decision.

Of course, Sam might make his prediction and then forget it. If so, he can still decide, at *t*, to do *A*. However, if Sam not only knows, before *t*, that he will decide to do *A*, but also *continues* to know this up until *t*, then Sam cannot, at *t*, decide to do *A*. In other words, the following three events are logically impossible: (a') Sam decides, at *t*, to do *A*, (b') Sam predicts (i.e., knows) that he will decide to do *A*, and (c') Sam continues to know this until *t*.

What do these two logical impossibilities prove? Do they prove that decisions and inventions are undetermined? Do they prove that voluntary actions, including the decisions which lead to them, have a special immunity to prediction? The answer is "No," I believe, to both questions.

Our examples of logical impossibilities do not establish any special status for human behavior, for precisely analogous impossibilities can be produced for physical phenomena. Let the expression "a tornado strikes *x* by surprise" mean: (1) a tornado strikes *x* at a certain time, and (2) before that time nobody ever thought of a tornado striking *x*. Now suppose that, as a matter of fact, a tornado strikes Peking by surprise. Then it is logically impossible for this event to have been predicted. That is, the set consisting in the tornado striking Peking by surprise and a prediction of the tornado striking Peking by surprise is a logically impossible set. In general it is logically impossible for tornadoes striking places by surprise to be predicted. For if anyone were to predict these events, they could no longer be described as "tornadoes striking places by surprise." Nevertheless, there certainly are (or could be) events correctly describable as "a tornado striking *x* by surprise."

I wish next to argue that the invention and decision impossibilities do not show that these human phenomena are undetermined. Notice first that the tornado case, though it has the same

<sup>5</sup> *Freedom: A New Analysis* (London, 1954), p. 169.

<sup>6</sup> "Can the Will Be Caused?," *The Philosophical Review*, vol. 71 (1962), pp. 49-55.

<sup>7</sup> One might challenge Ginet's argument by criticizing this definition of "deciding." This criticism has implicitly been made, along with other criticisms of Ginet's position, by various writers. For example, see John Canfield, "Knowing about Future Decisions," *Analysis*, vol. 22 (1962), and J. W. Roxbee Cox, "Can I Know Beforehand What I Am Going to Decide?," *The Philosophical Review*, vol. 72 (1963). Here I shall waive these criticisms, however, and accept Ginet's claim that it is impossible to predict one's own decisions. I shall then ask whether this proves that decisions are undetermined and whether they are intrinsically different from physical phenomena.

logical structure, does not bear on the question of determinism. Although it is logically impossible for anyone to predict the tornado striking Peking by surprise, I am in no way inclined to suppose that this event is not determined. Similarly, our logical impossibilities fail to show that inventions and decisions are undetermined. How could such logical impossibilities demonstrate that these events are not governed by laws of nature? The notion of a law is in no way involved in the concept of logical impossibility. And hence the presence of logical impossibilities sheds no light on the question of whether there are laws and antecedent conditions which entail inventions or decisions.

The critical error here is the assumption that if an event is determined (under a given description), it must be possible to predict it (under that description).<sup>8</sup> The falsity of this proposition should be adequately clear from the invention case. Suppose that Sam's thinking of the corkscrew at *t* is deducible from laws and antecedent conditions. And suppose that the fact that no one ever thinks of the corkscrew before *t* is also deducible from laws and antecedent conditions. Then, the event consisting in Sam's *inventing* the corkscrew at *t* would be determined; but it still would be logically impossible for it to have been predicted under that description. The lesson to be learned here is not that inventions are undetermined actions, but that the alleged entailment between determinism and predictability is not an entailment at all. At any rate, the fact that an event is determined under a given description does not entail that it is *logically compossible* for it to be predicted under that description.<sup>9</sup>

The case of decisions can be handled similarly. It seems to me quite possible that a person's passing into a state of knowing, or intending, to do *A* be deducible from laws and antecedent conditions. But although this event would be determined (under the given description) it would not be logically compossible for Sam to have predicted it (under that description) and continued to know it until *t*.

<sup>8</sup> That this is an error has also been claimed by Arnold S. Kaufman, in "Practical Decision," *Mind*, vol. 75 (1966), see p. 29.

<sup>9</sup> It is also an error—committed at least as frequently—to think that determinism entails the possibility of retrodicting or explaining every event under any description. Suppose that Sam thinks of the corkscrew at *t* and that no one ever thinks of the corkscrew after *t*. Suppose, moreover, that both of these events are deducible from laws and antecedent conditions. Now let us introduce the expression "postventing *x*" to mean "thinking of *x* for the last time" (just as "inventing *x*" means "thinking of *x* for the first time"). Clearly, we may say of Sam that he "postvented" the corkscrew and that this action of his is determined. However, it is logically impossible for anyone to *retrodict* Sam's postventing the corkscrew. To do so, the retrodictor would himself have to think of the corkscrew, and, *ex hypothesi*, Sam thought of the corkscrew for the last time at *t*.

<sup>10</sup> *Op. cit.*

## IV

I turn now to physical possibility. Is it physically possible to make scientific predictions of human actions? Here the emphasis should be placed on the qualifier "scientific." Although it may well be physically possible to make "lucky guess" predictions, or perhaps even predictions based on "intuition," it is not obvious that predictions can be made by deducing an action from laws and antecedent conditions. And this is the only kind of prediction which bears on the issue of determinism.

Anti-predictionists might claim that it is physically impossible for human beings to make scientific predictions of actions, because human beings cannot learn enough antecedent conditions to deduce what will be done. But it is inessential to the predictionist's position to restrict the range of predictors to human beings. In order to avoid theological or supernatural issues, we may require that any predictor be a finite entity operating within the causal order of the universe. But apart from this, no arbitrary limits should be placed on admissible predictors.

Karl Popper<sup>10</sup> has tried to show that there are certain limitations of the predictions which can be made by "classical mechanical calculating" machines. But to restrict the range of predictors to calculating machines is an important restriction; even if Popper is right about the prediction-limitations of machines of the sort he discusses, there may be other beings that can make predictions his machines cannot. Another limitation on Popper's discussion is that much of it is aimed at establishing the physical impossibility of a *single* being, like Laplace's demon, making scientific predictions of *all* events or of a very large number of events. But the fact that all events cannot be predicted by a single being is compatible with the proposition that every event can be predicted by some being or other.

