Hist-Analtic would like to thank the Editor of the Proceedings of the Aristotelian Society for permission to webpublish Russell's paper.

www.aristoteliansociety.org.uk

HERBERT FEIGL and MAY BRODBECK
Editors
UNIVERSITY OF MINNESOTA

Readings in
THE PHILOSOPHY OF SCIENCE

New York
APPLETON-CENTURY-CROFTS, INC.

On the Notion of Cause, with Applications to the Free-Will Problem*

BERTRAND RUSSELL

IN THE FOLLOWING paper I wish, first, to maintain that the word "cause" is so inextricably bound up with misleading associations as to make its complete extrusion from the philosophical vocabulary desirable; secondly, to inquire what principle, if any, is employed in science in place of the supposed "law of causality" which philosophers imagine to be employed; thirdly, to exhibit certain confusions, especially in regard to teleology and determinism, which appear to me to be connected with erroneous notions as to causality.

All philosophers, of every school, imagine that causation is one of the fundamental axioms or postulates of science, yet, oddly enough, in advanced sciences such as gravitational astronomy, the word "cause" never occurs. Dr. James Ward, in his *Naturalism and Agnosticism*, makes this a ground of complaint against physics: the business of those who wish to ascertain the ultimate truth about the world, he apparently thinks, should be the discovery of causes, yet physics never even seeks them. To me it seems that philosophy ought not to assume such legislative functions, and that the reason why physics has ceased to look for causes is that, in fact, there are no such things. The law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm.

In order to find out what philosophers commonly understand by "cause," I consulted Baldwin's *Dictionary*, and was rewarded beyond my expectations, for I found the following three mutually incompatible definitions:—

> "CAUSALITY. (1) The necessary connection of events in the time-series. . . .
>
> "CAUSE (notion of). Whatever may be included in the thought or perception of a process as taking place in consequence of another process. . . .

* Reprinted by kind permission of the author and the publishers from Bertrand Russell, *Mysticism and Logic* (George Allen & Unwin, Ltd.), pp. 180–205, and his *Our Knowledge of the External World* (W. W. Norton & Co., London, 1929, New York, 1929), pp. 247–256.

"CAUSE AND EFFECT. (1) Cause and effect . . . are correlative terms denoting any two distinguishable things, phases, or aspects of reality, which are so related to each other that whenever the first ceases to exist the second comes into existence immediately after, and whenever the second comes into existence the first has ceased to exist immediately before."

Let us consider these three definitions in turn. The first, obviously, is unintelligible without a definition of "necessary." Under this head, Baldwin's *Dictionary* gives the following:—

"NECESSARY. That is necessary which not only is true, but would be true under all circumstances. Something more than brute compulsion is, therefore, involved in the conception; there is a general law under which the thing takes place."

The notion of cause is so intimately connected with that of necessity that it will be no digression to linger over the above definition, with a view to discovering, if possible, *some* meaning of which it is capable; for, as it stands, it is very far from having any definite signification.

The first point to notice is that, if any meaning is to be given to the phrase "would be true under all circumstances," the subject of it must be a propositional function, not a proposition.[1] A proposition is simply true or false, and that ends the matter: there can be no question of "circumstances." "Charles I's head was cut off" is just as true in summer as in winter, on Sundays as on Mondays. Thus when it is worth saying that something "would be true under all circumstances," the something in question must be a propositional function, i.e. an expression containing a variable, and becoming a proposition when a value is assigned to the variable; the varying "circumstances" alluded to are then the different values of which the variable is capable. Thus if "necessary" means "what is true under all circumstances," then "if x is a man, x is mortal" is necessary, because it is true for any possible value of x. Thus we should be led to the following definition:—

"NECESSARY is a predicate of a propositional function, meaning that it is true for all possible values of its argument or arguments."

Unfortunately, however, the definition in Baldwin's *Dictionary* says that what is necessary is not only "true under all circumstances" but is also "true." Now these two are incompatible. Only propositions can be "true," and only propositional functions can be "true under all circumstances." Hence the definition as it stands is nonsense. What is meant seems to be this: "A proposition as necessary when it is a value of a propositional function which is true under all circumstances, i.e. for all values of its argument or arguments." But if we adopt this definition, the same prop-

[1] A propositional function is an expression containing a variable, or undetermined constituent, and becoming a proposition as soon as a definite value is assigned to the variable. Examples are: "A is A," "x is a number." The variable is called the *argument* of the function.

osition will be necessary or contingent according as we choose one or other of its terms as the argument to our propositional function. For example, "if Socrates is a man, Socrates is mortal," is necessary if Socrates is chosen as argument, but not if *man* or *mortal* is chosen. Again, "if Socrates is a man, Plato is mortal," will be necessary if either Socrates or *man* is chosen as argument, but not if Plato or *mortal* is chosen. However, this difficulty can be overcome by specifying the constituent which is to be regarded as argument, and we thus arrive at the following definition:

"A proposition is *necessary* with respect to a given constituent if it remains true when that constituent is altered in any way compatible with the proposition remaining significant."

We may now apply this definition to the definition of causality quoted above. It is obvious that the argument must be the time at which the earlier event occurs. Thus an instance of causality will be such as: "If the event e_1 occurs at the time t_1, it will be followed by the event e_2." This proposition is intended to be necessary with respect to t_1, i.e. to remain true however t_1 may be viewed. Causality, as a universal law, will then be the following: "Given any event e_1, there is an event e_2 such that, whenever e_1 occurs, e_2 occurs later." But before this can be considered precise, we must specify how much later e_2 is to occur. Thus the principle becomes:—

"Given any event, e_1, there is an event e_2 and a time interval τ such that, whenever e_1 occurs, e_2 follows after an interval τ."

