CHAPTER ONE
Analysis and Subsumption

A major contention of this study is that psychological phenomena are typically not explained by subsuming them under causal laws, but by treating them as manifestations of capacities that are explained by analysis. Thus, a contrast between two explanatory strategies—*subsumption* and *analysis*—is central to what follows. Since the analytical strategy is by no means peculiar to psychology or to the social sciences,[1] and since it has generally been neglected by philosophers and methodologically minded scientists, it will be useful to begin with an abstract characterization of the analytical strategy and its relation to the more familiar subsumptive strategy.

In order to see this matter clearly, we need to distinguish between two kinds of theorizing, one of which customarily achieves its goals via causal subsumption, the other one via analysis.

1.1. TRANSITION THEORIES

Many scientific theories are designed to explain change. The point of what I call a *transition theory* is to explain changes of state in a system as effects of previous causes—typically disturbances in the system. The emphasis is on what will happen *when* (i.e., under what conditions). Subsumption under causal law is the natural strategy: one tries to fix on a set of state variables for the system that will allow one to exhibit each change of state as a function of a disturbing event and the state of the system at the time of the disturbance. A transition law therefore requires a systematic way of representing the states of the target system S, and a systematic way of transforming these representations such that, given a representation R of S at t, the (or a) transformation of R will represent S at t′. We can picture the situation this way:

$$\begin{tikzcd} & R_s(e) \arrow[r, "T_s"] & T_s(R_s(e)) = R_s(e') & \\ {} \arrow[r, squiggly] & e \arrow[u, "R_s"] \arrow[r, squiggly] & e' \arrow[u, "R_s"] \arrow[r, squiggly] & {} \end{tikzcd}$$

(events in S)

The wavy line represents the temporal sequence of events in S. R_S is a function that maps events in the system S onto the canonical representations licensed by the theory, and T_S is a function that maps the domain of R_S into itself.

A venerable tradition to which I subscribe holds that transition theories are not genuinely explanatory unless the laws appealed to are causal laws—i.e., laws that subsume cause-effect pairs[2]—for the goal of such theories is to explain changes as effects. Subsumption under a generalization that is not causal merely summarizes our reasons for believing the change would occur—it justifies our expectations perhaps—but it doesn't explain why the change occurs.[3] Certain cases of subsumption under noncausal generalization are admittedly called explanations, and I have no desire to arbitrarily restrict use of the word. But it is important to see that the goals of a *transition theory* will not be satisfied by noncausal subsumption. Hence, transition theories are fairly criticized as nonexplanatory when noncausal generalizations are substituted for the genuine article. Throughout this book, I shall mean by "subsumption" *causal subsumption.* Causal subsumption is not, of course, the only sort of explanation that involves logical derivation from laws, but it is the only sort of derivation that achieves the goal of what I am calling transition theories.

The causal character of many standard subsumptive explanations is obscured by the equational form of most physical laws. For example, the pendulum law that tells us that the period of a pendulum, T, is equal to 2π times the square root of the pendulum's length, lh, divided by the constant of gravitation, g.

$$(1)\quad T = 2\pi\sqrt{lh/g}.$$

Here we have no obvious reference to events, nor do standard textbook applications of the law involve such references. Philosophers add to this obscurity when they give the following as an instance of a D-N (deductive-nomological) "explanation."

$$\begin{array}{ll} (2) & T = 2\pi\sqrt{lh/g}\,; \\ & \underline{lh = 2\text{ ft}\;;} \\ & T = 1.570\text{ sec}\,. \end{array}$$

Argument (2) is typically offered as a genuine explanation in contrast to (3) which, though it fits the usual D-N schema, is said to be nonexplanatory.

$$\begin{array}{ll} (3) & T = 2\pi\sqrt{lh/g}\,; \\ & T = 1.570\text{ sec}\;; \\ & lh = 2\text{ ft}\,. \end{array}$$

(2) is univerally preferred to (3) on the grounds that the length isn't two feet because the period is $\pi/2$ seconds: length causes period, not vice versa.

Anyone who shares this and like intuitions should, it seems, abandon the Humean doctrine that to be causally connected is simply to be an instance of a nomological generalization, for (1) will not distinguish between the claim that length causes period and the claim that period causes length. It is useless to reply that once (1) is rewritten as a statement subsuming cause-effect pairs (as below), only the former connection will appear as an instance. That is true, as we shall see shortly, but it begs the question: the decision to represent changes in length in the "cause" slot rather than in the "effect" slot is motivated solely by a prior conviction that changes in length cause changes in period, and not vice versa. Equation (1) exhibits no such asymmetry. A temporal priority condition will not help either: the changes are concomitant. Nor, finally, can we resort to the fact that we can alter ("control") the period by altering (controlling) the length, for this is just a misleading way of saying that the change in length causes the change in period, and not vice versa.

