PREDICTING DISCOVERIES AND THE RULE-DESCRIPTION ARGUMENT

Michael E. Levin

Professor of Philosophy
City University of New York
February, 1972

I

It has been claimed (most notably by Karl Popper (1)) that it is in principle impossible to predict discoveries. This in turn is taken to show that human behavior could never be the subject of a completely predictively adequate science. Without going into unnecessary detail about the relation of predictability to determinism — the larger thesis on which "predictivism" usually relies — I argue below that the plausibility of the Popperian thesis rests on an ambiguity in the phrase "predicting a discovery". I argue that in the one sense of this phrase in which "Discoveries are unpredictable" is true, that claim offers no obstacle to the claim that all human behavior is predictable.

I will also make what is essentially the same distinction from the predictivist's perspective. The predictivist holds that (all) aspects of human behavior are in principle predictable. I will argue that in the sense of "discovery" in which discoveries are not predictable, the predicate '— is a discovery' does not hold of instances of human behavior; and in the sense of "discovery" in which the predicate '— is a discovery' does hold of instances of human behavior, there is no logical incoherence in holding that discoveries are predictable.

1) Determinism, which I take to assert at least that all events are connected by laws with preceding events, has as a corollary the connection by laws of what I will do in my laboratory tomorrow with what is true of me today. A further commitment of determinism is that, unless some special reason is

presented to show that, unlike other law-governed connections in nature, this law-governed connection is undiscoverable, there is nothing in principle to prevent behavioral scientist S from enunciating the appropriate law today, and on its basis predicting what I will do in my laboratory tomorrow. If tomorrow I will discover a new isotope of manganese, S would be able to predict today that tomorrow I will discover that new isotope. To the degree that we accept determinism — and the pull of determinism is always strong — it appears that we shall have to admit that discoveries are in principle predictable. That implication of determinism is what I am concerned with here.

2) On the other hand, the supposition that a discovery is even in principle predictable appears to lead to absurdities. (The argument is essentially Popper's). To be able to predict that tomorrow I will become the first person ever to find out that there are atoms of atomic weight 142 and atomic number 58 (and that is what it is to discover this), S will have already to know that there are atoms of atomic weight 142 and atomic number 58, for S must know what my discovery consists in to be able to predict that I will make it. (Notice that it is not enough for S, at t, simply to describe my discovery with some locution like: «At t' [$t' > t$] Levin will find out something about the isotopes of manganese". Predictivism is committed to the much stronger claim that S could in principle specify my discovery with unlimited fineness of description. To claim less would be to concede that certain particular events are unpredictable). But then, when I find that there is such an isotope, it will not be the *first time* anyone has found this. So I didn't make a discovery after all, contrary to the hypothesis. The problem is that, to predict a discovery, S has to know something before anyone knows it; he has to have done something before the first person to do it has ever done it.

This argument might seem vulnerable to the observation that the word "discovery" is not used in ordinary discourse in a way that fits it. Do we not say "Leverrier discovered Neptune independently of Adams, and only a month after" and "There

are certain truths about life that each person must discover for himself," thereby allowing that a man can be said to discover what is already known ? In fact, however, this point of usage can be conceded, and Popper's original argument remounted. Let us call "X-ing" what someone does when he becomes the first person ever to find something out. The argument presented in (2) can now be reformulated to show that cases of X-ing are unpredictable. This becomes an argument against predictivism if we make the plausible assumption that there have been cases of X-ing (for example, Newton X-ed the law of gravity).

3) Let me now introduce a distinction that will, I think, point the way out of the dilemma. This distinction, however artificial it may first appear, will not, I hope, be intrinsically controversial.

The word "discovery", it seems to me, can have two senses; more to the point, whatever senses "discovery" may have, it is being used in at least two ways in the foregoing dialectic.

