

STUDIES IN THE LOGIC OF EXPLANATION

CARL G. HEMPEL AND PAUL OPPENHEIM¹

§1. *Introduction.*

To explain the phenomena in the world of our experience, to answer the question "why?" rather than only the question "what?", is one of the foremost objectives of all rational inquiry; and especially, scientific research in its various branches strives to go beyond a mere description of its subject matter by providing an explanation of the phenomena it investigates. While there is rather general agreement about this chief objective of science, there exists considerable difference of opinion as to the function and the essential characteristics of scientific explanation. In the present essay, an attempt will be made to shed some light on these issues by means of an elementary survey of the basic pattern of scientific explanation and a subsequent more rigorous analysis of the concept of law and of the logical structure of explanatory arguments.

The elementary survey is presented in Part I of this article; Part II contains an analysis of the concept of emergence; in Part III, an attempt is made to exhibit and to clarify in a more rigorous manner some of the peculiar and perplexing logical problems to which the familiar elementary analysis of explanation gives rise. Part IV, finally, is devoted to an examination of the idea of explanatory power of a theory; an explicit definition, and, based on it, a formal theory of this concept are developed for the case of a scientific language of simple logical structure.

PART I. ELEMENTARY SURVEY OF SCIENTIFIC EXPLANATION

§2. *Some illustrations.*

A mercury thermometer is rapidly immersed in hot water; there occurs a temporary drop of the mercury column, which is then followed by a swift rise. How is this phenomenon to be explained? The increase in temperature affects at first only the glass tube of the thermometer; it expands and thus provides a larger space for the mercury inside, whose surface therefore drops. As soon as by heat conduction the rise in temperature reaches the mercury, however, the latter expands, and as its coefficient of expansion is considerably larger than that of

¹ This paper represents the outcome of a series of discussions among the authors; their individual contributions cannot be separated in detail. The technical developments contained in Part IV, however, are due to the first author, who also put the article into its final form.

Some of the ideas presented in Part II were suggested by our common friend, Kurt Grelling, who, together with his wife, became a victim of Nazi terror during the war. Those ideas were developed by Grelling, in a discussion by correspondence with the present authors, of emergence and related concepts. By including at least some of that material, which is indicated in the text, in the present paper, we feel that we are realizing the hope expressed by Grelling that his contributions might not entirely fall into oblivion.

We wish to express our thanks to Dr. Rudolf Carnap, Dr. Herbert Feigl, Dr. Nelson Goodman, and Dr. W. V. Quine for stimulating discussions and constructive criticism.

glass, a rise of the mercury level results.—This account consists of statements of two kinds. Those of the first kind indicate certain conditions which are realized prior to, or at the same time as, the phenomenon to be explained; we shall refer to them briefly as antecedent conditions. In our illustration, the antecedent conditions include, among others, the fact that the thermometer consists of a glass tube which is partly filled with mercury, and that it is immersed into hot water. The statements of the second kind express certain general laws; in our case, these include the laws of the thermic expansion of mercury and of glass, and a statement about the small thermic conductivity of glass. The two sets of statements, if adequately and completely formulated, explain the phenomenon under consideration: They entail the consequence that the mercury will first drop, then rise. Thus, the event under discussion is explained by subsuming it under general laws, i.e., by showing that it occurred in accordance with those laws, by virtue of the realization of certain specified antecedent conditions.

Consider another illustration. To an observer in a row boat, that part of an oar which is under water appears to be bent upwards. The phenomenon is explained by means of general laws—mainly the law of refraction and the law that water is an optically denser medium than air—and by reference to certain antecedent conditions—especially the facts that part of the oar is in the water, part in the air, and that the oar is practically a straight piece of wood.—Thus, here again, the question “*Why* does the phenomenon happen?” is construed as meaning “according to what general laws, and by virtue of what antecedent conditions does the phenomenon occur?”

So far, we have considered exclusively the explanation of particular events occurring at a certain time and place. But the question “*Why?*” may be raised also in regard to general laws. Thus, in our last illustration, the question might be asked: Why does the propagation of light conform to the law of refraction? Classical physics answers in terms of the undulatory theory of light, i.e. by stating that the propagation of light is a wave phenomenon of a certain general type, and that all wave phenomena of that type satisfy the law of refraction. Thus, the explanation of a general regularity consists in subsuming it under another, more comprehensive regularity, under a more general law.—Similarly, the validity of Galileo’s law for the free fall of bodies near the earth’s surface can be explained by deducing it from a more comprehensive set of laws, namely Newton’s laws of motion and his law of gravitation, together with some statements about particular facts, namely the mass and the radius of the earth.

§3. *The basic pattern of scientific explanation.*

From the preceding sample cases let us now abstract some general characteristics of scientific explanation. We divide an explanation into two major constituents, the explanandum and the explanans². By the explanandum, we

² These two expressions, derived from the Latin *explanare*, were adopted in preference to the perhaps more customary terms “explicandum” and “explicans” in order to reserve the latter for use in the context of explication of meaning, or analysis. On explication in this sense, cf. Carnap, [Concepts], p. 513.—Abbreviated titles in brackets refer to the bibliography at the end of this article.

understand the sentence describing the phenomenon to be explained (not that phenomenon itself); by the explanans, the class of those sentences which are adduced to account for the phenomenon. As was noted before, the explanans falls into two subclasses; one of these contains certain sentences $C_1, C_2, \dots, C_k$ which state specific antecedent conditions; the other is a set of sentences $L_1, L_2, \dots, L_r$ which represent general laws.

If a proposed explanation is to be sound, its constituents have to satisfy certain conditions of adequacy, which may be divided into logical and empirical conditions. For the following discussion, it will be sufficient to formulate these requirements in a slightly vague manner; in Part III, a more rigorous analysis and a more precise restatement of these criteria will be presented.