As they stand, it seems to me that (2) and (3) are, in a crucial sense, on a par: neither is a causal explanation. Both are acceptable "problem solutions" ("Given T, find lh; given lh, find T"), hence explanations of why one thinks (or should think) that T = 1.570 sec. or lh = 2 ft. But if we mean by a causal explanation something that explains an effect by citing its cause, then neither (2) nor (3) is a causal explanation, for neither has an effect as explanandum, and neither invokes a causal law. Compare (2) and (3) with (4) and (5):

(4) A change of u to v in lh
causes a change of
$2\pi\ (\sqrt{v/g} - \sqrt{u/g})$ in T;[4]
lh increased from 1 to 2 ft;

T increased from 1.110 sec
to 1.570 sec .

(5) A change of u to v in T
causes a change of
$gv^2/4\pi - gu^2/4\pi$ in lh;[4]
T increased from 1.110 sec
to 1.570 sec ;

---

lh increased from 1 to 2 ft.

Notice that the information in the equational form of the simple pendulum law is neutral as between a construal that has changes in length causing changes in period and a construal that has changes in period causing changes in length. Since whatever supports the equation supports both construals equally, it follows that the equation is not a causal law but, as it were, an abstraction from the causal facts that make it true. This is why schemata like (2) tend to obscure the causal element in the explanation of a state transition: no causal element is present in (2), and it is precisely the reference to causes that is essential to the explanation of a state transition as an effect. (4), on the other hand, is a literal case of causal subsumption, with a change in T explained as an effect of a change in lh. Replacing the causal law in (4) with a noncausal generalization—e.g., an equation—unhinges the explanation while leaving the calculation (the transformation of representations) intact.

Causal laws—laws that do subsume cause-effect pairs—have two roles within transition explanations, both surprisingly limited in explanatory power.

First role: explanation of individual events. Individual events are explained by appeal to individual causes. The law may provide justification, since whatever reasons we have to accept the law will support our choice of cause/effect description in a particular case. Sometimes, especially when these descriptions are quantitative or heavily theoretical, we would not accept the explanation, let alone arrive at it, without this sort of support. But laws are not essential to the causal explanation of individual events. We can often causally explain an individual event with justified confidence even though no subsuming law is known. In any case, we should distinguish the explanation from its justification. Since causal laws play only a justifying role in the explanation of individual effects, they are not the source of explanatory force in such cases.

Often we need no guidance, and theory provides none, in finding causes. Theory will not help choose between (4) and (5)—Newtonian mechanics will simply yield the equation which is neutral between them—nor is it needed to explain why the window broke when I missed the nail and hit the pane with a hammer.

Of course, theory *is* required to explain why strikings cause breakings in glass but not putty. Hence, causal laws can be explananda. We could choose to say, with Hempel (1948), that we haven't "completely explained" a particular breaking by citing its cause unless we also explain how/why breakings—hence this breaking—are caused by strikings—hence by this striking. This would be all right were it not for the fact that this way of speaking tempts one to suppose that we don't "really know" what caused the breaking in the absence of this deeper theory. We are likely to be misled in just this way, however, for it is tempting to move from (a) we haven't completely explained the breaking, to (b) we don't really know why it occurred, to (c) we don't really know what caused it. It is only because we do know what caused it that we want a theory to explain the causal link!

Second role: explanation of event-types. More important than the explanation of individual events, which is seldom of scientific interest, is the explanation of event-types. A transition theory provides a recipe for explaining events of the types specified in its effect-descriptions. We have an especially important case of this when a transition theory for a system S can be derived from a transition theory for a more general type of system S* of which S is a special case. But even this last sort of explanation is limited in power. A transition theory for a system S simply specifies a dispositional property of S. Hence, derivation of a transition theory T_s for S from a transition theory T_{s*} for S* exhibits the dispositional property specified by T_s as a special case of the dispositional property specified by T_{s*}. It allows us to see manifestation of the T_s disposition as manifestation of the T_{s*} disposition.

Dispositions want explaining for reasons made famous if not clear by Molière in *Le malade imaginaire.* Asked why opium puts people to sleep, Molière's doctor replies that opium has a *virtus dormitiva.* The prospect of having to take seriously at the most fundamental level what we regard as a joke elsewhere is sufficiently unwelcome to lead us to ask whether there is some explanatory strategy to exploit other than more and more general subsumption of one transition theory to another. In particular, we are led to wonder whether there is some nonregressive way to explain the dispositional properties of a system, for transition theories only *specify* dispositions; they do not explain them. Not surprisingly: transition theories explain events or event-types, and dispositions are not events or event-types. The strategy of causal subsumption cannot stand alone.

Not all laws are causal laws, of course, and even causal laws play roles other than the roles they play in transition theories. To provide perspective, it will be useful to list and briefly identify a number of other sorts of laws, sorts that will loom large in my discussion of the analytical strategy.