I am not concerned as yet to consider whether this law is true or false. For the present, I am merely concerned to discover what the law of causality is supposed to be. I pass, therefore, to the other definitions quoted above.

The second definition need not detain us long, for two reasons. First, because it is psychological: not the "thought or perception" of a process, but the process itself, must be what concerns us in considering causality. Secondly, because it is circular: in speaking of a process as "taking place in consequence of" another process, it introduces the very notion of cause which was to be defined.

The third definition is by far the most precise; indeed as regards clearness it leaves nothing to be desired. But a great difficulty is caused by the temporal contiguity of cause and effect which the definition asserts. No two instants are contiguous, since the time-series is compact; hence either the cause or the effect or both must, if the definition is correct, endure for a finite time; indeed, by the wording of the definition it is plain that both are assumed to endure for a finite time. But then we are faced with a dilemma: if the cause is a process involving change within itself, we shall require (if causality is universal) causal relations between its earlier and later parts; moreover, it would seem that only the later parts can be relevant to the effect, since the earlier parts are not contiguous to the effect, and therefore (by the definition) cannot influence the effect. Thus we shall be led to diminish the duration of the cause without limit, and

however much we may diminish it, there will still remain an earlier part which might be altered without altering the effect, so that the true cause, as defined, will not have been reached, for it will be observed that the definition excludes plurality of causes. If, on the other hand, the cause is purely static, involving no change within itself, then, in the first place, no such cause is to be found in nature, and in the second place, it seems strange —too strange to be accepted, in spite of bare logical possibility—that the cause, after existing placidly for some time, should suddenly explode into the effect, when it might just as well have done so at any earlier time, or have gone on unchanged without producing its effect. This dilemma, therefore, is fatal to the view that cause and effect can be contiguous in time; if there are causes and effects, they must be separated by a finite time-interval τ, as was assumed in the above interpretation of the first definition.

What is essentially the same statement of the law of causality as the one elicited above from the first of Baldwin's definitions is given by other philosophers. Thus John Stuart Mill says:—

"The Law of Causation, the recognition of which is the main pillar of inductive science, is but the familiar truth, that invariability of succession is found by observation to obtain between every fact in nature and some other fact which has preceded it." [2]

And Bergson, who has rightly perceived that the law as stated by philosophers is worthless, nevertheless continues to suppose that it is used in science. Thus he says:—

"Now, it is argued, this law [the law of causality] means that every phenomenon is determined by its conditions, or, in other words, that the same causes produce the same effects." [3]

And again:—

"We perceive physical phenomena, and these phenomena obey laws. This means: (1) That phenomena *a*, *b*, *c*, *d*, previously perceived, can occur again in the same shape; (2) that a certain phenomenon P, which appeared after the conditions *a*, *b*, *c*, *d*, and after these conditions only, will not fail to recur as soon as the same conditions are again present." [4]

A great part of Bergson's attack on science rests on the assumption that it employs this principle. In fact, it employs no such principle, but philosophers—even Bergson—are too apt to take their views on science from each other, not from science. As to what the principle is, there is a fair consensus among philosophers of different schools. There are, however, a number of difficulties which at once arise. I omit the question of plurality of causes for the present, since other graver questions have to be considered. Two of these, which are forced on our attention by the above statement of the law, are the following:—

[2] *Logic*, Bk. III, Chap. V, §2.
[3] *Time and Free Will*, p. 199.
[4] *Time and Free Will*, p. 202.

(1) What is meant by an "event"?

(2) How long may the time-interval be between cause and effect?

(1) An "*event*," in the statement of the law, is obviously intended to be something that is likely to recur, since otherwise the law becomes trivial. It follows that an "event" is not a particular, but some universal of which there may be many instances. It follows also that an "event" must be something short of the whole state of the universe, since it is highly improbable that this will recur. What is meant by an "event" is something like striking a match, or dropping a penny into the slot of an automatic machine. If such an event is to recur, it must not be defined too narrowly: we must not state with what degree of force the match is to be struck, nor what is to be the temperature of the penny. For if such considerations were relevant, our "event" would occur at most once, and the law would cease to give information. An "event," then, is a universal defined sufficiently widely to admit of many particular occurrences in time being instances of it.

(2) The next question concerns the time-interval. Philosophers, no doubt, think of cause and effect as contiguous in time, but this, for reasons already given, is impossible. Hence, since there are no infinitesimal time-intervals, there must be some finite lapse of time τ between cause and effect. This, however, at once raises insuperable difficulties. However short we make the interval τ, something may happen during this interval which prevents the expected result. I put my penny in the slot, but before I can draw out my ticket there is an earthquake which upsets the machine and my calculations. In order to be sure of the expected effect, we must know that there is nothing in the environment to interfere with it. But this means that the supposed cause is not, by itself, adequate to insure the effect. And as soon as we include the environment, the probability of repetition is diminished, until at last, when the whole environment is included, the probability of repetition becomes almost *nil*.

In spite of these difficulties, it must, of course, be admitted that many fairly dependable regularities of sequence occur in daily life. It is these regularities that have suggested the supposed law of causality; where they are found to fail, it is thought that a better formulation could have been found which would have never failed. I am far from denying that there may be such sequences which in fact never do fail. It may be that there will never be an exception to the rule that when a stone of more than a certain mass, moving with more than a certain velocity, comes in contact with a pane of glass of less than a certain thickness, the glass breaks. I also do not deny that the observation of such regularities, even when they are not without exceptions, is useful in the infancy of a science: the observation that unsupported bodies in air usually fall was a stage on the way to the law of gravitation. What I deny is that science assumes the existence of invariable uniformities of sequence of this kind, or that it aims at discovering them. All such uniformities, as we saw, depend upon a certain

vagueness in the definition of the "events." That bodies fall is a vague qualitative statement; science wishes to know how fast they fall. This depends upon the shape of the bodies and the density of the air. It is true that there is more nearly uniformity when they fall in a vacuum; so far as Galileo could observe, the uniformity is then complete. But later it appeared that even there the latitude made a difference, and the altitude. Theoretically, the position of the sun and moon must make a difference. In short, every advance in a science takes us farther away from the crude uniformities which are first observed, into greater differentiation of antecedent and consequent, and into a continually wider circle of antecedents recognised as relevant.

