MICHAEL SCRIVEN

AN ESSENTIAL UNPREDICTABILITY IN HUMAN BEHAVIOR*

IN THE HISTORY OF SCIENCE, there have been many memorable occasions when philosophers or scientists have laid it down that something cannot be done, or done in a certain way, for certain a priori reasons. We recall these cases chiefly for the dates when the claim was disproved.

But the practice is by no means as foolish as these failures are often thought to indicate. For the a priori is here no more than the level of fundamental theory, and to make predictions—including predictions about the impossibility of making predictions—is the proper task of fundamental theory. Indeed, the main problem with most fundamental theories, as Popper is prone to remark, is to squeeze any predictions out of them so that they *can*—in principle—be falsified. Furthermore, our memory is remarkably treacherous on this subject, since victories are sweeter than failures, and a careful survey of the history of science reveals a very different story. For example, as any new trend develops in a science, its prophets defend it by making pessimistic claims about the limited possibility of success with the older approach, and *these* impossibility claims have often proven well founded. A classic example is to be found in the history of psychology itself; with the advent of behaviorism and the experimental approach in general, the prediction was often made by the modernists that the old armchair introspectionist method would never get anywhere because it was incompatible with the quantitative objective tools of modern science. And so it came to pass, if we are to believe the current histories of psychology.

But even this coin has another side. For the actual success of the experimental method in psychology in producing an axiomatic theory has so far been very much less than its earlier proponents expected. It is perhaps worth examining pessi-

* Gratitude for stimulation or correction but no blame to the following, as well as those mentioned in the paper: Roger Buck, Dave Harrah, Frederic Mosteller (especially), Michael Polanyi, Meyer Shapiro, Lancelot Whyte, and the members of my X663 seminar at Indiana, especially John Cole, David Hull and Will Humphreys.

mistic prophecies about the prospects of the new Newtonian kind of psychology. Such prophecies are by no means new, and have attended the subject since birth, an event which—depending on one's point of view—occurred with Aristotle, Wundt, Pavlov, or Hull. Various reasons have been given for such pessimism. Human beings were alleged to transcend the kind of approach that was so successful with inanimate objects; they had souls, or free will, or required empathic understanding. We can safely say that these skeptical forecasts were *largely* based on beliefs which have not turned out to be well supported. But they are *partly* based on an instinctive appreciation of some sound points. In this chapter, I present a formal statement of one underlying truth that is closely related to these doctrines. To place it in context, I add that I think it is only one of several points that can be made in support of the view that the appropriate model for psychology is totally unlike Newtonian physics, that absurdly simple and atypical branch of science, and is much more like geology or geography with their limited developmental theories covering only the broad outlines of events and a mass of organized but non-lawlike information, much of it quite restricted in application.

DEVELOPMENT

The point I shall take up here arose from a study of the so-called self-defeating predictions (seldeps) of recent sociological theory (such as "Dewey will win": George Gallup) in the following very simple way. (Analogous remarks apply to self-fulfilling predictions (selfups) such as "There will be a depression": President Kennedy.)

The element in the seldeps that produces their falsification is not that they are predictions, nor is it their publication.[1] It is their comprehension and the assessment of their effect by listeners with a certain motivation and certain powers which brings about their defeat (perhaps unintentionally). Let us consider a family of examples related to the simple seldep. First example: instead of *telling* the subject exactly what I predict he will do, I so act as to make my predictive belief clear. To use individual cases for examples: as an insurance agent, I turn down, out of hand, the subject's application for automobile collision coverage after reviewing his extremely unsatisfactory record. The cognitive result is the same as if I announce my prediction; I believe, and he knows I believe, that he will probably have another

[1] Statements about the present can be affected in the-same way; in obvious cases. because they take time to utter, for example, "Although there is a sign up there which says 'No Smoking,' several people here *are* smoking." There are other cases, however: "I know you can't bear to live without me." The sequel establishes the dispensability of publication.

serious accident. Thus my belief, a prediction, may lead him to take more care and so prove me and it wrong. This is in many ways the same process that the simple seldep instigates.