I. *Logical conditions of adequacy.*

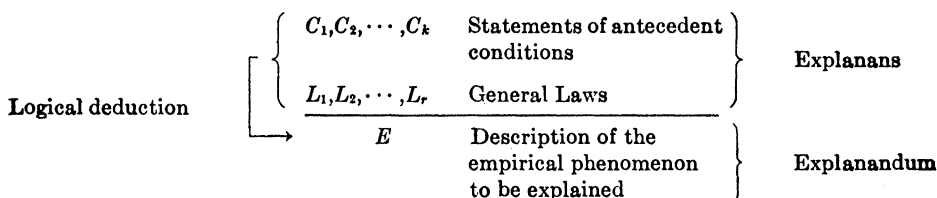
- (R1) The explanandum must be a logical consequence of the explanans; in other words, the explanandum must be logically deducible from the information contained in the explanans, for otherwise, the explanans would not constitute adequate grounds for the explanandum.
- (R2) The explanans must contain general laws, and these must actually be required for the derivation of the explanandum.—We shall not make it a necessary condition for a sound explanation, however, that the explanans must contain at least one statement which is not a law; for, to mention just one reason, we would surely want to consider as an explanation the derivation of the general regularities governing the motion of double stars from the laws of celestial mechanics, even though all the statements in the explanans are general laws.
- (R3) The explanans must have empirical content; i.e., it must be capable, at least in principle, of test by experiment or observation.—This condition is implicit in (R1); for since the explanandum is assumed to describe some empirical phenomenon, it follows from (R1) that the explanans entails at least one consequence of empirical character, and this fact confers upon it testability and empirical content. But the point deserves special mention because, as will be seen in §4, certain arguments which have been offered as explanations in the natural and in the social sciences violate this requirement.

II. *Empirical condition of adequacy.*

- (R4) The sentences constituting the explanans must be true.
That in a sound explanation, the statements constituting the explanans have to satisfy some condition of factual correctness is obvious. But it might seem more appropriate to stipulate that the explanans has to be highly confirmed by all the relevant evidence available rather than that it should be true. This stipulation however, leads to awkward consequences. Suppose that a certain phenomenon was explained at an earlier stage of science, by means of an explanans which was well supported by the evidence then at hand, but which had been highly disconfirmed by more recent empirical findings. In such a case, we

would have to say that originally the explanatory account was a correct explanation, but that it ceased to be one later, when unfavorable evidence was discovered. This does not appear to accord with sound common usage, which directs us to say that on the basis of the limited initial evidence, the truth of the explanans, and thus the soundness of the explanation, had been quite probable, but that the ampler evidence now available made it highly probable that the explanans was not true, and hence that the account in question was not—and had never been—a correct explanation. (A similar point will be made and illustrated, with respect to the requirement of truth for laws, in the beginning of §6.)

Some of the characteristics of an explanation which have been indicated so far may be summarized in the following schema:



Let us note here that the same formal analysis, including the four necessary conditions, applies to scientific prediction as well as to explanation. The difference between the two is of a pragmatic character. If E is given, i.e. if we know that the phenomenon described by E has occurred, and a suitable set of statements $C_1, C_2, \dots, C_k, L_1, L_2, \dots, L_r$ is provided afterwards, we speak of an explanation of the phenomenon in question. If the latter statements are given and E is derived prior to the occurrence of the phenomenon it describes, we speak of a prediction. It may be said, therefore, that an explanation is not fully adequate unless its explanans, if taken account of in time, could have served as a basis for predicting the phenomenon under consideration.^{2a}—Consequently, whatever will be said in this article concerning the logical characteristics of explanation or prediction will be applicable to either, even if only one of them should be mentioned.

It is this potential predictive force which gives scientific explanation its importance: only to the extent that we are able to explain empirical facts can we attain the major objective of scientific research, namely not merely to record the phenomena of our experience, but to learn from them, by basing upon them theoretical generalizations which enable us to anticipate new occurrences and to control, at least to some extent, the changes in our environment.

Many explanations which are customarily offered, especially in pre-scientific discourse, lack this predictive character, however. Thus, it may be explained

^{2a} The logical similarity of explanation and prediction, and the fact that one is directed towards past occurrences, the other towards future ones, is well expressed in the terms "post-dictability" and "predictability" used by Reichenbach in [Quantum Mechanics], p. 13.

that a car turned over on the road "because" one of its tires blew out while the car was travelling at high speed. Clearly, on the basis of just this information, the accident could not have been predicted, for the explanans provides no explicit general laws by means of which the prediction might be effected, nor does it state adequately the antecedent conditions which would be needed for the prediction.—The same point may be illustrated by reference to W. S. Jevons's view that every explanation consists in pointing out a resemblance between facts, and that in some cases this process may require no reference to laws at all and "may involve nothing more than a single identity, as when we explain the appearance of shooting stars by showing that they are identical with portions of a comet".³ But clearly, this identity does not provide an explanation of the phenomenon of shooting stars unless we presuppose the laws governing the development of heat and light as the effect of friction. The observation of similarities has explanatory value only if it involves at least tacit reference to general laws.

In some cases, incomplete explanatory arguments of the kind here illustrated suppress parts of the explanans simply as "obvious"; in other cases, they seem to involve the assumption that while the missing parts are not obvious, the incomplete explanans could at least, with appropriate effort, be so supplemented as to make a strict derivation of the explanandum possible. This assumption may be justifiable in some cases, as when we say that a lump of sugar disappeared "because" it was put into hot tea, but it is surely not satisfied in many other cases. Thus, when certain peculiarities in the work of an artist are explained as outgrowths of a specific type of neurosis, this observation may contain significant clues, but in general it does not afford a sufficient basis for a potential prediction of those peculiarities. In cases of this kind, an incomplete explanation may at best be considered as indicating some positive correlation between the antecedent conditions adduced and the type of phenomenon to be explained, and as pointing out a direction in which further research might be carried on in order to complete the explanatory account.