The principle "same cause, same effect," which philosophers imagine to be vital to science, is therefore utterly otiose. As soon as the antecedents have been given sufficiently fully to enable the consequent to be calculated with some exactitude, the antecedents have become so complicated that it is very unlikely they will ever recur. Hence, if this were the principle involved, science would remain utterly sterile.

The importance of these considerations lies partly in the fact that they lead to a more correct account of scientific procedure, partly in the fact that they remove the analogy with human volition which makes the conception of cause such a fruitful source of fallacies. The latter point will become clearer by the help of some illustrations. For this purpose I shall consider a few maxims which have played a great part in the history of philosophy.

(1) "Cause and effect must more or less resemble each other." This principle was prominent in the philosophy of occasionalism, and is still by no means extinct. It is still often thought, for example, that mind could not have grown up in a universe which previously contained nothing mental, and one ground for this belief is that matter is too dissimilar from mind to have been able to cause it. Or, more particularly, what are termed the nobler parts of our nature are supposed to be inexplicable, unless the universe always contained something at least equally noble which could cause them. All such views seem to depend upon assuming some unduly simplified law of causality; for, in any legitimate sense of "cause" and "effect," science seems to show that they are usually very widely dissimilar, the "cause" being, in fact, two states of the whole universe, and the "effect" some particular event.

(2) "Cause is analogous to volition, since there must be an intelligible *nexus* between cause and effect." This maxim is, I think, often unconsciously in the imaginations of philosophers who would reject it when explicitly stated. It is probably operative in the view we have just been considering, that mind could not have resulted from a purely material world. I do not profess to know what is meant by "intelligible"; it seems to mean "familiar to imagination." Nothing is less "intelligible," in any other sense, than the connection between an act of will and its fulfilment. But

obviously the sort of nexus desired between cause and effect is such as could only hold between the "events" which the supposed law of causality contemplates; the laws which replace causality in such a science as physics leave no room for any two events between which a nexus could be sought.

(3) "The cause *compels* the effect in some sense in which the effect does not compel the cause." This belief seems largely operative in the dislike of determinism; but, as a matter of fact, it is connected with our second maxim, and falls as soon as that is abandoned. We may define "compulsion" as follows: "Any set of circumstances is said to compel A when A desires to do something which the circumstances prevent, or to abstain from something which the circumstances cause." This presupposes that some meaning has been found for the word "cause"—a point to which I shall return later. What I want to make clear at present is that compulsion is a very complex notion, involving thwarted desire. So long as a person does what he wishes to do, there is no compulsion, however much his wishes may be calculable by the help of earlier events. And where desire does not come in, there can be no question of compulsion. Hence it is, in general, misleading to regard the cause as compelling the effect.

A vaguer form of the same maxim substitutes the word "determine" for the word "compel"; we are told that the cause *determines* the effect in a sense in which the effect does not *determine* the cause. It is not quite clear what is meant by "determining"; the only precise sense, so far as I know, is that of a function or one-many relation. If we admit plurality of causes, but not of effects, that is, if we suppose that, given the cause, the effect must be such and such, but, given the effect, the cause may have been one of many alternatives, then we may say that the cause determines the effect, but not the effect the cause. Plurality of causes, however, results only from conceiving the effect vaguely and narrowly and the cause precisely and widely. Many antecedents may "cause" a man's death, because his death is vague and narrow. But if we adopt the opposite course, taking as the "cause" the drinking of a dose of arsenic, and as the "effect" the whole state of the world five minutes later, we shall have plurality of effects instead of plurality of causes. Thus the supposed lack of symmetry between "cause" and "effect" is illusory.

(4) "A cause cannot operate when it has ceased to exist, because what has ceased to exist is nothing." This is a common maxim, and a still more common unexpressed prejudice. It has, I fancy, a good deal to do with the attractiveness of Bergson's "*durée*": since the past has effects now, it must still exist in some sense. The mistake in this maxim consists in the supposition that causes "operate" at all. A volition "operates" when what it wills takes place; but nothing can operate except a volition. The belief that causes "operate" results from assimilating them, consciously or unconsciously, to volitions. We have already seen that, if there are causes at all, they must be separated by a finite interval of time from their effects and thus cause their effects after they have ceased to exist.

It may be objected to the above definition of a volition "operating" that it only operates when it "causes" what it wills, not when it merely happens to be followed by what it wills. This certainly represents the usual view of what is meant by a volition "operating," but as it involves the very view of causation which we are engaged in combating, it is not open to us as a definition. We may say that a volition "operates" when there is some law in virtue of which a similar volition in rather similar circumstances will usually be followed by what it wills. But this is a vague conception, and introduces ideas which we have not yet considered. What is chiefly important to notice is that the usual notion of "operating" is not open to us if we reject, as I contend that we should, the usual notion of causation.

(5) "A cause cannot operate except where it is." This maxim is very widespread; it was urged against Newton, and has remained a source of prejudice against "action at a distance." In philosophy it has led to a denial of transient action, and thence to monism or Leibnizian monadism. Like the analogous maxim concerning temporal contiguity, it rests upon the assumption that causes "operate," i.e. that they are in some obscure way analogous to volitions. And, as in the case of temporal contiguity, the inferences drawn from this maxim are wholly groundless.

I return now to the question, What law or laws can be found to take the place of the supposed law of causality?