Second example: I act in a way that makes it clear that I have made *some* prediction, but not clear *what* prediction it is, for instance, I smirk and say knowingly, "I know what you are going to do about marrying that girl, even if *you* don't." The subject may ignore this and proceed as he would have done anyway. But it may be the case that the importance to him of showing me wrong is greater than any gains he can make by making his choice on the intrinsic merit of the alternatives (a common situation in cards, business, love, and war). I shall say that a subject whose utility-set is under this constraint is contrapredictive, or is contrapredictively motivated, or is a contrapredictive. (This is not at all the same as, though it overlaps with, being countersuggestible.)

Then a *good* strategy for him is to use a randomizer to determine his choice. If there are n alternatives open to him, this makes my chance of predictive success only $1/n$ which is presumably worse than it was. In principle, some randomizers are predictable (for instance, dice), but this is an uninteresting sense of "in principle" for the working scientist. Moreover, we can readily use a quantum randomizer, which is in principle unpredictable (on what I judge to be the best-supported current view of quantum theory). In the absence of access to such devices, it is comparatively easy to invent an *ad hoc* mental or physical randomizing procedure which will select a digit or letter in a way no more predictable than a roulette wheel.

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 (which includes the last class), a stronger conclusion is demonstrable.

Third example: It may be that I do not indicate to the subject that I have made a prediction about his behavior, but that he suspects I may have done so—and is contrapredictively motivated at the time. His best strategy here (and in the preceding case) is to replicate my prediction, if he can. He may already know, or be able to infer, what I know about him; from this he draws any predictive conclusions that are possible. Then, of course, he acts so as to falsify my prediction. This strategy yields a gain in expected utility Δu (over the randomizing strategy) which is of course given by the formula $\Delta u = u/n$, where u is the utility of surprise, and is thus diminishingly important for choices between an increasing number of alternatives.[2]

[2] If the utility of surprise varies from alternative to alternative, being u_i for the i^{th} alternative, $\Delta u = u_i/n$, and may therefore be larger for larger n.

SIGNIFICANCE

I shall try to give a precise statement of the conditions for this strategy in a moment. First we should examine the question whether such a case counts for anything against the predictability of human behavior as it is usually conceived. Psychologists have rarely been naive enough to suppose that one can always announce one's predictions of behavior to the subject of the prediction without subsequent falsification: determinism does not mean this. And surely I am merely pointing out that the same conclusions apply when the subject can work out the prediction without being told it?

But 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; unmakable ones are not.

The behavior of an intelligent contrapredictive with adequate computer resources will never be more predictable than the best randomizer he can get hold of; but it is *absolutely* unpredictable, that is, the available data yields *no* predictive conclusion at all, unless certain very special conditions are met, namely, (i) the contrapredictive incorrectly believes he knows all the relevant data the predictor possesses about him; (ii) this presumed data implies a definite prediction (which *usually* means it does not include the fact of contrapredictive motivation); (iii) the data the predictor *actually* has enables him to predict the subject's behavior under conditions (i) and (ii).

For in any other case, the fact of contrapredictivity automatically nullifies any prediction the remaining data may imply, and also any implied by that fact *and* the other data.

PRECISE FORMULATION

Assume a rational intelligent predictor, P, whose task is to infer from information I_c the choice of an individual C, where

(i) C is choosing rationally and intelligently (that is, so as to maximize expectations of utility) between alternatives $a_1, a_2, \cdots a_n$, $(n \geq 2)$ (that is, is physically capable of each [would do it if he decided to] and must do one).

(ii) C is a contrapredictive relative to P and a_i, that is, wishes to falsify any prediction made by P about his choice. Precisely, if $\bar{u}_i$ is the utility for C of a_i if a_i is predicted by P, and u is the utility for C of picking an unpredicted a, then "C is a contrapredictive" = "$u > [\max \bar{u}_i - \min \bar{u}_i]$."[3]

[3] The formula for the more general case, where u depends on the alternative chosen, is $\min(\bar{u}_i + u_i) > \max \bar{u}_i$. It is possible to regard this as the (.../cont.)

(iii) C knows that $\boldsymbol{I_c}$ is P's data and C has sufficient facilities to calculate any consequences of $\boldsymbol{I_c}$ with respect to C's forthcoming choice, prior to the time he must make the choice.

It immediately follows that $\boldsymbol{I_c}$ either implies an *incorrect* prediction as to which $\boldsymbol{a_i}$ will be chosen by C (that is, contains false information) or *none:* hence C's choice cannot be rationally predicted by P from $\boldsymbol{I_c}$.