(i) By "discovery" one might mean, literally, the words formulating a true novel claim about the world. In this sense, "$E = mc^2$" is Einstein's discovery. That is, in this usage we are quite divorcing the sentence formulating the discovery from the *extra information* that this sentence is true. In this sense, a list of discoveries would be a list of sentences or a series of utterances. In this sense, "P discovered D" describes the event of P's writing or uttering the sentence "D" ([2]). That this is an invented sense of "discovery" is no matter, for precisely what I will argue is that the predictivist aims to predict (the making of) discoveries in *this* sense.

(ii) By "discovery" one might mean *both* the words formulating (what is in fact) a true, novel claim about the world, *plus* the information that this sentence *is* a true, novel claim about the world. In this second sense, Einstein discovered that $E = mc^2$. I will eventually argue that it is in this sense that discoveries are unpredictable.

4) Notice first that in sense (ii) I can know "P made discovery

D" only if I know both that P produced the sentence "D" and that "D" is true, whereas this is not so in sense (i). Suppose that, quite ignorant of the truth of "D", I predict that tomorrow P will become the first person ever to say "D". P does, and "D" is true, and P is *a fortiori* the first person ever to notice this. In sense (i), I have predicted a discovery by P, since "discovery" here refers only to the sentence produced by P. I was able to predict that P would discover D without knowing that "D" is true. Thus, in sense (i), so long as I have no idea that "D" is true, it is possible for me to know that P will say "D" without beating him to D, without contradicting the hypothesis that P is the discoverer of D. In this sense even a computer could predict discoveries.

Notice too that the second clause of sense (ii) renders "made a discovery" more than a prediction of P's thinking or behavior, more than a characteristic of P. Clearly, that "D" is a true sentence is not a trait of P or P's act of uttering "D"; P's behavior in uttering "D" is invariant to "D"'s truth-value.

5) These observations make it easier to sort out what the predictivist and the Popperian are each entitled to. The predictivist claims that if S knew enough, S could now completely enumerate everything I will ever do or say, all my vocal and inscriptional productions. (And clearly '— produced incription or utterance "D"' is a property of persons.) In no way will S's prediction compromise the novelty of D, if D is a discovery. For 'being a discovery' is a property that holds of "D" in virtue of a relation between "D" and the world. There is no need for S to know *this* fact about "D" if he knows that, at t', I will produce "D". In sense (i), S has predicted a discovery. S does not have to know that "D" satisfies the semantic description "is true" for him to know that I will say "D" at t'. And this is all the predictivist claims he can do. This is compatible with "D"'s being (unbeknowst to S at t) a true, novel claim at t'. Since S at t has no idea if "D" is true, "D" is being *claimed* for the first time at t'. What is impossible — and here the Popperian is right — is that (i) S know that at some future time I will utter "D", (ii) S know that in so doing I will

be saying something true, and (iii) I discover D. But this is not what the predictivist claims is possible.

6) There are objections to this way of meeting Popper's argument. (i) The anti-predictivist might say that we ought not describe what S has done as "predicting a discovery" if S does not know that the utterance whose occurence he predicts is true. But this is a verbal dodge, conceding that the predictivist might well accomplish his most controversial goal, the prediction of my behavior. The realizability of *this* program is the philosophical issue at stake. If we are left with quibbling about how to describe what the predictivist has done, the predictivist has won. Certainly Popper thought his argument showed that there could be no exact predicting of what I will *do* tomorrow (4).

(ii) Someone might also object that it is unlikely that S will be able to predict that I will utter "D" in my laboratory unless S has $\vdash$ D among his premises, that it is extravagant fiction to suppose that S could predict that I will seriously say "Manganese 142 is radioactive" by any argument other than that Manganese 142 *is* radioactive and that I will notice this. Perhaps so. But what is crucial is that this argument no longer purports to show a logical incompatibility between "discovery" and "prediction." It is an inductive argument about the dim prospects for predicting the behavior of an organism without certain information about the organism's environment. This argument abandons the distinctively Popperian claim that there is a *conceptual* incoherence in the supposition that S predicts my "D" — utterance at t', where in uttering "D" at t' I make a discovery.