First, without passing beyond such uniformities of sequence as are contemplated by the traditional law, we may admit that, if any such sequence has been observed in a great many cases, and has never been found to fail, there is an inductive probability that it will be found to hold in future cases. If stones have hitherto been found to break windows, it is probable that they will continue to do so. This, of course, assumes the inductive principle, of which the truth may reasonably be questioned; but as this principle is not our present concern, I shall in this discussion treat it as indubitable. We may then say, in the case of any such frequently observed sequence, that the earlier event is the *cause* and the later event the *effect*.

Several considerations, however, make such special sequences very different from the traditional relation of cause and effect. In the first place, the sequence, in any hitherto unobserved instance, is no more than probable, whereas the relation of cause and effect was supposed to be necessary. I do not mean by this merely that we are not sure of having discovered a true case of cause and effect; I mean that, even when we have a case of cause and effect in our present sense, all that is meant is that on grounds of observation, it is probable that when one occurs the other will also occur. Thus in our present sense, A may be the cause of B even if there actually are cases where B does not follow A. Striking a match will be the cause of its igniting, in spite of the fact that some matches are damp and fail to ignite.

In the second place, it will not be assumed that *every* event has some antecedent which is its cause in this sense; we shall only believe in causal sequences where we find them, without any presumption that they always are to be found.

In the third place, *any* case of sufficiently frequent sequence will be causal in our present sense; for example, we shall not refuse to say that night is the cause of day. Our repugnance to saying this arises from the ease with which we can imagine the sequence to fail, but owing to the fact that cause and effect must be separated by a finite interval of time, *any* such sequence *might* fail through the interposition of other circumstances in the interval. Mill, discussing this instance of night and day, says:—

"It is necessary to our using the word cause, that we should believe not only that the antecedent always *has* been followed by the consequent, but that as long as the present constitution of things endures, it always *will* be so." [5]

In this sense, we shall have to give up the hope of finding causal laws such as Mill contemplated; any causal sequence which we have observed may at any moment be falsified without a falsification of any laws of the kind that the more advanced sciences aim at establishing.

In the fourth place, such laws of probable sequence, though useful in daily life and in the infancy of a science, tend to be displaced by quite different laws as soon as a science is successful. The law of gravitation will illustrate what occurs in any advanced science. In the motions of mutually gravitating bodies, there is nothing that can be called a cause, and nothing that can be called an effect; there is merely a formula. Certain differential equations can be found, which hold at every instant for every particle of the system, and which, given the configuration and velocities at one instant, or the configurations at two instants, render the configuration at any other earlier or later instant theoretically calculable. That is to say, the configuration at any instant is a function of that instant and the configurations at two given instants. This statement holds throughout physics, and not only in the special case of gravitation. But there is nothing that could be properly called "cause" and nothing that could be properly called "effect" in such a system.

No doubt the reason why the old "law of causality" has so long continued to pervade the books of philosophers is simply that the idea of a function is unfamiliar to most of them, and therefore they seek an unduly simplified statement. There is no question of repetitions of the "same" cause producing the "same" effect; it is not in any sameness of causes and effects that the constancy of scientific law consists, but in sameness of relations. And even "sameness of relations" is too simple a phrase; "sameness of differential equations" is the only correct phrase. It is impossible to state this accurately in non-mathematical language; the nearest approach would be as follows: "There is a constant relation between the state of the

[5] *Loc. cit.*, §6.

universe at any instant and the rate of change in the rate at which any part of the universe is changing at that instant, and this relation is many-one, i.e. such that the rate of change in the rate of change is determinate when the state of the universe is given." If the "law of causality" is to be something actually discoverable in the practice of science, the above proposition has a better right to the name than any "law of causality" to be found in the books of philosophers.

In regard to the above principle, several observations must be made—

(1) No one can pretend that the above principle is *a priori* or self-evident or a "necessity of thought." Nor is it, in any sense, a premiss of science: it is an empirical generalisation from a number of laws which are themselves empirical generalisations.

(2) The law makes no difference between past and future: the future "determines" the past in exactly the same sense in which the past "determines" the future. The word "determine," here, has a purely logical significance: a certain number of variables "determine" another variable if that other variable is a function of them.

(3) The law will not be empirically verifiable unless the course of events within some sufficiently small volume will be approximately the same in any two states of the universe which only differ in regard to what is at a considerable distance from the small volume in question. For example, motions of planets in the solar system must be approximately the same however the fixed stars may be distributed, provided that all the fixed stars are very much farther from the sun than the planets are. If gravitation varied directly as the distance, so that the most remote stars made the most difference to the motions of the planets, the world might be just as regular and just as much subject to mathematical laws as it is at present, but we could never discover the fact.

(4) Although the old "law of causality" is not assumed by science, something which we may call the "uniformity of nature" is assumed, or rather is accepted on inductive grounds. The uniformity of nature does not assert the trivial principle "same cause, same effect," but the principle of the permanence of laws. That is to say, when a law exhibiting, e.g. an acceleration as a function of the configuration has been found to hold throughout the observable past, it is expected that it will continue to hold in the future, or that, if it does not itself hold, there is some other law, agreeing with the supposed law as regards the past, which will hold for the future. The ground of this principle is simply the inductive ground that it has been found to be true in very many instances; hence the principle cannot be considered certain, but only probable to a degree which cannot be accurately estimated.