The type of explanation which has been considered here so far is often referred to as causal explanation. If E describes a particular event, then the antecedent circumstances described in the sentences $C_1, C_2, \dots, C_k$ may be said jointly to "cause" that event, in the sense that there are certain empirical regularities, expressed by the laws $L_1, L_2, \dots, L_r$, which imply that whenever conditions of the kind indicated by $C_1, C_2, \dots, C_k$ occur, an event of the kind described in E will take place. Statements such as $L_1, L_2, \dots, L_r$, which assert general and unexceptional connections between specified characteristics of events, are customarily called causal, or deterministic, laws. They are to be distinguished from the so-called statistical laws which assert that in the long run, an explicitly stated percentage of all cases satisfying a given set of conditions are accompanied by an event of a certain specified kind. Certain cases of scientific explanation involve "subsumption" of the explanandum under a set of laws of which at least some are statistical in character. Analysis of the peculiar logical structure

³ [Principles], p. 533.

of that type of subsumption involves difficult special problems. The present essay will be restricted to an examination of the causal type of explanation, which has retained its significance in large segments of contemporary science, and even in some areas where a more adequate account calls for reference to statistical laws.⁴

§4. *Explanation in the non-physical sciences. Motivational and teleological approaches.*

Our characterization of scientific explanation is so far based on a study of cases taken from the physical sciences. But the general principles thus obtained apply also outside this area.⁵ Thus, various types of behavior in laboratory animals and in human subjects are explained in psychology by subsumption under laws or even general theories of learning or conditioning; and while frequently, the regularities invoked cannot be stated with the same generality and precision as in physics or chemistry, it is clear, at least, that the general character of those explanations conforms to our earlier characterization.

Let us now consider an illustration involving sociological and economic factors. In the fall of 1946, there occurred at the cotton exchanges of the United States a price drop which was so severe that the exchanges in New York, New Orleans, and Chicago had to suspend their activities temporarily. In an attempt to explain this occurrence, newspapers traced it back to a large-scale speculator in New Orleans who had feared his holdings were too large and had therefore begun to liquidate his stocks; smaller speculators had then followed his example

⁴ The account given above of the general characteristics of explanation and prediction in science is by no means novel; it merely summarizes and states explicitly some fundamental points which have been recognized by many scientists and methodologists.

Thus, e.g., Mill says: "An individual fact is said to be explained by pointing out its cause, that is, by stating the law or laws of causation of which its production is an instance", and "a law of uniformity in nature is said to be explained when another law or laws are pointed out, of which that law itself is but a case, and from which it could be deduced." ([*Logic*], Book III, Chapter XII, section 1). Similarly, Jevons, whose general characterization of explanation was critically discussed above, stresses that "the most important process of explanation consists in showing that an observed fact is one case of a general law or tendency." ([*Principles*], p. 533). Ducasse states the same point as follows: "Explanation essentially consists in the offering of a hypothesis of fact, standing to the fact to be explained as case of antecedent to case of consequent of some already known law of connection." ([*Explanation*], pp. 150-51). A lucid analysis of the fundamental structure of explanation and prediction was given by Popper in [*Forschung*], section 12, and, in an improved version, in his work [*Society*], especially in Chapter 25 and in note 7 referring to that chapter.—For a recent characterization of explanation as subsumption under general theories, cf., for example, Hull's concise discussion in [*Principles*], chapter I. A clear elementary examination of certain aspects of explanation is given in Hospers, [*Explanation*], and a concise survey of many of the essentials of scientific explanation which are considered in the first two parts of the present study may be found in Feigl, [*Operationism*], pp. 284 ff.

⁵ On the subject of explanation in the social sciences, especially in history, cf. also the following publications, which may serve to supplement and amplify the brief discussion to be presented here: Hempel, [*Laws*]; Popper, [*Society*]; White, [*Explanation*]; and the articles *Cause* and *Understanding* in Beard and Hook, [*Terminology*].

in a panic and had thus touched off the critical decline. Without attempting to assess the merits of the argument, let us note that the explanation here suggested again involves statements about antecedent conditions and the assumption of general regularities. The former include the facts that the first speculator had large stocks of cotton, that there were smaller speculators with considerable holdings, that there existed the institution of the cotton exchanges with their specific mode of operation, etc. The general regularities referred to are—as often in semi-popular explanations—not explicitly mentioned; but there is obviously implied some form of the law of supply and demand to account for the drop in cotton prices in terms of the greatly increased supply under conditions of practically unchanged demand; besides, reliance is necessary on certain regularities in the behavior of individuals who are trying to preserve or improve their economic position. Such laws cannot be formulated at present with satisfactory precision and generality, and therefore, the suggested explanation is surely incomplete, but its intention is unmistakably to account for the phenomenon by integrating it into a general pattern of economic and socio-psychological regularities.

We turn to an explanatory argument taken from the field of linguistics.⁶ In Northern France, there exist a large variety of words synonymous with the English “bee,” whereas in Southern France, essentially only one such word is in existence. For this discrepancy, the explanation has been suggested that in the Latin epoch, the South of France used the word “apicula”, the North the word “apis”. The latter, because of a process of phonologic decay in Northern France, became the monosyllabic word “é”; and monosyllables tend to be eliminated, especially if they contain few consonantic elements, for they are apt to give rise to misunderstandings. Thus, to avoid confusion, other words were selected. But “apicula”, which was reduced to “abelho”, remained clear enough and was retained, and finally it even entered into the standard language, in the form “abbaille”. While the explanation here described is incomplete in the sense characterized in the previous section, it clearly exhibits reference to specific antecedent conditions as well as to general laws.⁷

While illustrations of this kind tend to support the view that explanation in biology, psychology, and the social sciences has the same structure as in the physical sciences, the opinion is rather widely held that in many instances, the causal type of explanation is essentially inadequate in fields other than physics and chemistry, and especially in the study of purposive behavior. Let us ex-

⁶ The illustration is taken from Bonfante, [Semantics], section 3.