Anti-predictionists might proffer the following arguments for saying that it is physically impossible for *any* finite being, not just human beings, to make

scientific predictions of human behavior. Scientific predictions, they might claim, require knowledge of infinitely many facts, but it is physically impossible for a finite being to know infinitely many facts. The infinity requirement seems necessary because in order to *deduce* that even a certain finite system will yield a given result, one must know that no interfering factors will intrude from outside the system. And knowing this may involve knowing *all* states of the world at least at one time.

This argument is of questionable force. It is far from clear that the deduction of actions from antecedent conditions and laws requires knowledge of infinitely many facts. Nor is it clear that no finite being could know infinitely many facts. Even if the argument is correct, however, it would seem to prove *too much*. For if the knowledge of infinitely many facts is required in order to make scientific predictions of actions, the same would be true for scientific predictions of physical events. Thus, the above argument would fail to establish any special immunity of human action to prediction. Finally, even if it is physically impossible for any finite being to make scientific predictions of actions, this would not prove that actions are undetermined. Here too, as above, we have a sense of "possibility" in which determinism does *not* entail the possibility of prediction. The proposition that an event is (formally) deducible from laws and antecedent conditions does not entail that it is physically possible for any being to come to know these laws and antecedent conditions and to deduce the event from them. Hence, even if the anti-predictionist could establish that it is physically impossible to predict actions scientifically, he would not thereby establish that actions are undetermined.

We have not conclusively shown either that it is physically possible for some beings to predict actions scientifically or that it is not. But unless we assume that this is physically possible, we cannot turn to the other interesting issues that surround the problem of the prediction of human behavior. Unless we assume this, the question of the causal compossibility of predicting actions cannot even arise. In order to explore these important issues, therefore, I shall henceforth assume that scientific predictions of actions (like scientific predictions generally) are physically possible.

## V

Perhaps the anti-predictionist would think it obvious that it is causally impossible to predict

actions scientifically. He might argue as follows: "Let us grant, as is likely, that there have never been any genuine scientific predictions of voluntary human actions. If, as my opponent claims, determinism is true, then it is causally impossible for any predictions to have been made of these actions. For every actual action  $A$ , there is an actual event  $\bar{P}_A$ , the *absence* of a prediction of  $A$ . Since each of these events  $\bar{P}_A$  is actual, and since determinism is true, each of these events  $\bar{P}_A$  must be causally necessitated by some set of actual events prior to it. But if each of these events  $\bar{P}_A$  is causally necessitated by actual prior events, then each event  $\bar{P}_A$ —the prediction of  $A$ —is causally impossible relative to some actual events. In other words, for each actual action  $A$ , it is causally impossible for  $A$  to have been predicted."

This argument, like a previous one, proves too much. The anti-predictionist is right in saying that non-actual predictions of actions are causally impossible with the actual prior events in the world. But this is true simply because, assuming determinism, every non-actual event whatever is causally impossible with some set of actual prior events. Thus, using the notion of causal-compossibility-relative-to-all-actual-events, we can establish the impossibility of predicting physical phenomena as well as human behavior. We can point to an action that was never predicted and say that, in this sense, it "could not" have been predicted, since its non-prediction was causally necessitated by other actual events. But by the same token, we can point to a physical event which was never predicted and say that it "could not" have been predicted, since its non-prediction was also causally necessitated by other actual events. Using this notion of "possibility of prediction," the anti-predictionist again fails to establish any special immunity of action to prediction.

Apart from this point, however, the notion of "causal-compossibility-relative-to-all-actual-events" does not seem to be a pertinent kind of possibility for our discussion. We have seen that determinism does not entail the possibility of predicting actions in *every* sense of "possible." And here, I believe, we have still another sense of "possible" in which determinism does not entail that it is possible for every action to be predicted. Determinism does not say that, relative to all actual prior events, it is causally compossible for a prediction of an action to be made *even if* those actual prior events causally necessitate that no prediction occur. Thus, the fact that it is impossible, in this sense, for

action:  
the t  
surpr  
discu  
"cau  
even  
only  
and  
ing t  
W  
one  
taki  
We  
ing  
we s  
Sup  
coul  
wou  
clau  
coul  
Nov  
fron  
to s  
had  
subs  
ever  
in t  
wou  
case  
or v  
wor  
way  
A  
pur  
we:  
"Co  
The  
eve:  
it d  
ima  
arg  
tha  
eve:  
an  
succ  
it n  
hav  
11  
even  
cha:  
will  
12  
that

actions to be predicted does not conflict with the thesis that actions are determined. Nor is it surprising that the sense of "possible" here under discussion is not important. Using the notion of "causally-compossible-relative-to-all-actual-prior-events" it turns out, assuming determinism, that only actual events are possible. But it is a strange and unduly restrictive notion of "possible" according to which only actual events are possible!

We need, then, a broader notion of possibility, one which allows for non-actual possibles while also taking into account the notion of causal necessity. We can discover a more relevant notion by examining what is often meant in ordinary contexts when we say, counterfactually, "*e* could have occurred." Suppose we say, counterfactually, "The picnic could have been a success." This sort of statement would normally be made with a suppressed "if"-clause. We might mean, for example, "The picnic could have been a success if it had not rained." Now if the only thing which prevented the picnic from being a success was the rain, we are also likely to say, "The picnic *would* have been a success if it had not rained." In the first case we mean that the substitution of non-rain for rain in the course of events would have *allowed* the picnic to be a success; in the second case we mean that this substitution would have *ensured* the success of the picnic. In both cases we are saying that a certain event could have or would have occurred *if* the prior course of the world had differed from its actual course in specified ways.

Although in ordinary contexts we might not pursue the matter further, in order to be systematic we must inquire further: "Could it *not* have rained?" "Could non-rain have occurred instead of rain?" The actual rain was causally necessitated by actual events prior to the rain. If we are to suppose that it did not rain, we must also make changes (in our imagination) of still earlier events. Carrying this argument to its logical conclusion, it is obvious that whenever a determinist says that a non-actual event *e* "could have" occurred, he must imagine *an entirely new world*. For the picnic to have been a success, it is required that it not have rained. For it not to have rained, the cloud formation would have had to be different. For the cloud formation

to have been different, it is required that the wind velocity (or some other factor) have been different. Etc.

Not only must we change conditions prior to *e*, if we are to suppose *e* occurs, but we probably<sup>11</sup> must change events after *e* as well. Had it not rained, a certain other picnic group near us would not have ended their picnic just then. And had they not ended their picnic just then, they would not have left for home just then. And had they not left for home just then, they would not have had an automobile accident when they did.<sup>12</sup> Etc.