For assume the contrary: If C($\boldsymbol{a_m}$) = C will choose $\boldsymbol{a_m}$ (definition) then the assumption gives us: (a) $\boldsymbol{I_c} \rightarrow$ C($\boldsymbol{a_m}$) for some m; (b) C($\boldsymbol{a_m}$) is true. Now C will know that P, since rational and accepting $\boldsymbol{I_c}$, is making the prediction C($\boldsymbol{a_m}$) (from (i) and (iii)). Hence C will in fact choose another alternative $\boldsymbol{a_p}$, $p \neq m$. Hence C($\boldsymbol{a_m}$) is false, contrary to the assumption. Hence either $\boldsymbol{I_c}$ does not imply C($\boldsymbol{a_m}$) for any m or C($\boldsymbol{a_m}$) is false.

QED

COMMENTS AND CLARIFICATIONS

A. "But much behavior is *already* known to be predictable, even when the prediction is known, for example, that of the compulsive, or the rational benevolent man, etc."

Granted: as long as condition (ii) does not apply, the theorem does not apply. (Other unpredictabilities: (i) K. R. Popper has suggested interesting similar results in other cases; (ii) quantum-uncertainty-dependent behavior.)

B. "Something *must* cause the eventual decision in the contrapredictive; and if we had enough knowledge, we would know what it is, and the effect it has."

Something does (in so far as determinism is true), but *what* it is, or rather what *effect* it has, we cannot know *in advance*. We could know it later, and this capacity for explanation is what I take determinism to mean: if it means "inferential predictability in principle," the theorem proves it false (see 4H and SD below).

C. "But *we* can precisely predict C's behavior; C will always do the opposite of what P predicts."

1. The prediction task of the theorem is prediction of the precise alternative which C selects. Of course, for other prediction tasks, this result does not apply. For example, we can always predict that C will pick *some* $\boldsymbol{a_i}$; this is guaranteed by (i). And we can predict with great confidence (for large n) that C will pick an $\boldsymbol{a_i}$ such that $1 \leq i \leq n - 1$; or that he will not pick $\boldsymbol{a_n}$. As the information content goes down, the confidence level goes up, and vice versa. But the theorem tells us we can never achieve *high* confidence in a *highly specific* prediction, which is of course the kind

(*cont.*) definition of strong contrapredictivity and discuss a weaker notion involving the probabilities P_i that P will predict $\boldsymbol{a_i}$.

we would usually prefer to be able to make. If P predicts C(a_m), we as third parties can still only predict "Not C(a_m)," which has vanishingly small information content for increasing n.

2. Moreover, P can't predict *anything* if he knows C is contrapredictive, unless he also knows C to be badly wrong as to what I_c is, so we never even get situation 1 in the important cases.

3. Even if P does predict something in the sense of selecting some a_m without proof, C can still actually do a_m since C knows P can't be rationally *sure* of C(a_m).

D. "The proof assumes C's rationality, which is unrealistic."

1. The objection is irrelevant since the theorem can be taken as a limit theorem for rational methods; one does not raise a serious objection to the third law of thermodynamics (that absolute zero is unattainable) by saying "It practically never gets that cold."

2. If C is likely to be irrational, so is P; hence C is still likely to be unpredicted because P does not fully utilize his data.

3. There are in fact many practical occasions when the degree of rationality required is available and the theorem then represents a relevant practical consideration. .

E. "Stochastic strategies are immune to publicity; hence the theorem only applies to exact prediction."

But C can falsify predictions about the statistical properties of his choice behavior just as easily as individual predictions: for example, if P concludes C will, *r% of the time,* choose an a_m where $1 \leq m \leq n/2$, then C can choose such an a_m just (100 - r)% of the time.

F. "One can't *prove* C is unpredictable; for P may just guess correctly or be a precognitive or prophet."

1. Guessing is not a procedure of predictive inference, since we cannot know when the guess is correct, that is, can have no confidence in predictions generated in this way. The theorem only refers to the impossibility of rational (correct) prediction, predictions in which we can have confidence.