⁷ While in each of the last two illustrations, certain regularities are unquestionably relied upon in the explanatory argument, it is not possible to argue convincingly that the intended laws, which at present cannot all be stated explicitly, are of a causal rather than a statistical character. It is quite possible that most or all of the regularities which will be discovered as sociology develops will be of a statistical type. Cf., on this point, the suggestive observations by Zilsel in [Empiricism] section 8, and [Laws]. This issue does not affect, however, the main point we wish to make here, namely that in the social no less than in the physical sciences, subsumption under general regularities is indispensable for the explanation and the theoretical understanding of any phenomenon.

amine briefly some of the reasons which have been adduced in support of this view.

One of the most familiar among them is the idea that events involving the activities of humans singly or in groups have a peculiar uniqueness and irrepeatability which makes them inaccessible to causal explanation because the latter, which its reliance upon uniformities, presupposes repeatability of the phenomena under consideration. This argument which, incidentally, has also been used in support of the contention that the experimental method is inapplicable in psychology and the social sciences, involves a misunderstanding of the logical character of causal explanation. Every individual event, in the physical sciences no less than in psychology or the social sciences, is unique in the sense that it, with all its peculiar characteristics, does not repeat itself. Nevertheless, individual events may conform to, and thus be explainable by means of, general laws of the causal type. For all that a causal law asserts is that any event of a specified kind, i.e. any event having certain specified characteristics, is accompanied by another event which in turn has certain specified characteristics; for example, that in any event involving friction, heat is developed. And all that is needed for the testability and applicability of such laws is the recurrence of events with the antecedent characteristics, i.e. the repetition of those characteristics, but not of their individual instances. Thus, the argument is inconclusive. It gives occasion, however, to emphasize an important point concerning our earlier analysis: When we spoke of the explanation of a single event, the term "event" referred to the occurrence of some more or less complex characteristic in a specific spatio-temporal location or in a certain individual object, and not to *all* the characteristics of that object, or to all that goes on in that space-time region.

A second argument that should be mentioned here⁸ contends that the establishment of scientific generalizations—and thus of explanatory principles—for human behavior is impossible because the reactions of an individual in a given situation depend not only upon that situation, but also upon the previous history of the individual.—But surely, there is no *a priori* reason why generalizations should not be attainable which take into account this dependence of behavior on the past history of the agent. That indeed the given argument "proves" too much, and is therefore a *non sequitur*, is made evident by the existence of certain physical phenomena, such as magnetic hysteresis and elastic fatigue, in which the magnitude of a specific physical effect depends upon the past history of the system involved, and for which nevertheless certain general regularities have been established.

A third argument insists that the explanation of any phenomenon involving purposive behavior calls for reference to motivations and thus for teleological rather than causal analysis. Thus, for example, a fuller statement of the suggested explanation for the break in the cotton prices would have to indicate the large-scale speculator's motivations as one of the factors determining the event

⁸ Cf., for example, F. H. Knight's presentation of this argument in [Limitations], pp. 251–52.

in question. Thus, we have to refer to goals sought, and this, so the argument runs, introduces a type of explanation alien to the physical sciences. Unquestionably, many of the—frequently incomplete—explanations which are offered for human actions involve reference to goals and motives; but does this make them essentially different from the causal explanations of physics and chemistry? One difference which suggests itself lies in the circumstance that in motivated behavior, the future appears to affect the present in a manner which is not found in the causal explanations of the physical sciences. But clearly, when the action of a person is motivated, say, by the desire to reach a certain objective, then it is not the as yet unrealized future event of attaining that goal which can be said to determine his present behavior, for indeed the goal may never be actually reached; rather—to put it in crude terms—it is (a) his desire, present before the action, to attain that particular objective, and (b) his belief, likewise present before the action, that such and such a course of action is most likely to have the desired effect. The determining motives and beliefs, therefore, have to be classified among the antecedent conditions of a motivational explanation, and there is no formal difference on this account between motivational and causal explanation.

Neither does the fact that motives are not accessible to direct observation by an outside observer constitute an essential difference between the two kinds of explanation; for also the determining factors adduced in physical explanations are very frequently inaccessible to direct observation. This is the case, for instance, when opposite electric charges are adduced in explanation of the mutual attraction of two metal spheres. The presence of those charges, while eluding all direct observation, can be ascertained by various kinds of indirect test, and that is sufficient to guarantee the empirical character of the explanatory statement. Similarly, the presence of certain motivations may be ascertainable only by indirect methods, which may include reference to linguistic utterances of the subject in question, slips of the pen or of the tongue, etc.; but as long as these methods are “operationally determined” with reasonable clarity and precision, there is no essential difference in this respect between motivational explanation and causal explanation in physics.

A potential danger of explanation by motives lies in the fact that the method lends itself to the facile construction of ex-post-facto accounts without predictive force. It is a widespread tendency to “explain” an action by ascribing it to motives conjectured only after the action has taken place. While this procedure is not in itself objectionable, its soundness requires that (1) the motivational assumptions in question be capable of test, and (2) that suitable general laws be available to lend explanatory power to the assumed motives. Disregard of these requirements frequently deprives alleged motivational explanations of their cognitive significance.