The determinist who says, counterfactually, "*e* could have occurred," must construct a whole world to justify his claim. Nevertheless, this gives him a sense of "possible" that allows non-actual possibles. For a determinist, "*e* could have occurred" may be translated as "a causally compossible world can be imagined in which *e* occurs." Normally the determinist will be able to construct worlds resembling the real one to a large extent. But these worlds will never be exactly like our world except for one event only. Any such imagined world will differ from the real world by at least one event for every moment of time. This will be true, at any rate, if the laws governing these imagined worlds are identical with those of the real world. And I shall assume throughout that these laws (whatever they are, exactly) are held constant.

## VI

We can now give what I regard as a reasonable formulation of the question: "Is it possible, in principle, to make scientific predictions of voluntary actions?" The formulation is: "Can one construct causally compossible worlds in which scientific predictions are made of voluntary actions?" In saying that this is a "reasonable" formulation of the question, I do not mean that a negative answer to this question would entail that voluntary actions are not determined. I have already pointed out that determinism does not entail that it is physically possible to make scientific predictions of events, including actions. Hence, neither does determinism entail that there are causally compossible worlds in which scientific predictions of

<sup>11</sup> I say "probably" because the definition of determinism does not entail that every event is a determinant of some subsequent event. Thus, if not-*e* actually occurred but had no effect on any subsequent event, then we might substitute *e* for not-*e* without changing any subsequent events. However, though determinism does not require it, it is reasonable to assume that every event will have some differential effect on *some* later event or events.

<sup>12</sup> This is all plausible, at any rate, if we deny fatalism. Fatalism, which is by no means implied by determinism, is the view that certain events will happen at certain times *no matter what* antecedent conditions obtain.

actions occur. However, since we are assuming that scientific predictions are physically possible, it would be an important negative result to discover that one cannot construct causally compossible worlds in which scientific predictions are made of voluntary actions. This might not prove that actions are undetermined, but it would suggest a disparity between actions and physical phenomena. For, assuming that scientific predictions are physically possible, it does seem that there are causally compossible worlds in which scientific predictions are made of physical events.

Similar comments are in order on the question, "Can one construct causally compossible worlds in which scientific predictions are made of voluntary actions and in which the agent learns beforehand of the prediction?" Determinism does not entail that there must be such causally compossible worlds. But if no such worlds are constructible—worlds in which "books of life" are found, or things comparable to books of life—one might well claim a disparity between voluntary actions and physical phenomena.

Fortunately, I believe that there *are* causally compossible worlds in which scientific predictions are made of voluntary actions and in which, moreover, the agent learns of (some of these) predictions before he performs the predicted actions. I believe that there are causally compossible worlds in which books of life are written before a man's birth. Inscribed in these books are predictions of the agent's actions, predictions based on laws and antecedent conditions. These predictions are correct even though the agent sometimes reads them before he performs the predicted actions. I shall support my claim that there are such causally compossible worlds by giving a sketch of such a world. Before giving my sketch, however, I wish to examine the structure of prediction-making where the prediction itself has a causal effect on the predicted event. This will be essential in understanding how a "book of life" could be written, even though the writer knows that the agent will read it.

Consider the problem of an election predictor. He may know what the precise results of the upcoming election are going to be, if he makes no public prediction of the election. If he publishes a prediction, however, some of the voters, having found

out what the results will be, may change their votes and thereby falsify his prediction. How, then, can a pollster make a genuinely scientific and accurate prediction of an election? Can he take into account the effect of the prediction itself? Herbert Simon has shown that, under specifiable conditions, a predictor can do this.<sup>13</sup> Essentially, what the predictor must know is the propensity of the voters in the community to *change* their voting intention in accordance with their expectations of the outcome. If persons are more likely to vote for a candidate when they expect him to win than when they expect him to lose, we have a "bandwagon" effect; if the opposite holds, we have an "underdog" effect.

Let us suppose that a given pollster has ascertained that, two days before the election, 60 percent of the electorate plans to vote for candidate *A* and 40 percent for *B*. He also knows that, unless he publishes a prediction, the percentages will be the same on election day. Further suppose he knows that there is a certain "bandwagon" effect obtaining in the voting community.<sup>14</sup> When the original intention of the electorate is to vote 60 percent for *A*, this bandwagon effect can be expressed by the equation,  $V = 60 + .2(P - 50)$ , where *P* is the percentage vote for *A* publicly predicted by a pollster, and *V* is the actual resultant vote for *A*. Clearly, if the pollster publicly predicts that *A* will receive 60 percent of the vote, his prediction will be falsified. Putting  $P = 60$ , the equation tells us that  $V = 62$ . In other words, the effect of the prediction, combined with the original voting intention of the electorate, would result in a 62 percent vote for *A*. However, the pollster can easily calculate a value for *P* which will make  $P = V$ . He need only solve the two equations,  $P = V$  and  $V = 60 + .2(P - 50)$ . Such a solution yields  $P = 62.5$ . Thus, the pollster can publish a prediction saying that 62.5 percent of the electorate will vote for *A*, knowing that his own prediction will bring an additional 2.5 percent of the electorate into the *A* column, and thereby make his prediction come true.

Notice that all the antecedent conditions relevant to the outcome cannot be known until it is known what prediction (if any) the pollster will make. His prediction (or lack of prediction) is itself an im-

<sup>13</sup> "Bandwagon and Underdog Effects of Election Predictions," reprinted in *Models of Man* (New York, 1957). The requisite condition is that the function relating the actual outcome of the voting to the predicted outcome, given the electorate's original voting intention, be *continuous*.

<sup>14</sup> That this bandwagon effect holds in the community could be discovered either by studying previous elections or by deducing it from "higher-level" generalizations found to be true of the community.



portant antecedent condition. However, one of the crucial determinants of the outcome—viz., the original voting intention of the electorate—is given independently of the pollster's prediction. Thus, while holding that factor constant, the pollster calculates what the outcome of the election *would* be, *if* he were to make certain predictions. By solving the equations given above, he discovers a prediction which, if published, would be confirmed. He thereupon forms an intention to publish that prediction and proceeds to fulfill that intention. Until he forms this intention, he does not know what prediction he will make, and therefore does not know all the requisite antecedent conditions from which to deduce the election outcome. But at the same time he makes the prediction (and perhaps even earlier), he does know all the relevant antecedent conditions and has deduced from these conditions what the results will be. Thus, his prediction of the outcome is a truly scientific prediction.