(iii) More generally, it might be suggested that, human behavior being supremely complicated, nothing short of omniscience could as a matter of fact be competent to make predictions about it. To be in a position to predict human behaviour would require that we know all, and this implies that there are no more discoveries to be made. But again, such an incompatibility between predictivisim and discovery-making would not be logical, but would rest instand on alleged facts about the difficulty of making a certain kind of prediction.

7) Let me restate the points made in (3) - (5) in a way which brings them into line with the distinction I drew in (3). The one sense of "discovery" in which S could not predict my discovery of D was found to be compatible with predictivism about human behavior. Predictivism says that S could know, today, everything that I will do or say or think tomorrow. Now in the sense in which "Levin discovered D" is unpredictable, this sentence asserts a relation between something I do — saying "D" — and its truth condition; the sentence "Levin discovered D" does more than describe something I did. To say "I made a discovery" in the unpredictable sense is to assert (i) that I said something, and (ii) that what I said was, among other things, true. The predictivist claims only to be able to predict facts of type (i). Put in terms of he distinction in (3): In sense (i) of "discovery," discovery-making is both an aspect of my behavior, and, for all the Popperian has shown, in principle predictable; in sense (ii) of "discovery," discoveries are not in principle predictable, but at the same time they are not cases of human behavior. When Popper said that discoveries are as a matter of logic unpredictable, he must have been using "discovery" in sense (ii), for it is only in that sense that discoveries *are* unpredictable a priori. However, the predicitivst is committed only to the claim that he can predict discoveries in sense (i): he is committed only to the claim that he could predict every aspect of my behavior. If my argument is correct, then, the Popperian claim rests on the fallacy of equivocation.

What I have said in no way positively supports predictivism. It merely shows that one popular and persuasive argument against it is fallacious.

II

If we grant that all bodily events are predictable from laws and initial conditions formulable entirely in the language of physics, in what sense are *discoveries* predictable ? I have suggested a precise answer to this question — discoveries are

predictable in sense (i). But this is not enough, for what is wanted is a more complete evaluation than I have offered of the *importance* of this one sense in which discoveries are predictable, and an evaluation of its bearing on the overall question of the predictability of human behavior. I preemptorily dismissed this question earlier by saying that predictability of discoveries in sense (i) is the predictvist's most controversial goal, and in so saying implied that predictability of discoveries in sense (i) would sanction calling discoveries predictable simpliciter. A possible objection to this is that "discovery" (i) is a philosopher's coinage, that the nuclear sense of "discovery" is (ii), and that therefore, properly speaking, discoveries are unpredictable. I want now to present the case against this objection.

In fact, the main motivation behind the claim that sense (ii) is the core of "discovery" is one instance of a more general claim that has achieved some currency in the literature on the theory of action. This enhances the interest of the present issue, since it can rapidly be generalized into a test case for an important general thesis. The generalization is this. It is commonly claimed that, if a predicate P occurs non-trivially in a statement S, the state of affairs referred to by S cannot be P. So, for example, the fact formulated as "Mercury expands when heated" cannot be explained by a theory in which the term "mercury" does not occur. (4) It is then argued that certain predicates which must appear in descriptions of human actions are not physical-object (or bodily) predicates; therefore, even if all my bodily events were predictable from the laws of physics, my actions, under any of these non-physical descriptions, will not be physically predictable. For instance: since "writes a check" is not a term of physics, "Levin will write a check tomorrow" could never be predicted on the basis of physical laws, even if all the movements of my body could be. Moreover, since of two (nearly) physically idental events one can be a case of check-writing and the other not (if, as is sometimes said, the "contexts" of the two events differ), and two physically quite dissimilar events could both be cases of check-writing (if the "contexts" are right), it seems reasonable