2. If P finds his "guesses" are *significantly* more effective than they would be by chance alone, that is, that he can rationally have confidence in them, he has discovered a new fact about himself *and about* C, that P is a good instrument for predicting C's choices. This will—it could be argued—be part of I_c , hence C will know it. But this alone does not enable C to evade the prediction. Now we can plausibly take reliable intuition as a limit case of inference where the content of the intuition is itself part of the data, and this will then reduce to the case of announcing the prediction to C. Of course, a Helmholtzian would argue that P must be performing an *unconscious* inference. This is a misleading account in some ways, but it suggests an important distinction between supernatural precognition and supernormal prescience. The expert clinician's prognoses are not inferences, nor are

they normal perceptions—but nor are they supernatural. His brain is absorbing information and converting it into a prediction using a learned though not known transformation principle; if we cut him off from perceptual access to the patient or the patient's charts or his own past experience, he will fail. A prognostic computer operated by a medically naive technician has the same dependency on stored information, so we may argue that the clinician and the technician are using this data in getting to their prediction, though not in a process of explicit inference which they themselves perform. I shall henceforth regard this as encompassed by the phraseology of the theorem preamble: "infer from information $\boldsymbol{I_c}$" is to mean "explicitly deduce or infer from data $\boldsymbol{I_c}$ by means of laws, skills, or devices that incorporate transforming principles $\boldsymbol{I_c}$ ($\boldsymbol{I_c = I'_c + I''_c}$)."

The crystal ball, however, operates in a different way. It does not require a data input, merely reflecting the shadows that coming events cast before them. And the same is true of the "true prophet," or the parapsychologist's precognitive. There is an *experimentum crucis* to distinguish the precognition from prescience; it is that used by the parapsychologists: can the agent predict events that are either fully random themselves, or determined by random events? If so, his powers are supernatural. The theorem, thus interpreted, applies to all prediction other than supernatural. As long as P's skill is inexplicable (but not supernatural), $\boldsymbol{I_c}$ will be unknown and hence the theorem cannot be applied since C cannot be in possession of $\boldsymbol{I_c}$. Of course, he may still be able to duplicate P's prediction by finding a matching predictor P′; and he is highly secure in a random strategy. Thus, one consequence of replacing the idea of inferential prediction with that of using information to generate predictions is to put the P with a mysterious skill in the same category as the P with mysterious data; he can succeed only so long as the mystery is continued.

We conclude by stressing that making a rational prediction using a mysterious skill or instrument requires a testing period to establish the reliability of the "instrument," even if it is oneself. Supposing the instrument has a "pointer error," for example, indicates C($\boldsymbol{a_j}$) whenever C actually chooses $\boldsymbol{a_{j+i}}$. This emerges in such a trial period and is no handicap at all: and one might plausibly argue that, pointer error or none, the trial period establishes a correlational law from which together with subsequent pointer readings, the later predictions are *inferred*. If so, law and readings must be regarded as part of $\boldsymbol{I_c}$, and the earlier argument is unnecessary.

3. The above considerations have three interesting incidental consequences. (a) The term "data" should be distinguished from the term "information" since the latter is here used to include laws that are not inferable from data in any systematic way. A computer may have the same data as another and not be able to generate the same predictions because it has not hit upon the same generalizations. In the theorem, $\boldsymbol{I_c}$ includes all information, including any well-confirmed generalizations;

which is essentially required for generating P's predictions. The fact remains that a brilliant computer or psychologist is the best weapon for getting one-up in the prediction war: a new insight is a new predicting algorithm, and like a new theorem, it tells us something we did not know and hence gives us new information, though it is not based on new data. (b) In the philosophical discussion of determinism, that doctrine has frequently been defined as predictability-in-principle. Now, a precognitive can by definition reliably predict anything, including the most random possible sequences, yet I think it unsatisfactory to suppose the existence of such a being would demonstrate determinism (an argument can be given, but it can be made very implausible by experimental ingenuity). I think determinism is much more nearly connected with predictions that use information. It is too narrow to require that the process be explicit: the existence of a prescient as defined above still shows determinism to be true. I give reasons below for thinking even nonsupernatural predictability-in-principle too strong a definition of determinism, but for forward-looking philosophers the present distinction is of some importance. (c) A mysterious predicting gadget may work perfectly during the trial period, but yet must fail thereafter, if $\boldsymbol{I_c}$ includes all the information from which P predicts. For during the trial, its readings are not an adequate basis for prediction and hence are not part of $\boldsymbol{I_c}$: but thereafter, when they would yield good predictions in normal circumstances, they will not under the conditions of the theorem. The instrument, because it works, now must fail. This *ad hoc* way of falsifying predictions is reminiscent of the Maxwell demon's properties, and the peculiar defining characteristics of the Einstein-Podolski-Rosen *Gedankenexperiment.*

G. "A *third* person P′ watching the P-C affair, can—in principle—predict everything that goes on, using the fact of contrapredictivity plus P's predictability."