The explanation of an action in terms of the motives of the agent is sometimes considered as a special kind of teleological explanation. As was pointed out above, motivational explanation, if adequately formulated, conforms to the conditions for causal explanation, so that the term “teleological” is a misnomer if it is

meant to imply either a non-causal character of the explanation or a peculiar determination of the present by the future. If this is borne in mind, however, the term "teleological" may be viewed, in this context, as referring to causal explanations in which some of the antecedent conditions are motives of the agent whose actions are to be explained.⁹

Teleological explanations of this kind have to be distinguished from a much more sweeping type, which has been claimed by certain schools of thought to be indispensable especially in biology. It consists in explaining characteristics of an organism by reference to certain ends or purposes which the characteristics are said to serve. In contradistinction to the cases examined before, the ends are not assumed here to be consciously or subconsciously pursued by the organism in question. Thus, for the phenomenon of mimicry, the explanation is sometimes offered that it serves the purpose of protecting the animals endowed with it from detection by its pursuers and thus tends to preserve the species.—Before teleological hypotheses of this kind can be appraised as to their potential explanatory power, their meaning has to be clarified. If they are intended somehow to express the idea that the purposes they refer to are inherent in the design of the universe, then clearly they are not capable of empirical test and thus violate the requirement (R3) stated in §3. In certain cases, however, assertions about the purposes of biological characteristics may be translatable into statements in non-teleological terminology which assert that those characteristics function in a specific manner which is essential to keeping the organism alive or to preserving the species.¹⁰ An attempt to state precisely what is meant by this latter assertion—or by the similar one that without those characteristics, and other things being equal, the organism or the species would not survive—encounters considerable difficulties. But these need not be discussed here. For even if we assume that biological statements in teleological form can be adequately translated into descriptive statements about the life-preserving function of certain biological characteristics, it is clear that (1) the use of the concept of purpose is not essential in these contexts, since the term "purpose" can be completely eliminated from the statements in question, and (2) teleological assumptions, while now endowed with empirical content, cannot serve as explanatory principles in the customary contexts. Thus, e.g., the fact that a

⁹ For a detailed logical analysis of the character and the function of the motivation concept in psychological theory, see Koch, [Motivation].—A stimulating discussion of teleological behavior from the standpoint of contemporary physics and biology is contained in the article [Teleology] by Rosenblueth, Wiener and Bigelow. The authors propose an interpretation of the concept of purpose which is free from metaphysical connotations, and they stress the importance of the concept thus obtained for a behavioristic analysis of machines and living organisms. While our formulations above intentionally use the crude terminology frequently applied in philosophical arguments concerning the applicability of causal explanation to purposive behavior, the analysis presented in the article referred to is couched in behavioristic terms and avoids reference to "motives" and the like.

¹⁰ An analysis of teleological statements in biology along these lines may be found in Woodger, [Principles], especially pp. 432 ff; essentially the same interpretation is advocated by Kaufmann in [Methodology], chapter 8.

given species of butterflies displays a particular kind of coloring cannot be inferred from—and therefore cannot be explained by means of—the statement that this type of coloring has the effect of protecting the butterflies from detection by pursuing birds, nor can the presence of red corpuscles in the human blood be inferred from the statement that those corpuscles have a specific function in assimilating oxygen and that this function is essential for the maintenance of life.

One of the reasons for the perseverance of teleological considerations in biology probably lies in the fruitfulness of the teleological approach as a heuristic device: Biological research which was psychologically motivated by a teleological orientation, by an interest in purposes in nature, has frequently led to important results which can be stated in non-teleological terminology and which increase our scientific knowledge of the causal connections between biological phenomena.

Another aspect that lends appeal to teleological considerations is their anthropomorphic character. A teleological explanation tends to make us feel that we really “understand” the phenomenon in question, because it is accounted for in terms of purposes, with which we are familiar from our own experience of purposive behavior. But it is important to distinguish here understanding in the psychological sense of a feeling of empathic familiarity from understanding in the theoretical, or cognitive, sense of exhibiting the phenomenon to be explained as a special case of some general regularity. The frequent insistence that explanation means the reduction of something unfamiliar to ideas or experiences already familiar to us is indeed misleading. For while some scientific explanations do have this psychological effect, it is by no means universal: The free fall of a physical body may well be said to be a more familiar phenomenon than the law of gravitation, by means of which it can be explained; and surely the basic ideas of the theory of relativity will appear to many to be far less familiar than the phenomena for which the theory accounts.

“Familiarity” of the explicans is not only not necessary for a sound explanation—as we have just tried to show—, but it is not sufficient either. This is shown by the many cases in which a proposed explicans sounds suggestively familiar, but upon closer inspection proves to be a mere metaphor, or an account lacking testability, or a set of statements which includes no general laws and therefore lacks explanatory power. A case in point is the neovitalistic attempt to explain biological phenomena by reference to an entelechy or vital force. The crucial point here is not—as it is sometimes made out to be—that entelechies cannot be seen or otherwise directly observed; for that is true also of gravitational fields, and yet, reference to such fields is essential in the explanation of various physical phenomena. The decisive difference between the two cases is that the physical explanation provides (1) methods of testing, albeit indirectly, assertions about gravitational fields, and (2) general laws concerning the strength of gravitational fields, and the behavior of objects moving in them. Explanations by entelechies satisfy the analogue of neither of these two conditions. Failure to satisfy the first condition represents a violation of (R3); it renders all statements about entelechies inaccessible to empirical test and thus devoid of empirical meaning. Failure to comply with the second condition involves a

violation of (R2). It deprives the concept of entelechy of all explanatory import; for explanatory power never resides in a concept, but always in the general laws in which it functions. Therefore, notwithstanding the flavor of familiarity of the metaphor it invokes, the neovitalistic approach cannot provide theoretical understanding.

The preceding observations about familiarity and understanding can be applied, in a similar manner, to the view held by some scholars that the explanation, or the understanding, of human actions requires an empathic understanding of the personalities of the agents¹¹. This understanding of another person in terms of one's own psychological functioning may prove a useful heuristic device in the search for general psychological principles which might provide a theoretical explanation; but the existence of empathy on the part of the scientist is neither a necessary nor a sufficient condition for the explanation, or the scientific understanding, of any human action. It is not necessary, for the behavior of psychotics or of people belonging to a culture very different from that of the scientist may sometimes be explainable and predictable in terms of general principles even though the scientist who establishes or applies those principles may not be able to understand his subjects empathically. And empathy is not sufficient to guarantee a sound explanation, for a strong feeling of empathy may exist even in cases where we completely misjudge a given personality. Moreover, as the late Dr. Zilsel has pointed out, empathy leads with ease to incompatible results; thus, when the population of a town has long been subjected to heavy bombing attacks, we can understand, in the empathic sense, that its morale should have broken down completely, but we can understand with the same ease also that it should have developed a defiant spirit of resistance. Arguments of this kind often appear quite convincing; but they are of an *ex post facto* character and lack cognitive significance unless they are supplemented by testable explanatory principles in the form of laws or theories.