The uniformity of nature, in the above sense, although it is assumed in the practice of science, must not, in its generality, be regarded as a kind of major premiss, without which all scientific reasoning would be in error. The assumption that *all* laws of nature are permanent has, of course, less

probability than the assumption that this or that particular law is permanent; and the assumption that a particular law is permanent for all time has less probability than the assumption that it will be valid up to such and such a date. Science, in any given case, will assume what the case requires, but no more. In constructing the *Nautical Almanac* for 1915 it will assume that the law of gravitation will remain true up to the end of that year; but it will make no assumption as to 1916 until it comes to the next volume of the almanac. This procedure is, of course, dictated by the fact that the uniformity of nature is not known *a priori*, but is an empirical generalisation, like "all men are mortal." In all such cases, it is better to argue immediately from the given particular instances to the new instance, than to argue by way of a major premiss; the conclusion is only probable in either case, but acquires a higher probability by the former method than by the latter.

In all science we have to distinguish two sorts of laws: first, those that are empirically verifiable but probably only approximate; secondly, those that are not verifiable, but may be exact. The law of gravitation, for example, in its applications to the solar system, is only empirically verifiable when it is assumed that matter outside the solar system may be ignored for such purposes; we believe this to be only approximately true, but we cannot empirically verify the law of universal gravitation which we believe to be exact. This point is very important in connection with what we may call "relatively isolated systems." These may be defined as follows:—

A system relatively isolated during a given period is one which, within some assignable margin of error, will behave in the same way throughout that period, however the rest of the universe may be constituted.

A system may be called "practically isolated" during a given period if, although there *might* be states of the rest of the universe which would produce more than the assigned margin of error, there is reason to believe that such states do not in fact occur.

Strictly speaking, we ought to specify the respect in which the system is relatively isolated. For example, the earth is relatively isolated as regards falling bodies, but not as regards tides; it is *practically* isolated as regards economic phenomena, although, if Jevons' sunspot theory of commercial crises had been true, it would not have been even practically isolated in this respect.

It will be observed that we cannot prove in advance that a system is isolated. This will be inferred from the observed fact that approximate uniformities can be stated for this system alone. If the complete laws for the whole universe were known, the isolation of a system could be deduced from them; assuming, for example, the law of universal gravitation, the practical isolation of the solar system in this respect can be deduced by the help of the fact that there is very little matter in its neighbourhood. But it should be observed that isolated systems are only important as providing a possibility of *discovering* scientific laws; they have no theoretical importance in the finished structure of a science.

The case where one event A is said to "cause" another event B, which philosophers take as fundamental, is really only the most simplified instance of a practically isolated system. It may happen that, as a result of general scientific laws, whenever A occurs throughout a certain period, it is followed by B; in that case, A and B form a system which is practically isolated throughout that period. It is, however, to be regarded as a piece of good fortune if this occurs; it will always be due to special circumstances, and would not have been true if the rest of the universe had been different though subject to the same laws.

The essential function which causality has been supposed to perform is the possibility of inferring the future from the past, or, more generally, events at any time from events at certain assigned times. Any system in which such inference is possible may be called a "deterministic" system. We may define a deterministic system as follows:—

A system is said to be "deterministic" when, given certain data, e_1, e_2, . . . , e_n, at times t_1, t_2, . . . , t_n respectively, concerning this system, if E_t is the state of the system at any time t, there is a functional relation of the form

$$E_t = f(e_1, t_1, e_2, t_2, \ldots, e_n, t_n, t). \qquad (A)$$

The system will be "deterministic throughout a given period" if t, in the above formula, may be any time within that period, though outside that period the formula may be no longer true. If the universe, as a whole, is such a system, determinism is true of the universe; if not, not. A system which is part of a deterministic system I shall call "determined"; one which is not part of any such system I shall call "capricious."

The events e_1, e_2, . . . , e_n I shall call "determinants" of the system. It is to be observed that a system which has one set of determinants will in general have many. In the case of the motions of the planets, for example, the configurations of the solar system at any two given times will be determinants.

We may take another illustration from the hypothesis of psycho-physical parallelism. Let us assume, for the purposes of this illustration, that to a given state of brain a given state of mind always corresponds, and vice versa, i.e. that there is a one-one relation between them, so that each is a function of the other. We may also assume, what is practically certain, that to a given state of a certain brain a given state of the whole material universe corresponds, since it is highly improbable that a given brain is ever twice in exactly the same state. Hence there will be a one-one relation between the state of a given person's mind and the state of the whole material universe. It follows that, if n states of the material universe are determinants of the material universe, then n states of a given man's mind are determinants of the whole material and mental universe—assuming, that is to say, that psycho-physical parallelism is true.

The above illustration is important in connection with a certain con-

fusion which seems to have beset those who have philosophised on the relation of mind and matter. It is often thought that, if the state of the mind is determinate when the state of the brain is given, and if the material world forms a deterministic system, then mind is "subject" to matter in some sense in which matter is not "subject" to mind. But if the state of the brain is also determinate when the state of the mind is given, it must be exactly as true to regard matter as subject to mind as it would be to regard mind as subject to matter. We could, theoretically, work out the history of mind without ever mentioning matter, and then, at the end, deduce that matter must meanwhile have gone through the corresponding history. It is true that if the relation of brain to mind were many-one, not one-one, there would be a one-sided dependence of mind on brain, while conversely, if the relation were one-many, as Bergson supposes, there would be a one-sided dependence of brain on mind. But the dependence involved is, in any case, only logical; it does not mean that we shall be compelled to do things we desire not to do, which is what people instinctively imagine it to mean.

As another illustration we may take the case of mechanism and teleology. A system may be defined as "mechanical" when it has a set of determinants that are purely material, such as the positions of certain pieces of matter at certain times. It is an open question whether the world of mind and matter, as we know it, is a mechanical system or not; let us suppose, for the sake of argument, that it is a mechanical system. This supposition—so I contend—throws no light whatever on the question whether the universe is or is not a "teleological" system. It is difficult to define accurately what is meant by a "teleological" system, but the argument is not much affected by the particular definition we adopt. Broadly, a teleological system is one in which purposes are realised, i.e. in which certain desires—those that are deeper or nobler or more fundamental or more universal or what not—are followed by their realisation. Now the fact—if it be a fact—that the universe is mechanical has no bearing whatever on the question whether it is teleological in the above sense. There might be a mechanical system in which all wishes were realised, and there might be one in which all wishes were thwarted. The question whether, or how far, our actual world is teleological, cannot, therefore, be settled by proving that it is mechanical, and the desire that it should be teleological is no ground for wishing it to be not mechanical.