If someone wishes to predict a single person's behavior and yet let him learn of the prediction, the predictor must employ the same sort of strategy as the pollster. He must take into account what the agent's reaction will be to the prediction. There are several kinds of circumstances in which, having made the appropriate calculations, he will be able to make a correct prediction. (A) The agent learns of the prediction but does not want to falsify it. (B) Upon hearing the prediction, the agent decides to falsify it. But later, when the time of the action approaches, he acquires preponderant reasons for doing what was predicted after all. (C) Having decided to refute the prediction, the agent performs the action conforming with it because he doesn't realize that he is conforming with it. (D) At the time of the action the agent lacks either the ability or the opportunity to do anything but conform with the prediction, though he may have believed that he would be able to falsify it. In any of these four kinds of cases, a predictor would be able to calculate that his prediction, together with numerous other antecedent conditions, would causally necessitate that the agent perform the predicted action. In a case of kind (B), for example, the predictor may be able to foresee that the agent will first read his prediction and decide to falsify it. But other factors will crop up—ones which the agent did not originally count on—which will make him change his mind and perform the predicted action after all. And the predictor also foresees this.

In the first three kinds of cases, (A), (B), and (C),

the agent performs the predicted action *voluntarily* (though in (C) he does not realize that what he is doing falls under the description "what was predicted"). In other words, in each of these three kinds of cases, the agent *could have* acted otherwise, in at least one sense of "could have" which some philosophers think is relevant to free will. Thus, the possibility of a scientific prediction does not require that the agent be *unable* to act in any way different from the prediction. All that is required is that the agent will not *in fact* act in any way different from the prediction. A predictor might know that an agent will in fact act in a certain way, not because he knows the agent will be incapable of doing otherwise, but because he knows that the agent will *choose* or *decide* to act as predicted. This point will be clarified at the end of the paper in a brief discussion of the indicated sense of "could have."

I shall now give a sketch of a causally compossible world in which a large number of correct predictions are made of an agent's behavior. Since I imagine this world to be governed by the same laws as those of the real world, and since I do not know all the laws of the real world, I cannot *prove* that my imagined world really is causally compossible. But as far as I can tell from common-sense knowledge of psychological and physical regularities, it certainly seems to be causally compossible. In this world, predictions of a man's life are made in great detail and inscribed in a "book of life," (parts of) which the agent subsequently reads. Obviously, I cannot describe the whole of this world, but I shall describe some of its most important and problematic features, namely the interaction between the agent and the book. Unfortunately, I shall have to omit a description of another important part of the world, the part in which the predictor (or predictors) gathers his data and makes his calculations. I am unable to describe this part of the world, first, because I do not know all the laws which the predictor would have at his disposal, and secondly, because I am not able to say just what the structure of this being would be. However, the main features of his *modus operandi* should be clear from our discussion of the pollster, whose technique is at the heart of such predicting.

## VII

And now to the description of the world.

While browsing around the library one day, I

noticed an old dusty tome, quite large, entitled "Alvin I. Goldman." I take it from the shelf and start reading. In great detail, it describes my life as a little boy. It always gibes with my memory and sometimes even revives my memory of forgotten events. I realize that this purports to be a book of my life and I resolve to test it. Turning to the section with today's date on it, I find the following entry for 2:36 p.m. "He discovers me on the shelf. He takes me down and starts reading me. . . ." I look at the clock and see that it is 3:03. It is quite plausible, I say to myself, that I found the book about half an hour ago. I turn now to the entry for 3:03. It reads: "He is reading me. He is reading me. He is reading me." I continue looking at the book in this place, meanwhile thinking how remarkable the book is. The entry reads: "He continues to look at me, meanwhile thinking how remarkable I am."

I decide to defeat the book by looking at a future entry. I turn to an entry 18 minutes hence. It says: "He is reading this sentence." Aha, I say to myself, all I need do is refrain from reading that sentence 18 minutes from now. I check the clock. To ensure that I won't read that sentence, I close the book. My mind wanders; the book has revived a buried memory and I reminisce about it. I decide to reread the book there and relive the experience. That's safe, I tell myself, because it is an earlier part of the book. I read that passage and become lost in reverie and rekindled emotion. Time passes. Suddenly I start. Oh yes, I intended to refute the book. But what was the time of the listed action?, I ask myself. It was 3:19, wasn't it? But it's 3:21 now, which means I have already refuted the book. Let me check and make sure. I inspect the book at the entry for 3:17. Hmm, that seems to be the wrong place for there it says I'm in a reverie. I skip a couple of pages and suddenly my eyes alight on the sentence: "He is reading this sentence." But it's an entry for 3:21, I notice! So I made a mistake. The action I had intended to refute was to occur at 3:21, not 3:19. I look at the clock, and it is still 3:21. I have not refuted the book after all.

I now turn to the entry for 3:28. It reads, "He is leaving the library, on his way to the President's office." Good heavens, I say to myself, I had completely forgotten about my appointment with the President of the University at 3:30. I suppose I could falsify the book by not going, but it is much more important for me not to be late for that appointment. I'll refute the book some other time!

Since I do have a few minutes, however, I turn back to the entry for 3:22. Sure enough, it says that my reading the 3:28 entry has reminded me about the appointment. Before putting the book back on the shelf, and leaving, I turn to an entry for tomorrow at 3:30 p.m. "He's still riding the bus bound for Chicago," it reads. Well, I say to myself, *that* prediction will be easy to refute. I have absolutely no intention of going to Chicago tomorrow.

Despite my decision to refute the book, events later induce me to change my mind and to conform to it. For although I want to refute the book on this matter, stronger reasons arise for not refuting it. When I get home that evening I find a note from my wife saying that her father (in Chicago) is ill and that she had to take the car and drive to Chicago. I call her there and she explains what has happened. I tell her about the book. Next morning she calls again with news that her father's condition is deteriorating and that I must come to Chicago immediately. As I hang up I realize that the book may turn out right after all, but the situation nevertheless demands that I go to Chicago. I might still refute it by going by plane or train. However, I call the airlines and am told that the fog is delaying all flights. The railroad says that there are no trains for Chicago till later in the day. So, acquiescing, I take a bus to Chicago, and find myself on it at 3:30.

### VIII

Let me interrupt my narrative here. I have given several cases in which the book is not refuted, and the reader should be convinced that I could easily continue this way. But it is important now to reply to several objections which the anti-predictionist is anxious to make against my procedure.