to conclude that no physical predicate is even extensionally equivalent to the predicate "x is a case of checkwriting," and even that there is no physical predicate T such that for any case *a* of check-writing there is some physical event *b* satisfying T which is necessary and sufficient for the occurrence of *a*. Now, all the usual examples invoked by proponents of this argument are descriptions involving rules (and the parenthetically-noted contexts are what determine if the norms or quasi-norms of the rule are satisfiable): "writing a check" involves complying with certain legal norms, and (presumably) explained or predicted by statements which do not contain the concept of a legal norm is no part of physics nor extensionally equivalent to any part. Hence I will (somewhat misleadingly) call the argument just stated the rule-description (r-d) argument.

This same pattern of argument can be made to apply to "discovery" in sense (ii); since a semantical condition must obtain for "discovery (ii)" to apply, the occurrence of a discovery is unpredictable relative to the class of physical-object predicates. That is, to say that P is a "discovery (ii)" is to say, among other things, that the world is as P, via the conventions governing its constituent terms, says it is: and semantical conventions are a kind of rule. Of course, this observation has anti-predictivist force only if sense (ii) of "discovery" is taken to be central; but this is what I am granting arguendo to present the anti-predictivist case at its most advantageous, and to make its fortunes the fortunes of the r-d pattern of argument in general.

There is a difference between the general r-d argument and the "discovery" case worth noting. The general r-d argument does not rule out the possibility that laws couched in some vocabulary richer than that of physics could be predictively adequate for human behavior under rule-descriptions, although it is frequently urged that such laws would, for this very reason, have to be non-causal. Popper's argument aims to show that no vocabulary, however rich, could express laws competent to predict the occurrence of events describable as "discoveries". This, however, does not attenuate the connection bet-

ween the two arguments that I want to make, since my upcoming argument, if correct, will show that these vocabulary discrepancies are largely irrelevant to the question of predictability. Thus, more and less extreme interpretations of such vocabulary discrepancies collapse together. In any case, I hope to have this neutralization of the Popperian argument fall out as a special case of the neutralization of the r-d argument — so I want to concentrate on that version of the one which rests on the other. (Indeed, one might even argue that "discovery (i)" is a rule-description, for it is not at all clear that in saying "X asserted P" or "X said P and meant it" we are saying merely that something physical, involving X's body, occurred. Again, the considerations immediately below will accommodate this complication.)

I want now to argue that even if the r-d argument is cogent, the predictability of bodily events via physics would diminish (perhaps to vanishing) the importance of the sense in which our actions are unpredictable.

1) Let us say that a set of predicates A is predictively closed just in case any event describable wholly in terms of A is predictable from laws and initial conditions statable wholly in terms of A. For example, (position, momentum) is not predictively closed, since magnetic phenomena can affect the position and momentum of a body. Physical determinism claims that physical science will (someday) yield a predictively closed set of predicates. One suspects that it is part of the motivation of the r-d argument that rule-descriptions are necessary for the explanation of certain *physical events* ([5]), that the world describable in physical terms is not a "closed system", and that, therefore, certain physical events are physically unpredictable. Suppose a certain human body is moving up and down, and this is described as "Jerry Lucas rebounding." The r-d argument, I suspect, wants to say that something physical here is physically unpredictable, describable as it is only via the rules of basketball. I take this something to be: Jerry Lucas moving up and down. For surely, if that *is* physically predictable, then "what Lucas did," his "rebounding", to whatever extent it involved Lucas' bodily motion, is not unpredictable

either. In traditional terminology, we do not want a will that is transcendental, and not a necessary condition for a least some events in the physical world. A transcendentally unpredictable will is of no use to anyone who hopes (perhaps because of the supposed demands of ethics) that actions that necessarily involve his body — like "being in New York for a friend's wedding" — are in some sense physically unpredictable.