There is, in all these questions, a very great difficulty in avoiding confusion between what we can infer and what is in fact determined. Let us consider, for a moment, the various senses in which the future may be "determined." There is one sense—and a very important one—in which it is determined quite independently of scientific laws, namely, the sense that it will be what it will be. We all regard the past as determined simply by the fact that it has happened; but for the accident that memory works backward and not forward, we should regard the future as equally de-

termined by the fact that it will happen. "But," we are told, "you cannot alter the past, while you can to some extent alter the future." This view seems to me to rest upon just those errors in regard to causation which it has been my object to remove. You cannot make the past other than it was—true, but this is a mere application of the law of contradiction. If you already know what the past was, obviously it is useless to wish it different. But also you cannot make the future other than it will be; this again is an application of the law of contradiction. And if you happen to know the future—e.g. in the case of a forthcoming eclipse—it is just as useless to wish it different as to wish the past different. "But," it will be rejoined, "our wishes can *cause* the future, sometimes, to be different from what it would be if they did not exist, and they can have no such effect upon the past." This, again, is a mere tautology. An effect being *defined* as something subsequent to its cause, obviously we can have no *effect* upon the past. But that does not mean that the past would not have been different if our present wishes had been different. Obviously, our present wishes are conditioned by the past, and therefore could not have been different unless the past had been different; therefore, if our present wishes were different, the past would be different. Of course, the past cannot be different from what it was, but no more can our present wishes be different from what they are; this again is merely the law of contradiction. The facts seem to be merely (1) that wishing generally depends upon ignorance, and is therefore commoner in regard to the future than in regard to the past; (2) that where a wish concerns the future, it and its realisation very often form a "practically independent system," i.e. many wishes regarding the future are realised. But there seems no doubt that the main difference in our feelings arises from the accidental fact that the past but not the future can be known by memory.

Although the sense of "determined" in which the future is determined by the mere fact that it will be what it will be is sufficient (at least so it seems to me) to refute some opponents of determinism, notably M. Bergson and the pragmatists, yet it is not what most people have in mind when they speak of the future as determined. What they have in mind is a formula by means of which the future can be exhibited, and at least theoretically calculated, as a function of the past. But at this point we meet with a great difficulty, which besets what has been said above about deterministic systems, as well as what is said by others.

If formulæ of any degree of complexity, however great, are admitted, it would seem that any system, whose state at a given moment is a function of certain measurable quantities, *must* be a deterministic system. Let us consider, in illustration, a single material particle, whose co-ordinates at time t are x_t, y_t, z_t. Then, however the particle moves, there must be, theoretically, functions f_1, f_2, f_3, such that

$$x_t = f_1(t), \qquad y_t = f_2(t), \qquad z_t = f_3(t).$$

It follows that, theoretically, the whole state of the material universe at time t must be capable of being exhibited as a function of t. Hence our universe will be deterministic in the sense defined above. But if this be true, no information is conveyed about the universe in stating that it is deterministic. It is true that the formulæ involved may be of strictly infinite complexity, and therefore not practically capable of being written down or apprehended. But except from the point of view of our knowledge, this might seem to be a detail: in itself, if the above considerations are sound, the material universe *must* be deterministic, *must* be subject to laws.

This, however, is plainly not what was intended. The difference between this view and the view intended may be seen as follows. Given some formula which fits the facts hitherto—say the law of gravitation—there will be an infinite number of other formulæ not empirically distinguishable from it in the past, but diverging from it more and more in the future. Hence, even assuming that there are persistent laws, we shall have no reason for assuming that the law of the inverse square will hold in future; it may be some other hitherto indistinguishable law that will hold. We cannot say that *every* law which has held hitherto must hold in the future, because past facts which obey one law will also obey others, hitherto indistinguishable but diverging in future. Hence there must, at every moment, be laws hitherto unbroken which are now broken for the first time. What science does, in fact, is to select the *simplest* formula that will fit the facts. But this, quite obviously, is merely a methodological precept, not a law of Nature. If the simplest formula ceases, after a time, to be applicable, the simplest formula that remains applicable is selected, and science has no sense that an axiom has been falsified. We are thus left with the brute fact that, in many departments of science, quite simple laws have hitherto been found to hold. This fact cannot be regarded as having any *a priori* ground, nor can it be used to support inductively the opinion that the same laws will continue; for at every moment laws hitherto true are being falsified, though in the advanced sciences these laws are less simple than those that have remained true. Moreover it would be fallacious to argue inductively from the state of the advanced sciences to the future state of the others, for it may well be that the advanced sciences are advanced simply because, hitherto, their subject-matter has obeyed simple and easily ascertainable laws, while the subject-matter of other sciences has not done so.

The difficulty we have been considering seems to be met partly, if not wholly, by the principle that the *time* must not enter explicitly into our formulæ. All mechanical laws exhibit acceleration as a function of configuration, not of configuration and time jointly; and this principle of the irrelevance of the time may be extended to all scientific laws. In fact we might interpret the "uniformity of nature" as meaning just this, that no scientific law involves the time as an argument, unless, of course, it is given in an integrated form, in which case *lapse* of time, though not absolute time, may appear in our formulæ. Whether this consideration suffices to

overcome our difficulty completely, I do not know; but in any case it does much to diminish it.

It will serve to illustrate what has been said if we apply it to the question of free will.