1. CAUSAL LAWS. These are nomic[5] correlations whose instances are cause-effect pairs. (4) is an example. In addition to the two roles just discussed, causal laws define—or, better, *specify*—dispositional properties of the systems whose state transitions they subsume. They are thus candidate explananda of theories that seek to explain the properties (as opposed to the transitions) of those systems.
2. NOMIC CORRELATIONS. An example is the law correlating

thermal and electrical conductivity. Nomic correlations have no explanatory role at all, contrary to what is predicted by the deductive-nomological model of explanation. But they are important nevertheless. First, they serve as predictive rules; they justify expectations by summarizing inductive evidence. As such, they are important to science mainly as aids to experimental design, including devising tests of an uncontrolled sort, as in astronomy. Second, nomic correlations are among the facts that explanatory theories must explain. The law correlating thermal and electrical conductivity is striking and important not as a candidate explanans, but as a candidate explanandum.

3. NOMIC ATTRIBUTIONS. These are predications, lawlike statements to the effect that all x's have a certain property P. An example is the statement that photons have gravitational mass. The law of inertia appears to be a nomic attribution in classical mechanics, and the law of gravitation appears to be a nomic attribution in the theory of general relativity. The explanatory role of nomic attributions, along with the roles of laws in the next two categories, will be taken up in the next section.
4. INSTANTIATION LAWS. These are lawlike statements specifying how a property is instantiated in a specified type of system. An example is the statement that temperature is instantiated in a gas as the average mean kinetic energy of the molecules in the gas.
5. COMPOSITION LAWS. These are lawlike statements specifying the (or an) analysis of a specified type of system. An example (greatly simplified) is the statement that water molecules are made of two hydrogen atoms ionically bonded to one oxygen atom. The most celebrated recent example is the double helix model of DNA.

One of the more unfortunate consequences of the dominance of the deductive-nomological model of explanation is that it has focused attention on causal laws and their associated explanatory roles, i.e., on transition theories, to the exclusion of more important alternatives. This consequence is not surprising: subsumptive explanation of the sort featured in transition theories is the only sort of explanation for which the model is initially plausible. Analytical explanation, as we shall see shortly, can be given a deductive-nomological format, but the result is completely uninformative, obscuring rather than illuminating the nature of analytical explanation. It is only in the case of causal subsumption that the model provides any hint as to the nature of explanation, and that hint is limited by failure to take causation explicitly into account (and the consequent failure to distinguish (2) and (3)).[6]

Philosophers and methodologically minded scientists have focused their attention on transition theorizing. As a result, the methodology of causal subsumption has received a great deal of attention from scientists and philosophers, and is therefore relatively well understood. As I use the term, a methodology is a set of adequacy conditions on the application of an explanatory strategy. One has a methodology for an explanatory strategy when one has a set of principles for distinguishing legitimate from illegitimate applications of that strategy. A methodology is thus naturally construed as a canon of criticism. From this point of view, it is useful to think of a methodology as having two parts corresponding to two sorts of critical questions we can raise about a particular application of an explanatory strategy: (i) Could the strategy have any explanatory force in such an application? (ii) What sorts of evidential considerations would tend to support or undermine an application of the strategy in this case? Chapter V, section 1, illustrates the methodology of causal subsumption in action against introspective psychology.

A familiar empiricist ploy is to move from (ii) to (i) by arguing that a particular application could have no explanatory force because nothing could support or undermine it. No doubt an explanation is unacceptable if it is untestable, but this is quite a different sort of unacceptability than that intended by (i). Moving from (ii) to (i) has been a serious source of misunderstanding about explanatory force, for it encourages us to confuse the truth of a theory, or its justification, with its explanatory value, thus obscuring the very real difference between description and explanation. It is easy to overlook the fact that a theory with no chance of being true can have much greater explanatory force than a well-confirmed theory about the same phenomena. Contrast, for example, the theory that molecules are held together by "hooks and eyes" with the theory that molecules are held together by some unknown force. By the mid-nineteenth century the former theory was strongly disconfirmed, while the latter theory turned out to be true (because molecules *are* held together by a force unknown at the time). Still, the "hook-and-eyes" theory is at least capable of explaining such facts as the occurrence of H_2O but not H_4O, and the occurrence of H_2—e.g., on the hypothesis that each hydrogen atom has one free hook, and each oxygen atom has two free eyes. The "unknown force" theory, however, has no explanatory force whatever, being equivalent to the theory that the molecules that do occur do occur and the ones that don't don't. So explanatory force is quite independent of confirmation and truth.

Perhaps "explanatory potential" would be a better phrase than "explanatory force" or "explanatory value," for saying that T has explanatory potential leaves the door open for the remark that explanatory potential is proportional to how explanatory T would be if it were true. But I'm inclined to take a hard line on this largely verbal issue: false theories can explain things (and have); it's just that the explanations they provide are not true. It seems to me quite mistaken to suppose that one cannot understand something if one has false beliefs about it. The doctrine that only truth can explain would surely rule that much contemporary theory is not explanatory!