(1) "*Your story clearly presupposes determinism. But whether or not determinism is true is the central matter of dispute. Hence, you are begging the question.*" Admittedly, my story does presuppose determinism. Unless determinism were true, the imagined predictor could not have figured out what actions the agent would perform and then written them in the book. However, I do not think that this begs the question. For I am not here trying to prove that determinism is true. I am merely trying to show that the thesis of determinism is quite compatible with the world as we know it and with human nature as we know it. The world

depicted in my story seems to be very much like the real world except that it contains different antecedent conditions. The fact that this imagined world is determined and contains predictions of actions, and yet it resembles the real world very closely, suggests to me that the real world may also be determined. At any rate, this supposition seems quite tenable, and its tenability is what I seek to establish in this paper.

(2) "*The story you told was fixed. Events might have been different from the way you described them. For example, the fog might not have curtailed all air traffic.*" No, events could not be different in the world I am imagining. That is, in my world all the events I described were causally necessitated by prior antecedent conditions. I did not describe all the antecedent conditions, so perhaps the reader cannot see that each event I did describe was causally necessitated by them. But, since it is a deterministic world, that is so. No one can imagine my world and also substitute the negation of one of the events I described. I'm not "fixing" the story by saying that the fog curtailed air traffic; that just is the way my imagined world goes.

(3) "*But I can imagine a world in which some putative predictions of actions are refuted.*" I have no doubt that you can; that is very easy. You could even imagine a world somewhat like the one I have just described, but in which putative predictions are falsified. But this proves nothing at all. I would never deny that one can construct some causally compossible worlds in which putative scientific predictions of actions are not successful. I have only claimed that one can (also) construct some causally compossible worlds in which genuine scientific predictions of actions are made (and are successful). The situation with predictions of action is no different from the one with predictions of physical events. We can construct causally compossible worlds in which predictions of physical phenomena are correct. But we can also construct worlds in which putative scientific predictions of physical phenomena are incorrect. If our ability to construct worlds in which predictions are unsuccessful proves the inherent unpredictableness of the kind of phenomena unsuccessfully predicted, then we can prove the unpredictableness of physical phenomena as easily as the unpredictableness of human action.

(4) "*The world you have described, though possible, is a highly improbable world. Worlds in which putative predictions of actions are falsified are much more probable.*" The notion of one possible world being "more

probable" than another seems to me unintelligible. Surely the statistical sense of probability cannot be intended. There is no way of "sampling" from possible worlds to discover what features most of them have. Perhaps the anti-predictionist means that we can *imagine* more worlds in which putative predictions of actions are falsified. But this too is questionable. I can imagine indefinitely many worlds in which successful predictions of actions are made.

Perhaps the anti-predictionist means that it is improbable that any such sequence of events as I described would occur in the *real* world. He may well be right on this point. However, to talk about what is probable (in the evidential sense) in the real world is just to talk about what has happened, is happening, and will happen as a *matter of fact*. But the dispute between predictionists and anti-predictionists is, presumably, not about what *will* happen, but about what *could* happen in principle. This "in principle" goes beyond the particular facts of the actual world.

(5) "*The difference between physical phenomena and action is that predictions of actions can defeat themselves; but predictions of physical events cannot.*" This is not so. One can construct worlds in which the causal effect of a putative prediction of a physical event falsifies that prediction. Jones calculates the position of a speck of dust three inches from his nose and the direction and velocity of wind currents in the room. He then announces his prediction that five seconds thence the speck will be in a certain position. He had neglected to account for the wind expelled from his mouth when he made the prediction, however, and this factor changes the expected position of the speck of dust. Perhaps one can imagine a wider variety of cases in which predictions affect human action more than physical phenomena. But this is only a difference of *degree*, not of *kind*.

(6) "*Predictions of physical events can refute themselves because the predictor may fail to account for the effect of his own prediction. But were he to take this effect into account, he would make a correct prediction. On the other hand, there are conditions connected with the prediction of action in which, no matter what prediction the predictor makes, his prediction will be falsified. Here there is no question of inaccurate calculation or insufficient information. Whatever he predicts will be incorrect. Yet this situation arises only in connection with human action, not physical events.*"

This is an important objection and warrants detailed discussion.

## IX

Suppose that I wish to predict what action you will perform 30 seconds from now, but that I shall not try to change or affect your behavior except by making my prediction. (Thus, I shall not, for example, predict that you will perform no action at all and then make that prediction come true by killing you.) Further suppose that the following conditions obtain. At this moment you want to falsify any prediction that I shall make of your action. Moreover, you will still have this desire 30 seconds from now, and it will be stronger than any conflicting desire you will have at that time. Right now you intend to do action  $A$ , but you are prepared to perform  $\bar{A}$  (not- $A$ ) if I predict that you will perform  $A$ . Thirty seconds hence you will have the ability and opportunity to do  $A$  and the ability and opportunity to do  $\bar{A}$ . Finally, conditions are such that, if I make a prediction in English in your presence, you will understand it, will remember it for 30 seconds, and will be able to tell whether any of your actions will conform to it or not. Given all these conditions, whatever I predict—at least, if I make the prediction by saying it aloud, in your presence, in English, etc.—will be falsified. If I predict you will do  $A$ , then you will do  $\bar{A}$ , while if I predict that you will do  $\bar{A}$ , you will proceed to do  $A$ . In other words, in these conditions no prediction of mine is causally compossible with the occurrence of the event I predict. Let  $C_1, \dots, C_n$  be the (actual) conditions just delineated, let  $P_A$  be my predicting you will do  $A$  (announced in the indicated way), and let  $P_{\bar{A}}$  be my predicting you will do  $\bar{A}$  (announced in the same way). Then both sets  $\{C_1, \dots, C_n, P_A, A\}$  and  $\{C_1, \dots, C_n, P_{\bar{A}}, \bar{A}\}$  are causally impossible sets of events.

Notice that this example does not prove that it is causally impossible "simpliciter" for me to make a scientific prediction of your action. All that it proves is that I cannot make such a prediction in a certain manner, viz., by announcing it to you in English. The events  $P_A$  and  $P_{\bar{A}}$  include this particular manner, and that they do so is important. If I predict your action in some other manner, by thinking it to myself or by saying it aloud in Hindustani, for example, the effect on your action would not be the same as if I say it aloud in English. Assume that, if you do not hear me make any prediction or if you hear me say something you fail to understand, you will proceed to perform action  $A$ . Then it is causally com-

possible for me to predict your action correctly by announcing the prediction in Hindustani. In other words, letting  $P_A$  be my predicting that you will do  $A$  by announcing this in Hindustani, then the set of events  $\{C_1, \dots, C_n, P_A, A\}$  is a causally compossible set.