Not if C is contrapredictive relative to P′ and knows what P′ knows about C, that is, not unless P′ breaks condition (ii) or (iii). And objections from objection C above also apply. Of course, C will have to decide whether making P′ wrong is more important than making P wrong, for $n = 2$; in general, C can only falsify one less predictors than he has alternatives.

H. "None of this really proves unpredictability, became that means the possibility of predictability by *someone*—if necessary, someone who *has* knowledge C doesn't know he has."

But Gödel's theorem is not a proof that a certain specific formulas not provable in any system; it is a proof that for certain common and important types of system, however they are strengthened, there is *at every given stage* an unprovable formula. The present proof is that under certain common and socially important conditions, as well as under certain ideal conditions (see 5D), however much is known about a

subject's behavior, there are certain parts of it that will not be predictable. Increases in knowledge will make predictions possible; but with respect to each such increase in $\boldsymbol{I_c}$, there will be new predictions which will be impossible. It seems not unreasonable to say this is an essential unpredictability in the same way that an essential unprovability is demonstrated by Gödel's proof.

I. "The proof begs the question; everything is packed into the 'free will' assumption that someone ***can*** do the opposite of whatever is covertly predicted of him by another. If we really knew ***all*** about a person, he would ***have*** to do what we predicted—or else we ***wouldn't*** know all about him."

I now undertake to prove the "free will assumption" for my own case. Notice first that it is here only necessary to demonstrate that, no matter what prediction is *announced,* I can do the opposite: for in the case where a prediction is made but not announced, I know what it is by replication (the possibility of which is guaranteed by the conditions) and hence am able to announce it myself, or have it announced to me, without change in the information parameters of the situation. Now we locate a predictor and we assign him the prediction task of saying whether I will turn my head left or right. He does not have very much information about me, but *even if he knew everything that can be known,* he would still have to predict either left or right. *By merely guessing, he has one chance in two of making the same prediction as if he were omniscient.* He makes his guess, and—to compensate for the fact that I would have been able to infer it—announces it. I then do the opposite. Now, if he had really had all the information, it might be that it would have led to the opposite prediction, which I would then have known to be his prediction. So the fact that I prove I can falsify his first prediction does not prove I could have falsified the correct prediction. But now there is only one chance in two of this. Let us continue the experiment, using the same prediction task. Many variables have changed significantly, but once again, there are only two possibilities. The predictor guesses again, or uses a randomizer to select one alternative. Once more I falsify his prediction. There is now only one chance in four he has not, on one of these occasions, made the same prediction as an omniscient psychologist. But I am not frozen by either prediction. We continue, and thus prove to any significance level that I have "free will" in this sense.

Of course, the point is obvious enough on other grounds since it is merely the claim that a man can be so sensitized as to respond with behavior B to the stimulus "You will do A" and conversely, which is a fairly trivial feat. But psychologists sometimes succumb to the "experimented demonstration" more readily.

It can here be noted that since full knowledge cannot lead to a correct prediction of a similarly informed contrapredictive, it cannot lead to any prediction at all (since false conclusions would imply false premises, that is, the "knowledge" would not be

knowledge). Hence, the maximum pos-sible knowledge the predictor can have of C cannot provide grounds for prediction. This is the limit version of the theorem.

J. "The result is easily circumvented since any psychologist knows that an observer can easily learn more about a subject than the subject knows about himself."

1. Knowing more may be easy, but knowing more in a direction that pays off in better predictions is more difficult and often impossible for the particular kind of choice we are trying to predict, for instance, whether to bluff at poker, what deployment of forces an unknown enemy strategist will select, and other nonneurotic cases.

2. What one observer can learn, another may; so the contrapredictive would obviously have to employ his own observers to study his own behavior, from whom he may be able to replicate the predictor's prediction, if any, and act so as to falsify it.

K. "To be realistic, C is quite likely to be wrong in his estimate of what P knows about him, and then the proof doesn't apply."