The problem of free will * is so intimately bound up with the analysis of causation that, old as it is, we need not despair of obtaining new light on it by the help of new views on the notion of cause. The free-will problem has, at one time or another, stirred men's passions profoundly, and the fear that the will might not be free has been to some men a source of great unhappiness. I believe that, under the influence of a cool analysis, the doubtful questions involved will be found to have no such emotional importance as is sometimes thought, since the disagreeable consequences supposed to flow from a denial of free will do not flow from this denial in any form in which there is reason to make it. It is not, however, on this account chiefly that I wish to discuss this problem, but rather because it affords a good example of the clarifying effect of analysis and of the interminable controversies which may result from its neglect.

Let us first try to discover what it is we really desire when we desire free will. Some of our reasons for desiring free will are profound, some trivial. To begin with the former: we do not wish to feel ourselves in the hands of fate, so that, however much we may desire to will one thing, we may nevertheless be compelled by an outside force to will another. We do not wish to think that, however much we may desire to act well, heredity and surroundings may force us into acting ill. We wish to feel that, in cases of doubt, our choice is momentous and lies within our power. Besides these desires, which are worthy of all respect, we have, however, others not so respectable, which equally make us desire free will. We do not like to think that other people, if they knew enough, could predict our actions, though we know that we can often predict those of other people, especially if they are elderly. Much as we esteem the old gentleman who is our neighbour in the country, we know that when grouse are mentioned he will tell the story of the grouse in the gun-room. But we ourselves are not so mechanical: we never tell an anecdote to the same person twice, or even once unless he is sure to enjoy it; although we once met (say) Bismarck, we are quite capable of hearing him mentioned without relating the occasion when we met him. In this sense, everybody thinks that he himself has free will, though he knows that no one else has. The desire for this kind of free will seems to be no better than a form of vanity. I do not believe that this desire can be gratified with any certainty; but the other, more respectable desires are, I believe, not inconsistent with any tenable form of determinism.

We have thus two questions to consider: (1) Are human actions

* The remainder of this selection is reprinted from pp. 247–56 of *Our Knowledge of the External World*, by Bertrand Russell by permission of W. W. Norton & Company, Inc. Copyright 1929 by W. W. Norton & Company, Inc.

theoretically predictable from a sufficient number of antecedents? (2) Are human actions subject to an external compulsion? The two questions, as I shall try to show, are entirely distinct, and we may answer the first in the affirmative without therefore being forced to give an affirmative answer to the second.

(1) *Are human actions theoretically predictable from a sufficient number of antecedents?* Let us first endeavour to give precision to this question. We may state the question thus: Is there some constant relation between an act and a certain number of earlier events, such that, when the earlier events are given, only one act, or at most only acts with some well-marked character, can have this relation to the earlier events? If this is the case, then, as soon as the earlier events are known, it is theoretically possible to predict either the precise act, or at least the character necessary to its fulfilling the constant relation.

To this question, a negative answer has been given by Bergson, in a form which calls in question the general applicability of the law of causation. He maintains that every event, and more particularly every mental event, embodies so much of the past that it could not possibly have occurred at any earlier time, and is therefore necessarily quite different from all previous and subsequent events. If, for example, I read a certain poem many times, my experience on each occasion is modified by the previous readings, and my emotions are never repeated exactly. The principle of causation, according to him, asserts that the same cause, if repeated, will produce the same effect. But owing to memory, he contends, this principle does not apply to mental events. What is apparently the same cause, if repeated is modified by the mere fact of repetition, and cannot produce the same effect. He infers that every mental event is a genuine novelty, not predictable from the past, because the past contains nothing exactly like it by which we could imagine it. And on this ground he regards the freedom of the will as unassailable.

Bergson's contention has undoubtedly a great deal of truth, and I have no wish to deny its importance. But I do not think its consequences are quite what he believes them to be. It is not necessary for the determinist to maintain that he can foresee the whole particularity of the act which will be performed. If he could foresee that A was going to murder B, his foresight would not be invalidated by the fact that he could not know all the infinite complexity of A's state of mind in committing the murder, nor whether the murder was to be performed with a knife or with a revolver. If the *kind* of act which will be performed can be foreseen within narrow limits, it is of little practical interest that there are fine shades which cannot be foreseen. No doubt every time the story of the grouse in the gun-room is told, there will be slight differences due to increasing habitualness, but they do not invalidate the prediction that the story will be told. And there is nothing in Bergson's argument to show that we can never predict what *kind* of act will be performed.

Again, his statement of the law of causation is inadequate. The law does not state merely that, if the *same* cause is repeated, the *same* effect will result. It states rather that there is a constant relation between causes of certain kinds and effects of certain kinds. For example, if a body falls freely, there is a constant relation between the height through which it falls and the time it takes in falling. It is not necessary to have a body fall through the *same* height which has been previously observed, in order to be able to foretell the length of time occupied in falling. If this were necessary, no prediction would be possible, since it would be impossible to make the height exactly the same on two occasions. Similarly, the attraction which the sun will exert on the earth is not only known at distances for which it has been observed, but at all distances, because it is known to vary as the inverse square of the distance. In fact, what is found to be repeated is always the *relation* of cause and effect, not the cause itself; all that is necessary as regards the cause is that it should be of the same *kind* (in the relevant respect) as earlier causes whose effects have been observed.

Another respect in which Bergson's statement of causation is inadequate is in its assumption that the cause must be *one* event, whereas it may be two or more events, or even some continuous process. The substantive question at issue is whether mental events are determined by the past. Now in such a case as the repeated reading of a poem, it is obvious that our feelings in reading the poem are most emphatically dependent upon the past, but not upon one single event in the past. All our previous readings of the poem must be included in the cause. But we easily perceive a certain law according to which the effect varies as the previous readings increase in number, and in fact Bergson himself tacitly assumes such a law. We decide at last not to read the poem again, because we know that this time the effect would be boredom. We may not know all the niceties and shades of the boredom we should feel, but we know enough to guide our decision, and the prophecy of boredom is none the less true for being more or less general. Thus the kind of cases upon which Bergson relies are insufficient to show the impossibility of prediction in the only sense in which prediction has practical or emotional interest. We may therefore leave the consideration of his arguments and address ourselves to the problem directly.