But it is just here that the r-d argument is vulnerable, for it is entirely impotent to show either that the physical event of Lucas' moving up and down was physically unpredictable, or that the predictability of this physical event is not, in some way that remains to be discussed, tantamount to the predictability of Lucas' rebounding.

This line of reasoning in no way assimilates action to bodily happening. I am suggesting only that the r-d argument, even if in some way correct, goes no way toward showing that human behavior, suitably described, is physically unpredictable. Specifically, I am willing to defend, or at least consider sympathetically, a thesis which brings actions and bodily events sufficiently close together to warrant calling the prediction of a bodily event the predication of an action: If *a* is an event which must be described by non-physical predicates, there is some event *b*, describable wholly by physical predicates, such that *a* cannot occur unless *b* occurs, and if *b* occurs, *a* occurs. I call this relation between *a* and b "rooting," and the thesis the "rooting thesis". R(x,y) can abbreviate "x is rooted in y", and the rooting thesis is just that (x) (Ey) R(x,y). For instance: I cannot perform the particular action of writing this check unless my hand moves, and if my hand moves (in these circumstances) then I have written a check. It is difficult to characterize the relation R(x,y), and I am even tempted to say that it is nothing less than *strict identity*. The points I want to make using R(x,y), however, do not depend on the implication $R(x,y) \rightarrow x = y$, but only on $R(x,y) \rightarrow$ (y's occurrence is necessary and sufficient for x's).

At this point proponents of the r-d argument may remind me of the concession that no physical event, is necessary or

sufficient for the occurrence of a typical (rule-describable) action. Given any arm-motion whatever, I can make it and not write a check (in case, say, my bank is defunct); and concerning no movement of my arm *m* is it true that "If Levin writes a check then Levin's arm undergoes *m*.» But this well-taken point, being about the relation between types of physical events and types of actions, does not impugn the rooting thesis. For what that says is that I cannot write a check without doing *something or other* with my arm or body. And indeed it is hard to imagine that I could remain perfectly motionless and write a check. I find it hard to suppoe that I could make no motion at all, avoid basketball courts, and omit every member of a large (and in many cases antecendently specifiable) set of physical events, and still be said to have "hit an outside shot." And this is not to say that hitting an outside shot, or writing a check, must always be rooted in some one type of physical event. Finally, by the same token, it is hard to see how I could assert "p" without *something* physical happening. That is to say, the rooting thesis is easily reconciled with the claim that no physical predicate is co-extensional with a predicate like "x is a case of check-writing" if we distinguish type-type rooting from particular-particular rooting. A type-type rooting between two types of things T and T' asserts that for every x such that x is a T, there is some y such that y is a T' and R(x,y). If Tx is some r-d predicate, then to say that "being a T" is type-wise rooted in being a T', where T' is a physical-event predicate, is to make the admittedly false claim

1. $(ET')\ (x)\ (Tx \rightarrow (Ey)\ (T'y\ \&\ R(x,y))$.

But surely a better rooting thesis is the weaker claim that for any particular event which satisfies Tx there is some physical event predicate T' and some physical event y such that T'y and x is (or is rooted in) y; where, if x_1 and x_2 are T, and $R(x_1,y_1)$ and $R(x_2,y_2)$, it is not required that $(ET')\ (T'y_1\ \&\ T'y_2)$. Each particular action is rooted in, or perhaps identical to, some particular physical event:

2a. $(x)\,(Tx \rightarrow (ET')\,(Ey)\,(T'y \,\&\, R(x,y))$; i.e.

2. $(x)\,(ET')\,(Tx \rightarrow (Ey)\,(T'y \,\&\, R(x,y))$.