Such errors are likely to reveal themselves if P acts on his predictions of C, say, in a limited war situation. C then improves his replication or defensively moves to a random strategy. (A rider of this is that if P does acquire covert, predictively effective data about C, he should hold off using it until a very large gain is available by doing so, since he sacrifices later success when he does.) Whenever P has, for example, the results of standard tests done by C, plus public psychological theory about interpretation of these, the proof applies in full force, as to many cases of less data and of more. C can frequently but not always pursue optimum strategy by giving P the benefit of any doubts as to the content $\boldsymbol{I_c}$, and including the doubtful items.

L. "In view of the risk of underestimating the predictor's knowledge, the minimax strategy for C will usually be a wholly random choice between $a_1 \cdots a_n$."

1. It is not only underestimating but overestimating that is risky. A relatively ignorant P may have data which happens to imply that C will do what he actually does do, because C has assumed that P had more knowledge which would have indicated a different choice.

2. C's best strategy will indeed depend on (a) the likelihood of error in his estimate of P's data, and also on (b) the disparity between u and the set $u_1 \cdots u_n$, (c) the relative sizes of the u_i and (d) the size of n. He does take a risk by switching from a fully random strategy; but he generally has a chance of gain by doing so.

3. In general, where $\boldsymbol{I_c}$ includes the fact that C is contrapredictive, it cannot imply any specific prediction, hence errors in estimating $\boldsymbol{I_c}$ by C are unimportant since it has no consequences over this range. P will have to guess, and C knows this. It is no solution for P to weight his guesses with C's utility weights since C can ignore them.

4. This leads to some useful considerations for C. It *is* usually possible for C to structure a situation so as to increase the number of alternatives open to him. By thus increasing n, he increases his expectation of u, (the utility of surprise). This usually gives a net gain, even though the procedure for doing this also usually leads to some decrease in u or in the $\bar{u}_i$. For example, in a war Russia might see a choice of H-bombing New York City or Los Angeles. If all the antimissile missiles available are concentrated in one of these areas, they would have a good chance (75 per cent) of handling the attack—if divided, they almost certainly fail (95 per cent). The U.S. thus has a choice between two defense strategies, one of which offers a 50% × 75% = 37½% chance of success, the other a 50% × 5% = 2½% chance. Now, if the Soviet Union extends its range of choice to include three other major urban areas, the best strategy available to the United States gives only 20% × 75% = 15% chance of success. There is a possible loss to the Russians due to the diminished utility of killing a smaller city, but the gain to them considerably outweighs this if their utility is population-proportioned. The relevant strategic inequalities to decide whether Russia should (a) use a weighted stochastic strategy, (b) increase the number of alternatives further, are easily calculated. The crucial data in this large subfamily of cases is thus simply the set $a_1 \cdots a_n$: C should (i) conceal it or misrepresent it, (ii) enlarge it, (iii) depending on u, u_i, etc., weight it, (iv) draw from it randomly, with or without weights.

5. Thus, in a large subclass of cases, but not all, the extent of the essential unpredictability is very close to being the same as that of the random unpredictability we first observed. But instead of being a way of avoiding prediction the strategy is adopted as a consequence of the impossibility of prediction.

6. One may ask: if prediction is impossible, why doesn't C simply pick the a_i which has maximum $\bar{u}_i$? I_c cannot contain the assertion that C, having seen P's impasse, will definitely select on this principle, because if it does, C's choice would be predictable by P, and this fact being known to C, he would do otherwise. Still, C *may* select on this basis; it has obvious merits. It cannot be definitely predicted that he will; but he may. Similarly for adjacent strategies. Similarly for any fixed ratio between them. Hence P can have no reliable grounds for supposing that C will deviate at all from a pure random choice between the a_i. This is not the same as saying that C will adopt this strategy. The situation is absolutely indeterminate. In practice, one might suppose that C should begin by using the high $\bar{u}_i$ strategies as much as he can until P bets on them and then changing: except that P can foresee this. The stable strategy for C will eventually be a mixture which is slightly biased toward the $\bar{u}_i$ weights by a factor which depends on what might be called P's sensitivity. This is a familiar proposition when applied to bluffing in poker.