Familiarity of the explanans, therefore, no matter whether it is achieved through the use of teleological terminology, through neovitalistic metaphors, or through other means, is no indication of the cognitive import and the predictive force of a proposed explanation. Besides, the extent to which an idea will be considered as familiar varies from person to person and from time to time, and a psychological factor of this kind certainly cannot serve as a standard in assessing the worth of a proposed explanation. The decisive requirement for every sound explanation remains that it subsume the explanandum under general laws.

PART II. ON THE IDEA OF EMERGENCE

§5. *Levels of Explanation. Analysis of Emergence.*

As has been shown above, a phenomenon may often be explained by sets of laws of different degrees of generality. The changing positions of a planet, for example, may be explained by subsumption under Kepler's laws, or by deriva-

¹¹ For a more detailed discussion of this view on the basis of the general principles outlined above, cf. Zilsel, [Empiricism], sections 7 and 8, and Hempel, [Laws], section 6.

Assertions of this type, then, appear to represent the rational core of the doctrine of emergence. In its revised form, the idea of emergence no longer carries with it the connotation of absolute unpredictability—a notion which is objectionable not only because it involves and perpetuates certain logical misunderstandings, but also because, not unlike the ideas of neo-vitalism, it encourages an attitude of resignation which is stifling for scientific research. No doubt it is this characteristic, together with its theoretical sterility, which accounts for the rejection, by the majority of contemporary scientists, of the classical absolutistic doctrine of emergence.¹⁸

PART III. LOGICAL ANALYSIS OF LAW AND EXPLANATION

§6. *Problems of the concept of general law.*

From our general survey of the characteristics of scientific explanation, we now turn to a closer examination of its logical structure. The explanation of a phenomenon, we noted, consists in its subsumption under laws or under a theory. But what is a law, what is a theory? While the meaning of these concepts seems intuitively clear, an attempt to construct adequate explicit definitions for them encounters considerable difficulties. In the present section, some basic problems of the concept of law will be described and analyzed; in the next section, we intend to propose, on the basis of the suggestions thus obtained, definitions of law and of explanation for a formalized model language of a simple logical structure.

The concept of law will be construed here so as to apply to true statements only. The apparently plausible alternative procedure of requiring high confirmation rather than truth of a law seems to be inadequate: It would lead to a relativized concept of law, which would be expressed by the phrase "sentence *S* is a law relatively to the evidence *E*". This does not seem to accord with the meaning customarily assigned to the concept of law in science and in methodological inquiry. Thus, for example, we would not say that Bode's general formula for the distance of the planets from the sun was a law relatively to the astronomical evidence available in the 1770s, when Bode propounded it, and that it ceased to be a law after the discovery of Neptune and the determination of its distance from the sun; rather, we would say that the limited original evidence had given a high probability to the assumption that the formula was a law, whereas more recent additional information reduced that probability so much as to make it practically certain that Bode's formula is not generally true, and hence not a law.^{18a}

cannot in the present state of our knowledge, i.e., prior to the working-out of many empirical correlations between behavior and its physiological correlates, be known even inferentially from a mere knowledge of the underlying, molecular, facts of physics and physiology." (l. c., pp. 7-8).—In a similar manner, Hull uses the distinction between molar and molecular theories and points out that theories of the latter type are not at present available in psychology. Cf. [Principles], pp. 19ff.; [Variables], p. 275.

¹⁸ This attitude of the scientist is voiced, for example, by Hull in [Principles], pp. 24-28.

^{18a} The requirement of truth for laws has the consequence that a given empirical statement *S* can never be definitely known to be a law; for the sentence affirming the truth of *S*

Apart from being true, a law will have to satisfy a number of additional conditions. These can be studied independently of the factual requirement of truth, for they refer, as it were, to all logically possible laws, no matter whether factually true or false. Adopting a convenient term proposed by Goodman¹⁹, we will say that a sentence is lawlike if it has all the characteristics of a general law, with the possible exception of truth. Hence, every law is a lawlike sentence, but not conversely.

Our problem of analyzing the concept of law thus reduces to that of explicating the meaning of "lawlike sentence". We shall construe the class of lawlike sentences as including analytic general statements, such as "A rose is a rose", as well as the lawlike sentences of empirical science, which have empirical content.²⁰ It will not be necessary to require that each lawlike sentence permissible in explanatory contexts be of the second kind; rather, our definition of explanation will be so constructed as to guarantee the factual character of the totality of the laws—though not of every single one of them—which function in an explanation of an empirical fact.

What are the characteristics of lawlike sentences? First of all, lawlike sentences are statements of universal form, such as "All robins' eggs are greenish-blue", "All metals are conductors of electricity", "At constant pressure, any gas expands with increasing temperature". As these examples illustrate, a lawlike sentence usually is not only of universal, but also of conditional form; it makes an assertion to the effect that universally, if a certain set of conditions, *C*, is realized, then another specified set of conditions, *E*, is realized as well. The standard form for the symbolic expression of a lawlike sentence is therefore the universal conditional. However, since any conditional statement can be transformed into a non-conditional one, conditional form will not be considered as essential for a lawlike sentence, while universal character will be held indispensable.