1 can be false and 2 true; to suppose $-1 \rightarrow -2$ is to commit a familiar quantifier-shift fallacy. ([6])

In general, the following definitions are useful in formulating various reductionistic theses. A relation R(x,y) is a *type-connector for* T if $(ET')\,(x)\,(Tx \rightarrow (Ey)\,(T'y \,\&\, R(x,y))$. R(x,y) is a *category-connector* between T and a (possibly infinite) set of predicates *A* if $(x)\,(Tx \rightarrow (ET')\,(Ey)\,(T' \in A \,\&\, T'y \,\&\, R(x,y))$. R(x,y) is a *weak* category-connector if it is a category-connector but not a type-connector. The rooting relation weakly category-connects rule-descriptions to the set of physicalistic predicates.

The rooting thesis, if accepted, severely limits anti-predictivism based solely on semantic considerations. If it is true, then for any rule-described event *a*, there is a physical event *b* in which *a* is rooted and which, we are assuming, is predictable on the basis of physically-expressed laws alone. This means that, in some important sense, *a* is predictable. The particular event *a* could not have occurred unless *b* did, while *b*'s occurrence was sufficient for *a*'s occurrence. Only *a-b*, whatever of the action is more than the root physical event, is even a candidate for unpredictability.

Without undertaking a full discussion of the Wittgensteinian conundrum "What remains of my raising my arm when the rising of my arm is subtracted?" we can cement by illustrations the intuition that, for purposes of assessing the prospects for predictivism, the answer is "not enough to warrant denying that an action is physically predictable if its root event is." Suppose scientist S correctly predicts that tomorrow the physical body denominated "Cleon Jones" will hit a baseball over a wall to the accompaniment of 125 decibels of laryngial noise. Has S predicted that Jones will hit a home run? Of course, a physically very similar event could fail to be a home run — remember batting practice. But this only shows that the predicate "is a home run" is not type-connected to any physica-

listic description. Surely, if S produced overwhelming evidence for his prediction, formulated in physicalistic terms, we would have to admit that it was likely that Jones would hit a home run tomorrow.

Before pulling the various strands together, it is worth mentioning an objection to which the rooting thesis, as a justification of predictivism, seems vulnerable. If R(x,y) entails that x is necessary and sufficient for y, it also entails that y is sufficient and necessary for x. Why then do we not use Jones' hitting a home run to explain its roof event of ash meeting horsehide instead of the other way around? The answer is just that root physical events have a place in a larger, better understood, system than the events they root. If we decided to explain x by the non-causal conditions of y given R(x,y), there would be many events z in the category to which x belongs but such that —(Ey)R(z,y); so z would be unexplainable. But if we explain and predict y via the causal conditions for x given R(x,y), then the thesis (y) (Ex)R(x,y) insures that *every* event belonging to y's category is predictable.

All the predictivist requires is that the set of physical-event predicates be predictively closed. He must, as a matter of logic, concede that facts formulated in terms of (say) rule-concepts could not be deduced from any statements in which those concepts do not appear. But what he insists on is that 1) those concepts are not needed to explain any physical event, and 2) every event that satisfies some rule-concept satisfies some other concept in fact available to him. These claims could never be overthrown by *a priori* reflection on the sorts of predicates available to the physical predictivist.

NOTES

(1) *The Poverty of Historicism,* London, 1957, pp. vii-viii. See also Peter Winch, *The Idea of a Social Science,* 1958, p. 94.

(2) If this is too behavioristic, we might include "the thought that D is the case" in this sense of discovery. What then comes wihin the purview of predictivism is the prediction of the occurrence of this thought in P's

mind. In general, when I speak of P's saying "D", we may add "or believing 'D'".

(3) Op. cit.: "We cannot, therefore, predict the future course of human history".

(4) Perhaps this principle is evident only on the nomological-deductive theory of explanation, a theory not without critics. But since my aim is to render nugatory an argument based on this principle, rejecting it from the start is just a short-cut to my eventual position.

(5) In a mixed material and formal mode: that some rule-described events are necessary conditions for the occurrence of certain physical events, paritcularly those involving the bodies of people.

(6) I would suggest that a number of objections to the mind-brain (or mind-body) identity thesis can be met by a parallel distinction between type-type and particular-particular identities.