It is arguable that an explanatory theory must be testable: if T is explanatory, it must have implications vis-à-vis its explananda, and observations relevant to the latter surely constitute a test of T. Something like this is probably correct, but it establishes only an uninformative connection between testability and explanatory force. Untestable theories fail to be explanatory not because they have no explananda, but because they say the wrong sort of thing about the (purported) explananda that they *do* have. The assumption to avoid is that a theory need only be well confirmed (or true) to explain its target phenomena.

The Methodology of Causal Subsumption

There is, among philosophers at any rate, no generally received account of how transition theories are to be tested, let alone a generally received account of their explanatory force. Still, there are common elements discernible in the many attempts in the tradition of Bacon, Hume, and Mill to articulate canons of "scientific method." These common elements correspond to uniformities in scientific practice, uniformities that evidence agreement on, and adherence to, a tacit methodology that Empiricist writers have sought to make explicit. When one canvasses methodological "casuistry," especially methodological chapters of undergraduate textbooks, one finds, I think, that the great bulk of this effort derives from three fundamental concerns: (i) *causality:* many of the rules and methods are designed to distinguish causal connections from mere correlations, whether these be lawlike or simply accidental; (ii) *determinism:* theory construction is constrained by the idea that one must not countenance uncaused or idle events; (iii) *justification* (Hume's condition): a program of research is regarded as fairly subject to criticism if the causes and effects it appeals to cannot be observed or measured independently of each other. I will briefly discuss each of these. The reader should keep in mind, however, that the point of this exercise is not to advance the discussion of the methodology of causal subsumption, but rather to make plausible the claim that there are commonly accepted canons of "scientific method" (however they should be formulated), and that these are, in fact, canons of causal subsumption.

(i) *Causality.* Does smoking cause cancer, or are smoking and cancer only highly correlated? In introductory social science classes, this example is as familiar as are deductions of Socrates' mortality in introductory logic classes. The practical implications of the problem are obvious: only the causal claim provides an immediate reason to give up smoking. But it is not the practical implication that the example is designed to bring out, but a methodological moral: only the causal claim has the sort of explanatory force for which good theory is valued. Smoking may be a good predictor of cancer and yet tell us nothing about why people get cancer. Hence the standard emphasis on control: if we can control cancer by controlling smoking, we have in smoking more than a good predictor of cancer.

The concern to distinguish genuine causal connections from other correlations is, as I noted above, central to good transition theory, since the point is to explain events as effects of causes. The universality of this concern is therefore a measure of the extent to which scientific methodology was identified with the methodology of causal subsumption. Nowhere is this concern more apparent than in Mill's famous discussion of the "four methods of experimental inquiry" (Bk. III, ch. viii). Like Hume before him, Mill assumes that scientific reasoning—indeed all *a posteriori* reasoning—is causal reasoning. He therefore never seriously questions the assumption that to explain a phenomenon is to subsume it under causal law. At the beginning of the famous chapter entitled "Of the Four Methods of Experimental Inquiry" (in *Philosophy and Scientific Method*) Mill writes:

> The simplest and most obvious modes of singling out from among the circumstances which precede or follow a phenomenon those with which it is really connected by an invariable law are two in number. One is by comparing together different instances in which the phenomenon occurs. The other is by comparing instances in which the phenomenon does occur with instances in other respects similar in which it does not. These two methods may be respectively denominated the method of agreement and the method of difference.
>
> In illustrating these methods, it will be necessary to bear in mind the two fold character of inquiries into the laws of phenomena, which may be either inquiries into the cause of a given effect or into the effects or properties of a given cause.

According to Mill, there are, at bottom, just four methods of experimental inquiry, and all four are concerned with the problem of distinguishing genuine causal connections!

The same concern apparently dominates Hume's work: how are we to tell genuine causal connection from mere constant conjunction? Mill's four methods can be thought of as an expansion of Hume's fourth rule for judging of causes and effects (*Treatise of Human Nature,* I.III.xv): "The same cause always produces the same effect, and the same effect never arises but from the same cause. This principle we derive from experience, and is the source of most of our philosophical [i.e. scientific] reasonings." Indeed, Hume's rules five through eight essentially embody Mill's four methods, and Hume claims that his fifth and sixth rules, at least, are corollaries of rule four. After stating the eight rules by which to judge of causes and effects, Hume remarks, "Here is all the Logic I think proper to employ in my reasoning . . . ," thus confirming (what is obvious anyway) that the rules by which to judge of causes and effects exhaust, for Hume at least, scientific methodology.

(ii) *Determinism.* Refusals to countenance uncaused or idle events in the theoretical treatment of a given domain are, perhaps, familiar enough in explicit methodological writings, but in that context such refusals have always been controversial. In actual practice, however, a theory is always subject to criticism on the grounds that it allows for uncaused or idle events. The motivation here is that a theory designed to explain events as effects of causes is in trouble if it countenances uncaused events or ineffectual causes. (See Cummins, 1976, for further discussion of this principle in the context of a particular historical application.)