But the requirement of universal form is not sufficient to characterize lawlike sentences. Suppose, for example, that a certain basket, *b*, contains at a certain time *t* a number of red apples and nothing else.²¹ Then the statement

(*S*₁) Every apple in basket *b* at time *t* is red

is both true and of universal form. Yet the sentence does not qualify as a law; we would refuse, for example, to explain by subsumption under it the fact

is logically equivalent with *S* and is therefore capable only of acquiring a more or less high probability, or degree of confirmation, relatively to the experimental evidence available at any given time. On this point, cf. Carnap, [Remarks].—For an excellent non-technical exposition of the semantical concept of truth, which is here applied, the reader is referred to Tarski, [Truth].

¹⁹ [Counterfactuals]. p. 125.

²⁰ This procedure was suggested by Goodman's approach in [Counterfactuals].—Reichenbach, in a detailed examination of the concept of law, similarly construes his concept of nomological statement as including both analytic and synthetic sentences; cf. [Logic], chapter VIII.

²¹ The difficulty illustrated by this example was stated concisely by Langford ([Review]), who referred to it as the problem of distinguishing between universals of fact and causal universals. For further discussion and illustration of this point, see also Chisholm [Conditional], especially pp. 301f.—A systematic analysis of the problem was given by Goodman

that a particular apple chosen at random from the basket is red. What distinguishes S_1 from a lawlike sentence? Two points suggest themselves, which will be considered in turn, namely, finite scope, and reference to a specified object.

First, the sentence S_1 makes, in effect, an assertion about a finite number of objects only, and this seems irreconcilable with the claim to universality which is commonly associated with the notion of law.²² But are not Kepler's laws considered as lawlike although they refer to a finite set of planets only? And might we not even be willing to consider as lawlike a sentence such as the following?

(S_2) All the sixteen ice cubes in the freezing tray of this refrigerator have a temperature of less than 10 degrees centigrade.

This point might well be granted; but there is an essential difference between S_1 on the one hand and Kepler's laws as well as S_2 on the other: The latter, while finite in scope, are known to be consequences of more comprehensive laws whose scope is not limited, while for S_1 this is not the case.

Adopting a procedure recently suggested by Reichenbach²³, we will therefore distinguish between fundamental and derivative laws. A statement will be called a derivative law if it is of universal character and follows from some fundamental laws. The concept of fundamental law requires further clarification; so far, we may say that fundamental laws, and similarly fundamental lawlike sentences, should satisfy a certain condition of non-limitation of scope.

It would be excessive, however, to deny the status of fundamental lawlike sentence to all statements which, in effect, make an assertion about a finite class of objects only, for that would rule out also a sentence such as "All robins' eggs are greenish-blue", since presumably the class of all robins' eggs—past, present, and future—is finite. But again, there is an essential difference between this sentence and, say, S_1 . It requires empirical knowledge to establish the finiteness of the class of robins' eggs, whereas, when the sentence S_1 is construed in a manner which renders it intuitively unlawlike, the terms "basket b " and "apple" are understood so as to imply finiteness of the class of apples in the basket at time t . Thus, so to speak, the meaning of its constitutive terms alone—without additional factual information—entails that S_1 has a finite scope.—Fundamental laws, then, will have to be construed so as to satisfy what we have called a condition of non-limited scope; our formulation of that condition however, which refers to what is entailed by "the meaning" of certain expressions, is too vague and will have to be revised later. Let us note in passing that the stipulation here envisaged would bar from the class of fundamental lawlike sentences also such undesirable candidates as "All uranic objects are spherical", where "uranic" means the property

in [Counterfactuals], especially part III.—While not concerned with the specific point under discussion, the detailed examination of counterfactual conditionals and their relation to laws of nature, in Chapter VIII of Lewis's work [Analysis], contains important observations on several of the issues raised in the present section.

²² The view that laws should be construed as not being limited to a finite domain has been expressed, among others, by Popper ([Forschung], section 13) and by Reichenbach ([Logic], p. 369).

²³ [Logic], p. 361.—Our terminology as well as the definitions to be proposed later for the two types of law do not coincide with Reichenbach's, however.

of being the planet Uranus; indeed, while this sentence has universal form, it fails to satisfy the condition of non-limited scope.

In our search for a general characterization of lawlike sentences, we now turn to a second clue which is provided by the sentence S_1 . In addition to violating the condition of non-limited scope, this sentence has the peculiarity of making reference to a particular object, the basket b ; and this, too, seems to violate the universal character of a law.²⁴ The restriction which seems indicated here, should however again be applied to fundamental lawlike sentences only; for a true general statement about the free fall of physical bodies on the moon, while referring to a particular object, would still constitute a law, albeit a derivative one.

It seems reasonable to stipulate, therefore, that a fundamental lawlike sentence must be of universal form and must contain no essential—i.e., uneliminable—occurrences of designations for particular objects. But this is not sufficient; indeed, just at this point, a particularly serious difficulty presents itself. Consider the sentence

(S_3) Everything that is either an apple in basket b at time t or a sample of ferric oxide is red.

If we use a special expression, say “ x is ferple”, as synonymous with “ x is either an apple in b at t or a sample of ferric oxide”, then the content of S_3 can be expressed in the form

(S_4) Everything that is ferple is red.

The statement thus obtained is of universal form and contains no designations of particular objects, and it also satisfies the condition of non-limited scope; yet clearly, S_4 can qualify as a fundamental lawlike sentence no more than can S_3 .

As long as “ferple” is a defined term of our language, the difficulty can readily be met by stipulating that after elimination of defined terms, a fundamental lawlike sentence must not contain essential occurrences of designations for particular objects. But this way out is of no avail when “ferple”, or another term of the kind illustrated by it, is a primitive predicate of the language under consideration. This reflection indicates that certain restrictions have to be imposed upon those predicates—i.e., terms for properties or relations,—which may occur in fundamental lawlike sentences.²⁵

²⁴ In physics, the idea that a law should not refer to any particular object has found its expression in the maxim that the general laws of physics should contain no reference to specific space-time points, and that spatio-temporal coordinates should occur in them only in the form of differences or differentials.