In determining whether or not a certain set of events, including (1) a prediction, (2) the event predicted, and (3) certain other assumed conditions, is a causally compossible set, it is essential to specify the manner of the prediction. This is true in general, not just in the case of predictions of action. A prediction which is "embodied" or expressed in one way will not have the same causal effects as the same prediction expressed in another way. We can see this in the case of the speck of dust. Jones predicted the position of the dust by announcing it orally, and this resulted in the falsification of the prediction. But had he made the same prediction in another fashion—say, by moving his toes in a certain conventional pattern—his prediction would not have been falsified, for the position of the dust would not have been affected.

What is the significance of the fact that it is causally impossible, in some circumstances, for a (correct) prediction of an action to be made in a specified manner? First, this unpredictability does not prove that these actions are undetermined. Indeed, the very construction of the case in which no prediction is possible presupposed the existence of laws of nature which, together with a given prediction, would result in a certain action. In short, the case under discussion should, if anything, support rather than defeat the thesis that actions are determined. The only reason one might have for thinking the contrary is the assumption—which should by now appear very dubious—that determinism entails predictability. What our present case shows, I think, is that under some circumstances, even a determined event may not be susceptible of being correctly predicted in a specified manner. This fact can be further supported by adducing a similar case connected with purely physical events. And this brings me to my second point: the case produced above does not reflect a peculiarity of human action, since parallel examples can be found among physical phenomena.

Imagine a certain physical apparatus placed in front of a piano keyboard. A bar extends from the apparatus and is positioned above a certain key. (Only white keys will be considered.) If the apparatus is not disturbed, the bar will strike that key at a certain time. Now let us suppose that the

apparatus is sensitive to sound, and, in particular, can discriminate between sounds of varying pitches. If the apparatus picks up a certain sound, the position of the bar will move to the right and proceed to strike the key immediately to the right of the original one (if there is one). Specifically, if the sound has the same pitch as that of the key over which the bar is poised, the bar will move. If the monitored sound has any other pitch, the bar will remain in its position and proceed to strike that key.

Now suppose that someone (or something) wishes to make predictions of the behavior of the apparatus. He wishes to predict what key the bar will strike. But the following restriction is made on the *manner* in which the prediction is to be made. The prediction must be expressed according to a specific set of conventions or symbols. To predict that the bar will strike middle *C*, for example, the predictor must emit a sound with the pitch of middle *C*. To predict that the bar will strike *D*, he must emit a sound with the pitch of that key, etc. All sound emissions are to be made in the neighborhood of the apparatus. Given this restriction on the manner of prediction, it will be causally impossible for the predictor to make a correct prediction. For suppose that the bar is poised above middle *C*. If he predicts that it will strike middle *C*—that is, if he emits a sound of that pitch—the bar will move and proceed to strike *D*. But if he predicts any other behavior of the bar, for example, that it will strike *D*, the bar will remain in its original position and strike middle *C*.

Admittedly, the manner of prediction I have allowed to the predictor of this physical phenomenon is much more narrowly restricted than the manner of prediction allowed to the predictor of human action. But we could imagine physical apparatuses with a greater degree of complexity, able to "refute" predictions made in any of a wider variety of manners. In any case, the principle of the situation is the same for both physical phenomena and human actions, though the manners of prediction which affect one phenomenon may be different from the manners of prediction which affect the other. The latter difference simply reflects that fact that physical objects and human beings do not respond in precisely the same ways to the same causes. But this is equally true of different kinds of physical objects and of different pairs of human beings.

<sup>15</sup> "An Essential Unpredictability in Human Behavior" in *Scientific Psychology: Principles and Approaches*, ed. by B. B. Wolman and E. Nagel (New York, 1965).

The reader should not suppose that the present discussion in any way vitiates my description of the book of life in Sect. VII. Our present discussion shows that under *some* conditions it is *not* causally compossible to predict a man's action in a way which allows him to learn of the prediction. But there are *other* conditions, such as the ones described in Sect. VII, in which such predictions *are* causally compossible. The existence of the latter conditions suffices to establish the possibility (in principle) of scientific predictions of voluntary actions which the agent hears or reads. Admittedly, it is not always possible to make predictions in this manner. But even when it is impossible to let one's prediction become known to the agent, it does not follow that it is impossible to make the prediction "privately." Thus, suppose you are trying to write a book of my life before I am born. Your calculations might show that if you inscribe certain predictions in the book they will be confirmed. For these calculations might reveal that I shall not read the book, or that I shall perform the actions despite the fact that I shall read the book. If so, you may proceed to write the book, having (scientific) knowledge that it will be correct. On the other hand, your calculations might reveal that, no matter what prediction you inscribe in the book, I shall refute it. In this case, you will be unable to write a book of my life. But you may nevertheless have scientific knowledge of what I shall do! Your calculations may reveal that I shall do a certain sequence of actions, as long as I do not come across any (putative) book of my life. If you decide not to write such a book yourself, and if you know that no one else will, you may conclude (deductively) that I shall perform the indicated sequence of acts.

## X

In the previous section we saw that, under certain conditions, it may not be causally compossible to predict a certain action in a specified manner. Recently, however, Michael Scriven has claimed that human behavior exhibits an even more important unpredictability.<sup>15</sup> Scriven writes: "So far we merely demonstrate that human choice behavior can be made at least as unpredictable as any physical system. In an important class of examples . . . , a stronger conclusion is demon-

strable."<sup>16</sup> Scriven's example consists in imagining an agent, *X*, who is contrapredictively motivated relative to a certain predictor, *Y*. That is, *X* wants to defeat any prediction *Y* makes about his actions. Scriven further supposes that *X* knows everything that *Y* knows about him. From this information, *X* figures out what *Y* will predict, or will have predicted, about *X*'s action. In other words, *X* "replicates" *Y*'s prediction; he comes to know what *Y*'s prediction is even though *Y* does not announce his prediction. After figuring out *Y*'s "secret" prediction, *X* proceeds to act otherwise. Scriven concludes: "... the present case is more interesting. The idea that human behavior is 'in principle' predictable is not seriously affected by the recognition that one may not be able to announce the predictions to the subjects with impunity (nor, more generally, can one allow them to be discovered). For one can make the predictions and keep them from the subjects. But in the present case, *one cannot make true predictions at all*. Secret predictions are still predictions; unmakeable ones are not."<sup>17</sup>

We must first note that Scriven has given a misleading account of his example in saying that "in the present case, *one cannot make true predictions at all*." (Italics his.) True, a particular person, *Y*, is unable to make correct predictions of *X*'s behavior. But *X*'s behavior is not completely immune to prediction. Scriven's case leaves open the possibility that there is, or was, some other being, *Z*, (who may have lived long before *X*) who predicted *X*'s behavior without *X* knowing of this. In order for *X*'s behavior to be *completely* immune to prediction, *X* would have to know with respect to *every* potential predictor—i.e., everyone who lived during or prior to *X*'s lifetime—what predictions, if any, be made about *X*'s behavior. Anything short of this state of knowledge by *X* would leave open the possibility that some being or other correctly predicted what *X* would do, indeed predicted it scientifically.