A classic example is the critique of behaviorist theories of language on the grounds that linguistic behaviors—i.e., distinct linguistic performances or "outputs"—cannot be predicted on the basis of environmental stimuli. (See, e.g., Chomsky, 1959a, and Fodor, Bever and Garrett, 1974.) Critics and advocates alike agree that a one-many mapping of stimuli to linguistic responses would, if it is the best we can do, refute stimulus-response accounts of linguistic behavior.

There are, of course, cases in which theoretical allowance for uncaused or idle events has not been thought crippling. The quantum theory, on certain interpretations, allows for uncaused events, and epiphenomenalism, popular in psychology in the early part of the century (and in some quarters still), allows for idle events. But both of these cases are controversial, and they are controversial, moreover, precisely because they countenance uncaused or idle events. The controversy generated by these cases is itself a measure of the extent to which the methodology of science *is,* in most minds, just the methodology of causal subsumption. Uncaused or idle events are explanatory lacunae in a transition theory, and since all theory tends to be seen as transition theory, uncaused or idle events are generally regarded as explanatory failures *tout court.* Thus both advocates and critics regard uncaused and idle events as embarrassments to the theories that allow them.

(iii) *Justification (Hume's condition).* Hume is unjustly famous[7] for the idea that causal claims are unjustified if there is no way to observe or measure the causes independently of the effects they are supposed to produce, and *vice versa.* In Chapter IV I show how this principle was used to great effect against the introspectionist psychology of Wundt and Titchener by their behaviorist critics.

Unlike the substitution of a noncausal correlation for a causal law, or theoretical allowance for uncaused or idle events, violation of Hume's condition does not strike directly at the explanatory power of a transition theory, but it does strike directly at its acceptability. Violation of Hume's condition is generally held to strike indirectly at the explanatory power of a theory, however, and if we examine the reasoning underlying this thought, we see once again that the assumed explanatory strategy is the strategy of causal subsumption.

It is often supposed that galvanic skin responses (gsr—the sort of thing measured by a "lie detector") provide a good measure of anxiety. Indeed, anxiety is sometimes "operationally defined" in terms of gsr. The motivation for this definitional maneuver is this: we cannot really explain gsr by appeal to anxiety because we cannot measure anxiety directly, but must have recourse to gsr or something comparably "external"—hence we shouldn't distinguish anxiety from gsr and the like, for this only encourages "pseudo-explanations." This line of thought is ubiquitous in the behavioral sciences. A (perhaps *the*) classic example is Skinner (1953). If we ask what underlies this line of reasoning, it seems clear that the following assumption is being made: to explain gsr means to explain changes in skin conductivity as effects of some cause. Given this assumption, Hume's condition applies directly: we cannot attribute these effects to changes in anxiety level, for, given that these very effects are our only way to observe or measure changes in anxiety level, we cannot justify the resulting causal claims. Hence, any attempt to explain gsr in terms of anxiety will be vitiated by appeal to an unjustified causal claim.

Whatever one may think of this reasoning, I don't see how one could understand it except on the assumption that all explanation is causal subsumption. One might suppose that it is only being assumed that the sort of explanation at issue in *cases of this sort* is causal subsumption, but in fact this is never even stated, but taken for granted. It could be, of course, that everyone assumes that the sort of explanation in question is causal subsumption because it is *obvious,* and not because there are no alternatives to rule out, but I don't think this can be right: the behaviorist conclusion is so controversial that the assumption would have been widely denied had there been any well-articulated alternative available. But the assumption (viz., that it is causal subsumption that is at issue) has not been widely denied and even when it has been denied, it was not denied because there is some widely recognized alternative explanatory strategy to assert in place of the strategy of causal subsumption. Some philosophers, of course, have argued that mental explanation is not a species of causal explanation, but a kind of rationalization. (See, for example, Meldin, 1961, and Ryle, 1949.) But no one, I think, has taken this line about anxiety, and in any case the doctrine that reasons are not causes has fallen on hard times. (See Davidson, 1963, Alston, 1967, Brandt and Kim, 1963.) What's more, it has fallen on hard times mainly because it seems clear that reasons do explain actions, and the underlying assumption here, once again, is surely that explanation is causal subsumption.

The concern to distinguish causal laws from noncausal correlations, to shun uncaused or idle events, and to make provision for independent access to causes and effects are, of course, not the only methodological concerns to manifest themselves in scientific practice and in writings on scientific method, but they are, perhaps, the most fundamental and pervasive. The universality of these concerns, and their status as *the* fundamental concerns, leaves little doubt that scientific methodology has often been identified (tacitly or explicitly) with the methodology of causal subsumption. It should become clear shortly, however, that these concerns are simply out of place in the context of property theories and the analytic strategy of explanation. This is, I think, something of a scandal: the analytic strategy is as old as atomism, yet its methodology is only now beginning to receive serious attention. Give or take a nicety of formulation, the canons of causal subsumption are widely recognized and honored; they have the status of truisms. None of this can be said concerning the methodology of analysis.