²⁵ The point illustrated by the sentences S_3 and S_4 above was made by Goodman, who has also emphasized the need to impose certain restrictions upon the predicates whose occurrence is to be permissible in lawlike sentences. These predicates are essentially the same as those which Goodman calls projectible. Goodman has suggested that the problems of establishing precise criteria for projectibility, of interpreting counterfactual conditionals, and of defining the concept of law are so intimately related as to be virtually aspects of a single problem. (Cf. his articles [Query] and [Counterfactuals].) One suggestion for an analysis of projectibility has recently been made by Carnap in [Application]. Goodman's note [Infirmities] contains critical observations on Carnap's proposals.

More specifically, the idea suggests itself of permitting a predicate in a fundamental lawlike sentence only if it is purely universal, or, as we shall say, purely qualitative, in character; in other words, if a statement of its meaning does not require reference to any one particular object or spatio-temporal location. Thus, the terms "soft", "green", "warmer than", "as long as", "liquid", "electrically charged", "female", "father of" are purely qualitative predicates, while "taller than the Eiffel Tower", "medieval", "lunar", "arctic", "Ming" are not.²⁶

Exclusion from fundamental lawlike sentences of predicates which are not purely qualitative would at the same time ensure satisfaction of the condition of non-limited scope; for the meaning of a purely qualitative predicate does not require a finite extension; and indeed, all the sentences considered above which violate the condition of non-limited scope make explicit or implicit reference to specific objects.

The stipulation just proposed suffers, however, from the vagueness of the concept of purely qualitative predicate. The question whether indication of the meaning of a given predicate in English does or does not require reference to some one specific object does not always permit an unequivocal answer since English as a natural language does not provide explicit definitions or other clear explications of meaning for its terms. It seems therefore reasonable to attempt definition of the concept of law not with respect to English or any other natural language, but rather with respect to a formalized language—let us call it a model language, *L*,—which is governed by a well-determined system of logical rules, and in which every term either is characterized as primitive or is introduced by an explicit definition in terms of the primitives.

This reference to a well-determined system is customary in logical research and is indeed quite natural in the context of any attempt to develop precise criteria for certain logical distinctions. But it does not by itself suffice to overcome the specific difficulty under discussion. For while it is now readily possible to characterize as not purely qualitative all those among the defined predicates in *L* whose definiens contains an essential occurrence of some individual name, our problem remains open for the primitives of the language, whose meanings are not determined by definitions within the language, but rather by semantical rules of interpretation. For we want to permit the interpretation of the primitives of *L* by means of such attributes as blue, hard, solid, warmer, but

²⁶ That laws, in addition to being of universal form, must contain only purely universal predicates was clearly argued by Popper ([Forschung], sections 14, 15).—Our alternative expression "purely qualitative predicate" was chosen in analogy to Carnap's term "purely qualitative property" (cf. [Application]).—The above characterization of purely universal predicates seems preferable to a simpler and perhaps more customary one, to the effect that a statement of the meaning of the predicate must require no reference to particular objects. For this formulation might be too exclusive since it could be argued that stating the meaning of such purely qualitative terms as "blue" or "hot" requires illustrative reference to some particular object which has the quality in question. The essential point is that no one specific object has to be chosen; any one in the logically unlimited set of blue or of hot objects will do. In explicating the meaning of "taller than the Eiffel Tower", "being an apple in basket *b* at time *t*", "medieval", etc., however, reference has to be made to one specific object or to some one in a limited set of objects.

not by the properties of being a descendant of Napoleon, or an arctic animal, or a Greek statue; and the difficulty is precisely that of stating rigorous criteria for the distinction between the permissible and the non-permissible interpretations. Thus the problem of setting up an adequate definition for purely qualitative attributes now arises again; namely for the concepts of the metalanguage in which the semantical interpretation of the primitives is formulated. We may postpone an encounter with the difficulty by presupposing formalization of the semantical meta-language, the meta-meta-language, and so forth; but somewhere, we will have to stop at a non-formalized meta-language, and for it a characterization of purely qualitative predicates will be needed and will present much the same problems as non-formalized English, with which we began. The characterization of a purely qualitative predicate as one whose meaning can be made explicit without reference to any one particular object points to the intended meaning but does not explicate it precisely, and the problem of an adequate definition of purely qualitative predicates remains open.

There can be little doubt, however, that there exists a large number of property and relation terms which would be rather generally recognized as purely qualitative in the sense here pointed out, and as permissible in the formulation of fundamental lawlike sentences; some examples have been given above, and the list could be readily enlarged. When we speak of purely qualitative predicates, we shall henceforth have in mind predicates of this kind.

In the following section, a model language L of a rather simple logical structure will be described, whose primitives will be assumed to be qualitative in the sense just indicated. For this language, the concepts of law and explanation will then be defined in a manner which takes into account the general observations set forth in the present section.

§7. Definition of law and explanation for a model language.

Concerning the syntax of our model language L , we make the following assumptions:

L has the syntactical structure of the lower functional calculus without identity sign. In addition to the signs of alternation (disjunction), conjunction, and implication (conditional), and the symbols of universal and existential quantification with respect to individual variables, the vocabulary of L contains individual constants (' a ', ' b ', ...), individual variables (' x ', ' y ', ...), and predicates of any desired finite degree; the latter may include, in particular, predicates of degree 1 (' P ', ' Q ', ...), which express properties of individuals, and predicates of degree 2 (' R ', ' S ', ...), which express dyadic relations among individuals.

For simplicity, we assume that all predicates are primitive, i.e., undefined in L , or else that before the criteria subsequently to be developed are applied to a sentence, all defined predicates which it contains are eliminated in favor of primitives.

The syntactical rules for the formation of sentences and for logical inference in L are those of the lower functional calculus. No sentence may contain free variables, so that generality is always expressed by universal quantification.

For later reference, we now define, in purely syntactical terms, a number of