Secondly, it is questionable whether Scriven's example shows that human behavior is more unpredictable than physical systems, as he suggests in the passage on page 413. Admittedly, Scriven's case goes beyond my previous example in one respect. In my example only certain manners of prediction lead to the performance of a different

action. In Scriven's example, *any* manner of prediction leads to a different action. This is because even the *minimal* manner of "making" a prediction—i.e., having a future-looking *belief*—is self-defeating. Nevertheless, it may still be possible to duplicate Scriven's human behavior case with physical systems. Suppose, for example, that we found neuro-physiological states that correlated with beliefs. That is, suppose we found one-one correlations between a person's believing certain propositions and his being in certain neuro-physiological states. We might then "hook up" a physical system to a potential predictor in such a way that the state of the system is causally affected by the beliefs (or their neuro-physiological correlates) of the predictor. The physical system would be arranged so that whenever the potential predictor had a belief about a future state of the physical system, this belief would cause the system to go into another state instead.

The third and most important point I wish to make is based on a criticism of Scriven by David K. Lewis and Jane Shelby Richardson.<sup>18</sup> In the competition between Scriven's agent and predictor, each is trying to get sufficient information about the other and to calculate from this data just what the other will do (or believe). Let us combine both of these factors—data and calculation—and call them "knowledge," since the function of the calculations is to add to the calculator's knowledge. Lewis and Richardson argue forcefully that it is impossible for *both* the predictor and the agent to have sufficient, or complete, knowledge of his opponent. We can construct two sorts of cases. We can endow the predictor with complete information about the agent, but this forces us to deny to the agent complete knowledge about the predictor. Or we can endow the agent with complete knowledge about the predictor, but in so doing we must deny complete knowledge about the agent to the predictor. Scriven does the latter, although he is not quite aware of this. In saying that the agent is able to "replicate" the predictor's prediction and then decide to act otherwise, he is in effect saying that there is some aspect of the agent's motivational structure which is *unknown* to the predictor. But if there is some fact relevant to the prediction which the predictor does not know, it is not surprising that he makes

<sup>16</sup> *Ibid.*, p. 413.

<sup>17</sup> *Ibid.*, p. 414.

<sup>18</sup> "Scriven on Human Unpredictability," *Philosophical Studies*, vol. 17 (1966).

an incorrect prediction. He simply is not in a position to make a scientific prediction. This hardly shows that the agent's behavior is inherently immune to scientific prediction. It merely shows that this particular predictor,  $\mathcal{Y}$ , does not know enough about  $X$  to make a scientific prediction.

For this reason Scriven's case is less interesting than the one I presented in the foregoing section. In my example, the predictor is unable to make a correct prediction (in a specified manner) even though he has all relevant information. In Scriven's example, the predictor lacks some relevant information. Scriven's example, then, hardly warrants us in concluding that human behavior is undetermined. The fact that someone with insufficient knowledge is unable to predict an event correctly does not at all suggest that the event is undetermined. Of course, Scriven does not claim that his example shows behavior to be undetermined. I say this merely to remind the anti-predictionist that he can take no comfort from Scriven's case.

## XI

I have shown that there are causally compossible worlds in which voluntary actions are scientifically predicted. Let us now see whether there are causally compossible worlds in which a person scientifically predicts one of his *own* actions. I think that there are such worlds and I shall illustrate by continuing the description of the world I was sketching earlier.

Having tested my book of life on a very large number of occasions during many months and failed to refute it, I become convinced that whatever it says is true. I have about as good inductive evidence for this proposition as I do for many another proposition I could be said to know. Finally, I get up enough courage to look at the very end of the book and, as expected, it tells when and how I shall die. Dated five years hence, it describes my committing suicide by jumping off the 86th floor observation deck of the Empire State Building. From a description of the thoughts which

will flash through my mind before jumping, it is clear that the intervening five years will have been terrible. As the result of those experiences, I shall have emotions and desires (and beliefs) which will induce me to jump. Since I trust the book completely, I now conclude that I *shall* commit suicide five years hence. Moreover, I can be said to *know* that I shall commit suicide.

As described so far, we cannot consider my prediction of my suicide a "scientific" prediction. To be a scientific prediction the predicted event must be *deduced* from laws and antecedent conditions, while, as I have described the case, no deduction was involved. However, we might supplement the situation so as to include a deduction. The book may be imagined to list the relevant physical and psychological laws (in a footnote, say) and the relevant conditions which determine my committing suicide (my intention to commit suicide, my proximity to the fence surrounding the observation deck, the absence of guards or other interfering factors, etc.). From these laws and conditions I actually deduce my future action.<sup>19</sup>

This example shows, contrary to the view of some authors, that we can have inductive knowledge of our own future actions, knowledge which is not based on having already made a decision or formed an intention to perform the future action. Stuart Hampshire, for example, has recently written, "... I cannot intelligibly justify a claim to certain knowledge of what I shall voluntarily do on a specific occasion by an inductive argument; if I do really know what I shall do, voluntarily, and entirely of my own free will, on a specific occasion, I must know this in virtue of a firm intention to act in a certain way."<sup>20</sup> The case outlined, I believe, shows that Hampshire is mistaken. In that case, there is a time at which I do have certain knowledge of what I shall do (at any rate, about as "certain" as one can be with inductive evidence) and yet I have formed no intention nor made any decision to perform that action. At the time I read the book's prediction, I do not intend to commit suicide. But although I do not intend to commit suicide, I fully believe and know that, five

<sup>19</sup> That these conditions will actually obtain is, of course, open to doubt. Moreover, I have not learned of *them* by scientific prediction. I have simply "taken the book's word" that these conditions will obtain; I have not deduced them from other, still earlier, conditions. However, there are no restrictions on the manner in which a predictor comes to know antecedent conditions. One way predictors might learn about antecedent conditions is by using various measuring devices and instruments, the reliability of which is supported by inductive evidence. My book of life may be regarded as such a device, and my inductive evidence supporting its reliability may be as strong as that supporting the reliability of various other devices which scientists commonly use for obtaining knowledge of antecedent conditions.