I.2. PROPERTY THEORIES

Many scientific theories are not designed to explain changes but are rather designed to explain properties. The point of what I call a property theory is to explain the properties of a system not in the sense in which this means "Why did S acquire P?" or "What caused S to acquire P?" but, rather, "What is it for S to instantiate P?", or, "In virtue of what does S have P?" Just as we can ask, "Why did the gas get hotter (or expand)?", we can ask, "In virtue of what does a gas have a temperature (volume)?" Understood as an answer to the latter questions, the kinetic theory of heat (and the molecular theory of gases that it presupposes) is not a transition theory but a property theory: it explains temperature in a gas by explaining how temperature is instantiated in a gas; it does not, by itself, explain changes in temperature.

Many of the most pressing and puzzling scientific questions are questions about properties, not about changes. We know a lot about what causes pain, but there is no very good theory of how pain is instantiated. Good property theories are wonderfully satisfying: we know how temperature is instantiated, how inheritance is instantiated, how electricity is instantiated, how solubility is instantiated. I think we are close to knowing how life and intelligence are instantiated, though we are a long way from understanding how consciousness or intentionality are instantiated.

The characteristic question answered by a transition theory is: Why does system S change states from s-1 to s-2? The characteristic question answered by a property theory is: What is it for system S to have property P?[8]

The natural strategy for answering such a question is to construct an analysis of S that explains S's possession of P by appeal to the properties of S's components and their mode of organization. The process often has as a preliminary stage an analysis of P itself into properties of S or S's components. This step will loom large when we come to discuss complex dispositional properties such as information-processing capacities. Analysis of a *system* will be called compositional analysis, to distinguish it from analysis of a *property,* which will be called functional analysis when the property is dispositional, and property analysis when the property is not dispositional. Analysis is "recursive," since a given analysis may appeal to properties or components that themselves require analysis.

Historically, the most important property theories are applications of the doctrine of atomism. In its simplest form, atomism is the claim that all physical objects are collections of elementary parts (a part being elementary if it has no theoretically relevant parts itself). The explanatory interest of this doctrine derives from the further claim that the properties of every object are determined by its microconstitution—i.e., by the properties of its elementary parts and the way those parts are "put together" to constitute the object in question. Thus atomism promises to explain the properties of an object by exhibiting its elementary part constitution—e.g., the shape and density of a crystal is explained by the relative positions of its elementary parts, which, in turn are fixed by the properties of those parts; the temperature of a gas is explained by identifying it with the average kinetic energy of its molecules. Chemistry texts detail an enormous number of applications of the atomist strategy.

Atomism explains property instantiation in S by appeal to property instantiation in S's elementary parts. A crucial assumption of the theory is that all qualitative change is compositional change. Thus an elementary part cannot change in its categorical properties (though it might be created or destroyed); hence no pressure is generated to explain how an elementary part acquires its categorical properties. An elementary part cannot be the object of a transition theory.[9] It might seem, however, that questions about instantiation could arise at the level of elementary parts—e.g., what is it for an elementary part to have a spherical shape? But the question is actually quite different at this level. Whereas we can explain how something made of cubes can be spherical, the corresponding question about an elementary part can only be met with a definition of "spherical"—e.g., surface everywhere equidistant from a single point. It isn't that we cannot *explain* what makes an elementary part spherical: nothing *makes* it spherical. There is just nothing to be explained here, or, rather, there is only a concept to be explained here, not a property.[10]

Thus atomism envisages an end to the explanatory regress by arriving at a case in which state transition does not occur and property instantiation requires no theoretical treatment. The fundamental laws are nomic attributions. This is easily overlooked if one is working within the hypothetical-deductive tradition, for that tradition assimilates explaining p to providing a "scientific" justification for p. Since we can justify any property attribution "scientifically"—e.g., by appeal to inductive evidence or theoretical derivation—it will seem from within the H-D tradition that all property attributions are candidate explananda. This is mistaken: fundamental nomic attributions require justification but not explanation. Hempel (1966) pointed out that "narrow inductivism" fails to distinguish theory construction from theory testing. Since testing is inductive, narrow inductivism left no room in science for theories that were not generalizations of the data. We now require a comparable distinction between theory testing and theoretical explanation. To assimilate the logic of explanation to the logic of testing leaves us no conceptual space to delineate the differences between laws that do and laws that do not require explanation, for all laws require justification. It also forces us to see the explananda of a theory as (among) the data that support it. The pernicious effects of this last consequence are illustrated in the discussion of Clark Hull in section IV.2, below.