The law of causation, according to which later events can theoretically be predicted by means of earlier events, has often been held to be *a priori*, a necessity of thought, a category without which science would be impossible. These claims seem to me excessive. In certain directions the law has been verified empirically, and in other directions there is no positive evidence against it. But science can use it where it has been found to be true, without being forced into any assumption as to its truth in other fields. We cannot, therefore, feel any *a priori* certainty that causation must apply to human volitions.

The question how far human volitions are subject to causal laws is a purely empirical one. Empirically it seems plain that the great majority

of our volitions have causes, but it cannot, on this account, be held necessarily certain that all have causes. There are, however, precisely the same kinds of reasons for regarding it as probable that they all have causes as there are in the case of physical events.

We may suppose—though this is doubtful—that there are laws of correlation of the mental and the physical, in virtue of which, given the state of all the matter in the world, and therefore of all the brains and living organisms, the state of all the minds in the world could be inferred, while conversely the state of all the matter in the world could be inferred if the state of all the minds were given. It is obvious that there is *some* degree of correlation between brain and mind, and it is impossible to say how complete it may be. This, however, is not the point which I wish to elicit. What I wish to urge is that, even if we admit the most extreme claims of determinism and of correlation of mind and brain, still the consequences inimical to what is worth preserving in free will do not follow. The belief that they follow results, I think, entirely from the assimilation of causes to volitions, and from the notion that causes *compel* their effects in some sense analogous to that in which a human authority can compel a man to do what he would rather not do. This assimilation, as soon as the true nature of scientific causal laws is realised, is seen to be a sheer mistake. But this brings us to the second of the two questions which we raised in regard to free will, namely, whether, assuming determinism, our actions can be in any proper sense regarded as compelled by outside forces.

(2) *Are human actions subject to an external compulsion?* We have, in deliberation, a subjective sense of freedom, which is sometimes alleged against the view that volitions have causes. This sense of freedom, however, is only a sense that we can choose which we please of a number of alternatives: it does not show us that there is no causal connection between what we please to choose and our previous history. The supposed inconsistency of these two springs from the habit of conceiving causes as analogous to volitions—a habit which often survives unconsciously in those who intend to conceive causes in a more scientific manner. If a cause is analogous to a volition, outside causes will be analogous to an alien will, and acts predictable from outside causes will be subject to compulsion. But this view of cause is one to which science lends no countenance. Causes, we have seen, do not *compel* their effects, any more than effects *compel* their causes. There is a mutual relation, so that either can be inferred from the other. When the geologist infers the past state of the earth from its present state, we should not say that the present state *compels* the past state to have been what it was; yet it renders it necessary as a consequence of the data, in the only sense in which effects are rendered necessary by their causes. The difference which we *feel,* in this respect, between causes and effects is a mere confusion due to the fact that we remember past events but do not happen to have memory of the future.

The apparent indeterminateness of the future, upon which some

advocates of free will rely, is merely a result of our ignorance. It is plain that no desirable kind of free will can be dependent simply upon our ignorance; for if that were the case, animals would be more free than men, and savages than civilised people. Free will in any valuable sense must be compatible with the fullest knowledge. Now, quite apart from any assumption as to causality, it is obvious that complete knowledge would embrace the future as well as the past. Our knowledge of the past is not wholly based upon causal inferences, but is partly derived from memory. It is a mere accident that we have no memory of the future. We might—as in the pretended visions of seers—see future events immediately, in the way in which we see past events. They certainly will be what they will be, and are in this sense just as determined as the past. If we saw future events in the same immediate way in which we see past events, what kind of free will would still be possible? Such a kind would be wholly independent of determinism: it could not be contrary to even the most entirely universal reign of causality. And such a kind must contain whatever is worth having in free will, since it is impossible to believe that mere ignorance can be the essential conditon of any good thing. Let us therefore imagine a set of beings who know the whole future with absolute certainty, and let us ask ourselves whether they could have anything that we should call free will.

Such beings as we are imagining would not have to wait for the event in order to know what decision they were going to adopt on some future occasion. They would know now what their volitions were going to be. But would they have any reason to regret this knowledge? Surely not, unless the foreseen volitions were in themselves regrettable. And it is less likely that the foreseen volitions would be regrettable if the steps which would lead to them were also foreseen. It is difficult not to suppose that what is foreseen is fated, and must happen however much it may be dreaded. But human actions are the outcome of desire, and no foreseeing can be true unless it takes account of desire. A foreseen volition will have to be one which does not become odious through being foreseen. The beings we are imagining would easily come to know the causal connections of volitions, and therefore their volitions would be better calculated to satisfy their desires than ours are. Since volitions are the outcome of desires, a prevision of volitions contrary to desires could not be a true one. It must be remembered that the supposed prevision would not create the future any more than memory creates the past. We do not think we were necessarily not free in the past, merely because we can now remember our past volitions. Similarly, we might be free in the future, even if we could now see what our future volitions were going to be. Freedom, in short, in any valuable sense, demands only that our volitions shall be, as they are, the result of our own desires, not of an outside force compelling us to will what we would rather not will. Everything else is confusion of thought, due to

the feeling that knowledge *compels* the happening of what it knows when this is future, though it is at once obvious that knowledge has no such power in regard to the past. Free will, therefore, is true in the only form which is important; and the desire for other forms is a mere effect of insufficient analysis.