<sup>20</sup> *Freedom of the Individual* (London, 1965), see p. 54.

years later, I shall intend to commit suicide. I firmly believe that, at that later time, I shall feel certain emotions and have certain desires which will induce me to jump off the Empire State Building. At the time of my reading the book I do not feel those things, but I commiserate with my future self, much as I commiserate with and understand another person's desires, beliefs, feelings, intentions, etc. Still, my understanding of these states of mind and of the action in which they will issue is the understanding of a spectator; my knowledge of these states and of my future action is purely inductive. Moreover, this knowledge is of a particular *voluntary* act to be performed at a specified time. Though the suicide will be a "desperate" action, it will in no sense be "coerced" or done unknowingly; it will flow from a firm intention, an intention formed very deliberately. But that intention will not be formed until after I have had certain experiences, experiences which, at the time I am reading the book, I have not yet had.

We can imagine two alternative series of events to occur between my reading the book and my suicide. First, I might *forget* what I have learned from the book, and later decide to commit suicide. Secondly, while never forgetting the prediction, the knowledge of my future suicide may gradually change from more inductive knowledge to knowledge based on intention. In this second alternative, there is never any "moment" of decision. I never pass from a state of complete doubt about committing suicide into a sudden intention of committing suicide. Rather, there is a gradual change, over the five-year period, from mere inductive knowledge that I shall commit suicide to an intention to commit suicide. When I first read the book I am fully prepared to assent to the proposition that I shall commit suicide. But I am saddened by the thought; my heart isn't in it. Later, as a result of various tragic experiences, my *will* acquiesces in the idea. I begin to welcome the thought of suicide, to entertain the thought of committing suicide with pleasure and relief. By the time the appointed time comes around, I am *bent* on suicide. This gradual change in attitude constitutes the difference between the kind of knowledge of my future suicide, the difference between mere inductive knowledge and knowledge based on intention. Hampshire claims that the first kind of knowledge of one's own action is

impossible. The present case, I believe, shows this claim to be mistaken.

Many philosophers seem to be very uncomfortable with the idea of a book of life. They believe that the existence of such books—or of foreknowledge of actions in any form—would deprive us of all the essential characteristics of voluntary behavior: choice, decision, deliberation, etc.<sup>21</sup> I do not think this fear is warranted, however. I have just shown that even if a person reads what a book of life predicts, and believes this prediction, he can still perform the indicated action voluntarily. Moreover, the existence of predictions which the agent does *not* read leaves ample opportunity for deliberation and decision. An agent may know that a book of his life exists and yet proceed to make decisions and to deliberate as all of us do now. The agent's belief that there is such a book, and his belief that the book's existence implies that his actions are causally necessitated, is compatible with his deliberating whether to do one action or another. Although his future action is causally necessitated, one of the antecedent conditions which necessitate it is his deliberation. Indeed, the prediction in the book of life was made precisely because its writer knew that the agent would deliberate and then decide to do the predicted action. Thus, the book of life can hardly be said to preclude deliberation. Nor does the book of life imply that the agent's deliberation is "for naught," or "irrelevant." On the contrary, his deliberation is a crucial antecedent condition: were he not to deliberate, he probably would not perform the action he eventually does perform. Deliberation and decision are perfectly compatible with the existence of books of life; and they are perfectly compatible with the thesis that they, and the actions in which they issue, are determined.

## XII

If actions are determined, there is at least one sense in which an agent "cannot" act other than he does: his actual action is causally necessitated by actual prior events, and hence any other (incompatible) action is causally impossible with actual prior events. It is precisely because his action is causally necessitated that it is amenable to scientific prediction. But it is generally accepted that if an agent does an action *A* voluntarily, he

<sup>21</sup> One such philosopher is Richard Taylor. See his "Deliberation and Foreknowledge," *American Philosophical Quarterly*, vol. 1 (1964), pp. 73-80. Many others could also be named.



also "can" do otherwise. So far we have only provided a sense of "can" in which he *cannot* do otherwise. We are therefore obliged to identify some *other* sense of "can," or "could" in which an agent "can" do actions he does not in fact do. Such a sense of "can" has long been in the literature, having been defended by Hobbes, Edwards, Hume, Mill, Moore, Nowell-Smith, Stevenson, and others.

Suppose that John does *A* voluntarily. To say that he "could have" done otherwise indicates that he had another alternative open to him, the alternative of doing  $\bar{A}$ . But what is meant by saying that  $\bar{A}$  was "open" to John? What is meant, I believe, is that, *if*, contrary to fact, John had wanted or chosen or decided to do  $\bar{A}$ , then he would have succeeded in doing  $\bar{A}$ . The "alternative" open to John is not the alternative of doing  $\bar{A}$  given his (actual) desire or decision to do *A*, but the possibility of doing  $\bar{A}$  relative to a (non-actual) desire to do  $\bar{A}$ . Here, as elsewhere, the determinist's sense of "could have" involves the supposition of counterfactual conditions; indeed, if taken to its logical conclusion, a whole counterfactual world. In the "could have" pertaining to action, the main counterfactual feature pertains to the agent's desire, intention, or decision. But the analysis of "could have" is not wholly counterfactual. To say that John "could have" done  $\bar{A}$  is also to make a

categorical assertion about the real world, viz., that John *was able* to do  $\bar{A}$ , that he had the ability and the opportunity to do  $\bar{A}$ . John's actual ability and actual opportunity were such that, if his desire or intention to act had been different, his action would have been different. Thus, the analysis of "John could have done  $\bar{A}$ " would be formulated as: "John had the ability and opportunity to do  $\bar{A}$ , and (therefore) if he had decided to do  $\bar{A}$ , he would have done  $\bar{A}$ ."

□ A scientific predictor who predicts that someone will perform a voluntary action *A* may thus be justified in accepting the following two propositions: (1) The agent *will* (certainly) do *A*, and (2) The agent *could* do  $\bar{A}$ . (2) is warranted because, as the predictor realizes, the agent will have the ability and opportunity to do  $\bar{A}$ . (1) is also warranted, however, because, as the predictor knows, the agent will *in fact* choose to do *A*, not  $\bar{A}$ . The predictor knows that he will choose to do *A* because he has deduced that he will so choose from antecedent conditions and psychological laws having choices (or desires) as their dependent variables.

The question remains whether the specified sense of "could" is *the* sense relevant to the traditional problems of freedom and responsibility. My own opinion is that this is the relevant sense, but I have no new arguments to give on this score. Here I wish not to join the fray, but to leave it. □

The University of Michigan

Received March 22, 1967