Successful analysis yields an explanatory payoff when we come to see that something having the kinds of components specified, organized in the way specified, is bound to have the target property. Unfortunately, the fact that what we come to see is a generalization of this sort encourages an assimilation of the analytical strategy to the subsumptive strategy, for it provides what looks like a deductive nomological schema:

(6i) Anything having components $C_1 \ldots C_n$ organized in manner O—i.e., having analysis $[C_1 \ldots C_n, O]$—has property P;
(6ii) S has analysis $[C_1 \ldots C_n, O]$;

(6iii) S has property P.

There is nothing wrong with representing analytical explanations in this format provided we avoid the assumption that the explanation is achieved by a state-transition type of subsumption of (6iii) under (6i) given (6ii). Assimilating the analytical strategy to the subsumptive strategy obscures the difference between explaining changes and explaining properties and thereby leads one to misapply the methodology of causal subsumption to analysis.

Let us call "laws" of the form (6i) "instantiation laws," and laws of the form (6ii) "composition laws."[11] We can begin to see the difference between instances of the schema (6) and causal subsumption [e.g., (4)] by noting that instantiation laws are not causal laws at all, for they do not have cause-effect pairs as instances. Thus the role of an instantiation law cannot be to explain S's possession of P as an effect of some cause. Indeed, laws of instantiation need not even be empirical. If the analysandum is a symbolic capacity—a capacity to manipulate symbols—the instantiation law will typically be a bit of mathematics. For example, "Anything executing the bubble algorithm sorts numbers into order; S executes the bubble algorithm; hence S sorts numbers into order."

Instantiation laws are derived principles—or they *should* be in a full-dress theory—for they obviously call for explanation themselves. There is really only one available strategy for explaining an instantiation law: it must be derived from laws specifying the properties of the components. If we can carry out this derivation, our schema looks like this:

(6a) The properties of $C_1 \ldots C_n$
are <whatever>, respectively;
(6ii) S has analysis $[C_1 \ldots C_n, O]$;

(6iii) S has property P.

Instances of this schema will be valid when we can derive (6i) from (6a). When we can do this, we can understand how P is *instantiated* in S. Laws like (6a) I call *nomic attributions,* to emphasize that they do not state correlations or causal connections, but attribute properties.[12]

The most interesting property theories are aimed at explaining dispositional properties. To attribute a disposition d to an object x is to assert that the behavior of x is subject to (exhibits or would exhibit) a certain lawlike regularity: to say x has d is to say that x would manifest d (dissolve, shatter) were any of a certain range of events to occur (x is put in water, x is struck sharply). The regularity associated with a disposition is a regularity that is special to the behavior of a certain kind of object and obtains in virtue of some special facts about that kind of object. Not everything is water-soluble: such things behave in a special way in virtue of certain (microstructural) features special to water-soluble things. Thus it is that dispositions require explanation: if s has d, then s is subject to a law or regularity in behavior special to things having d, and such a fact needs to be explained. When we discover that not everything *is* water-soluble, we are led to ask why the things that *are* dissolve in water, while other things do not. To explain a dispositional regularity, then, we must explain how or why manifestations of the disposition are brought about given the requisite precipitating conditions.

As an example of a property-theoretic explanation, consider Einstein's explanation of the photoelectric effect. When light shines on a metal surface, electrons are emitted from the surface. Five facts are central to the specification of this dispositional property of light.

a. The number of electrons emitted per unit time is proportional to the intensity of the incident light.

b. The kinetic energy of an emitted electron does not depend on the intensity of the incident light.

c. For a given metal, the maximum kinetic energy of an emitted electron is solely a function of the frequency of incident light.

d. For a given metal, no electrons are emitted if the incident light is below a certain threshold frequency.

e. Electrons are emitted immediately, regardless of the intensity of incident light.

Facts (b) through (e) are radically at odds with classical conceptions according to which (i) the energy of a light beam is continuously distributed throughout the beam, and (ii) a beam of light may have any amount of energy. For example, (i) implies that a surface irradiated by a very weak beam (say, 10^{-10} w/m^2) should take years to accumulate enough energy to release electrons, contrary to (e). (ii) implies that the kinetic energy of emitted electrons is a function of the intensity of incident light, contrary to (b). (c) and (d) are simply "danglers": classical theory provides no link between frequency and available energy to explain (c) and (d).

Einstein explained the photoelectric effect by assuming that light consists of discrete photons, and that the energy of a photon is quantized, its energy being hv, where h is Planck's constant, and v is the frequency. These assumptions allowed Einstein to treat the photoelectric effect as a straightforward case of energy conservation.