7. The general conclusion here I take to be of considerable importance in

assessing models for psychological theories. I think we can establish on this and on other grounds that possibility-statements (here, the datum of the range of alternatives) are more important for psychology than universal statements (that is, exact laws). True and informative statements of the first kind are always more readily and sometimes solely available, and their availability more than compensates for their lack of intrinsic virtue.

M. "In the social sciences, we are rarely dealing with the behavior of a single person, and as soon as we switch to predictions of group behavior, H. A. Simon has shown that it is in principle possible to predict publicly not only the general behavior but the exact percentage of a group that will perform in a certain way, so the theorem has negligible impact."

1. In the social sciences, we are often dealing with the *effects* of the behavior of one or a few persons, the effects occurring to a large number of people.

2. What Simon's very interesting result shows is that *under some circumstances* precise selfups about group behavior are possible. This finding has been inaccurately summarized by himself and others (e.g., Simon, 1957, p. 86; Nagel, 1959, p. 142). There is a wide range of circumstances, including common electoral ones, where his results do not apply. A simple counterexample is this. Suppose that the underdog (or read "bandwagon") effect with respect to a particular candidate in a two-way fight operates in this way: the moment it appears from the announced prediction that he will win, a number of his supporters (or read "opponents") will decide to vote the other way. No predictions can be right in these circumstances unless the underdog or bandwagon motivated group is smaller than the uncontaminated majority would have been. (If both effects operate simultaneously, possibly for each candidate, the inequality must be generalized in an obvious way.) In Simon's terms, this corresponds to a discontinuity in the function giving the actual vote in terms of the predicted vote, at the point where the latter is 50 per cent; and it is highly plausible to suppose that just such a discontinuity exists at times since 50 per cent is the point at which the prediction's support switches from one candidate to the other. Other difficulties arise for Simon in the case of alliances, grapevine information and suspicion of manipulation.

3. Pure contrapredictive motivation is not usually involved in the band-wagon/underdog effects, only the reconsideration of one's own actions in the light of data about other people's actions, e.g., when wanting a candidate to know he has some support while not wanting him to get in, or preferring A to B and C, but being willing to vote for B against C if A has no chance at all, etc. The key to these effects lies in contrafactual belief, rather than contrapredictive motivation: the voters had voted as they did, because they did not think the facts about other people's votes (or the prediction) are as they subsequently turn out to be. It is also important that a

vote can serve more than one purpose: it can help to elect, or it can be used to show support when/because election is impossible.

4. The crucial reason for judging the Simon theorem irrelevant to the present theorem becomes clear when we realize that Simon's prediction succeeds *only* because he has two items of information about the electorate that they lack, namely, the uncontaminated vote and the functional relationship. The first datum is not hard to get, by replicating the poll if bribery fails; the second only a little harder. So a single (rich) contrapredictive in the group can foil the predictions.

5. A systematic study of published predictions referring to groups including some contrapredictives and/or some people with contrafactual beliefs reveals an extraordinary complexity. Some are susceptible to immediate disproof by a single individual's unaided act, others require collusion, others are highly vulnerable to random strategies, etc. No general conclusions emerge readily, but the mining rights look valuable.

6. Study of an apparently rather different kind of example is also illuminating and suggests conclusions about selfups. The family of paradoxical announcements—the paradox of the Class A blackout, the condemned prisoner, or the unpredictable examination (see, e.g., Martin Gardner, 1963) introduces an interesting consideration. If you tell someone you are going to give a party for him the following evening whose occurrence he will not be able to predict, he may conclude that the announcement is self-defeating. But is it? It looks as if he can predict the party from your announcement and hence it cannot be unpredictable. Since the only party your announcement guarantees is an *un*predictable one, it follows there can be no party at all. Having thus concluded that the party's non-occurrence is predictable, which entails that its occurrence is not predictable he is unfortunately vulnerable to the fact that if you give it, it will precisely fulfill your guarantee of an unpredictable party. One of the morals of this example is to stress the peculiar difficulty of inferences from what is announced by P to what is predicted by P: the main theorem of this paper does not involve any such inferences, whereas they are extremely important in this paradox and underdog effect. The secret of the success of Simonized predictions lies in almost the same peccadillo as the above paradox. The underdog voter may know the pollster's announcement is Simonized, but this cannot justify him in ignoring it since the likelihood of his ignoring it has also been taken into account. Of course, collusion can cause trouble in Simon's case and is irrelevant to the paradox.

SPECIAL APPLICATIONS

1. C = computer. The proof demonstrates that physical determinism is either false or does not imply predictability-in-principle of all systems (contra Laplace). The

motivational condition of contrapredictivity is a simple matter to program, and the parity of information easily arranged. K. R. Popper has given a most extensive—and a very interesting—treatment of limitations on determinism that arise from taking seriously the fact that the predicting computer is itself a physical system. His results are related to the present ones, but by no means the same; we develop these by taking seriously the fact that the predictee may be contrapredictively motivated.

In the early discussions of the contrapredictive effect, it was commonly thought to be unique to the social sciences. We have just observed its applicability to computers which might be thought to count against this; but I have argued elsewhere that the social sciences will ultimately include the study of molar computer behavior (Scriven, 1960). However, even as it *was originally defined,* the effect applies throughout science. The prediction that a comet will return on a certain date may provoke a successful effort to intercept and destroy it. What is unique to the social sciences is only that their predictions are often falsified by the action of those to whose behavior the prediction refers (directly, or indirectly as in bank and crop predictions). The contrapredictive effect, however, as I have tried to demonstrate, has some of its most interesting manifestations where no prediction is or can be generated at all; hence it is a more general concept than the self-defeating (published) prediction.

2. C = P. The problem whether one could ever know in advance how one was going to decide is a nice dilemma for determinists, and this aspect of the problem has been treated most illuminatingly by D. M. MacKay in many writings, with somewhat different conclusions from my own. His emphasis has been on the impossibility of the predictee's ***believing*** the predictions about his choices made by any observer including himself; I have discussed the impossibility of the observer even ***inferring*** a prediction under certain (narrower) conditions, a result which is stronger in one way, but more limited in its range (weaker) in another. The emphasis of the discussions by Popper, MacKay and myself is thus ***primarily*** on three kinds of ultimate limitations on predictions; computer limitations, belief limitations, and inference limitations. The existentialists have also sensed the logical indispensability of the act of genuine choice, despite determinism; and of course many others have felt this so acutely as to conclude that determinism must be false. But I do not think MacKay's denial that a single true account is possible, or the existentialist's nihilism, or the libertarian's antideterminism are required. The idea that determinism implies total predictability is, on this and other grounds indefensible and must be rejected. Taking it to mean only the universal rule of some kind of exact laws, it is falsified by quantum actions but not by the present argument (contrary to Paul Shiman's interesting suggestion which stimulated my thinking on this point). In particular, exact ***explanation*** of human and computer actions is perfectly possible in principle

although prediction is not, quantum effects apart. However, it is another question whether it is always worth searching for such explanations, or indeed practically possible to find them. I think the task of the social sciences is largely elsewhere, and (perforce) largely outside the prediction field. We could put the present point by saying that rational contrapredictive behavior is probably perpetually "emergent," that is, generates new "laws"; at any rate, new phenomena. At each stage we have to appeal to a higher element in the hierarchy of laws than any we have yet discovered; yet it can be an exact law, and it can be in turn explained. Atemporal laws are not necessary for explanation.

3. P = God. God can know what we will do only by preventing us from knowing what he knows, or depriving us of "free will" in the sense of the capacity for contrapredictive motivation. Thus his omniscience to some degree limits his exercise of omnipotence or benevolence. The theorem does show that monotheism is the only form of theism reconcilable with omniscience, under certain obvious conditions.

4. I_c = All possible information about C, that is, the ultimate goal of individual psychology. As for the mechanical and the supernatural predictors, so for the human ones. The present proof shows that psychology can never get to the point where its practitioners know enough to predict all the behavior of other psychologists, even if they could predict the behavior of all physical randomizers. Conversely, to insist that *predictive* determinism is true is to insist that it is impossible for some knowledge about C to be conveyed to him: which I take to be an empirical and false claim.

REFERENCES

GARDNER, M. Mathematical games. *Sci. Amer., 1963.*

NAGEL, E. *The structure of science.* Glencoe, Ill.: Free Press, 1959.

SCRIVEN, M. The compleat robot, a prolegomena to androidology. In S. Hook *Dimensions of mind.* New York: New York Univ. Press, 1960.

SIMON, H.A. *Models of man.* New York: Wiley, 1957.