Each photon interacts with at most one electron, providing that electron with as much energy as it will ever get from the beam. Since this is a fixed amount, given by the frequency, the electron will either be emitted immediately (e), or not at all (d), with a maximum kinetic energy less than hv (c). Increase in intensity or duration will increase the number of photons, hence the number of emitted electrons (a), but it will not affect the energy of the photons in the beam, hence will not affect the kinetic energy of emitted electrons (b). Here, we have a simple and elegant explanation of a dispositional property of light—the photoelectric property—a property that is completely mysterious from the point of view of classical wave mechanics. The explanation can be conceived in stages. *Stage one,* corresponding to (6ii), is an analysis of the incident light beam into a stream of particle-like photons. *Stage two,* corresponding to (6a), introduces two fundamental nomic attributions governing the photons: their energy is quantized (NA_1), and is directly proportional to frequency (NA_2).[13] *Stage three* consists in deriving the facts (a) through (e), characterizing the photoelectric property from the assumptions of stages one and two. This evidently amounts to deriving an instantiation law for the photoelectric effect: A system consisting of a light beam—i.e., consisting of photons satisfying (NA_1) and (NA_2)—incident to a metal surface will satisfy (a) through (c).

As a second illustration of analytical explanation, consider the standard derivation of Archimedes' Principle, which specifies the dispositional property of liquids to exert a force on submerged objects in the direction of the surface and equal in magnitude to the weight of displaced liquid.

Consider an object, O. Let v be an arbitrary volume of the liquid having the same size and shape as O. The volume is at rest with respect to the remainder of the liquid, hence the net force on v is zero. Let W_o and W_v be the weights of o and v, respectively. The net force NF_v on v is the upward force UF_v minus the downward force $DF_v = W_v + W_c$, where W_c is the weight of the column of liquid above v:

$$NF_v = UF_v - (W_v + W_c) = 0.$$

Hence,

$$UF_v = W_v + W_c.$$

Now imagine O in place of v. Evidently

$$UF_o = UF_v = W_v + W_e.$$

The force of LF_o exerted by the liquid on O is $UF_o = W_c$. But $LF_o = UF_o - W_c$, so,

$$LF_o = (W_v + W_e) - W_e = W_v.$$

Since we have assumed that the direction of the surface is positive, the fact that LF_o is positive indicates that it is in the direction of the surface.

Here we explain Archimedes' property of liquids by assuming (i) that a liquid consists of a collection of parts (or volumes) free to move with respect to one another, (ii) that a part (or volume) at (relative) rest experiences a net force of zero, (iii) that weight is a downward force, and (iv) that forces are additive. (i) is a composition law. (iv) is a nomic attribution. (That is easier to see if we think of a force on x as a disposition of x to accelerate. It is then a property of masses that these dispositions can be (algebraically) summed.)[14] (iii) is an instantiation law imported from Newtonian mechanics, and (ii) follows from construing forces as dispositions to accelerate. (ii) and (iii) and a clear thought experiment (substituting o for v) suffice to establish an instantiation law: systems satisfying (ii) will have Archimedes' Property.

By now it should be clear that the same equation can appear as part of a property theory and a transition theory—e.g., as part of an instantiation law and as part of a causal transition law. Thus it should come as no surprise that focusing on the role of nomic equations in mathematical derivations should obscure rather than reveal their explanatory functions.

Property theories and transition theories fit together in an important way when target properties are dispositional, for when a system manifests a disposition, we have cause and effect (precipitating event causing manifestation), hence state transition. When we derive an instantiation law for a dispositional property from underlying nomic attributions, what we derive is a (minimal) transition theory for the disposed system—viz., the dispositional regularity. Since what we derive is a transition law for the system, the derivation will often (though not always) invoke transition laws governing components of S. The explanatory theory will not be "complete," of course, until we have discharged all the dispositions. Thus, as emphasized above, transition theories specify explananda for property theories. The pressure for explanation *increases* as more and more general transition theories are devised, for a very general transition theory specifies a dispositional property that a wide variety of systems have. A perfectly general transition theory is suspicious, for it would specify a disposition that every system has. In such a case we should ask whether the property specified is really dispositional—i.e., really supports a distinction between having the property and manifesting it. If not, the theory is not really a transition theory, but a set of nomic attributions. When we notice that everything is subject to gravitation, we are led to notice further that everything gravitates all the time. Thus the law of universal gravitation is a fundamental nomic attribution, not a transition law.

I.3. REDUCTION AND INSTANTIATION

It is important to distinguish the claim that a theory identifies instantiations from the claim that it licenses reductions. As I use the term, reduction requires that the true statements one can make about a domain in a vocabulary v can all be formulated in a different reducing vocabulary v′. For example, physicalistic reductionism in psychology is the claim that the truths of psychology are all formulatable in the language of physics. It is now a commonplace (I hope) that one can hold that everything is physical—has some physical description—without holding that everything worth saying in science can be said in the language of physics. (For example, see Fodor, 1975.) Thus, since systems not satisfying (i) could have Archimedes' Property as well as systems that do, we cannot *reduce* having Archimedes' Property to satisfying (i), for there are truths about Archimedes' Property that are not truths about satisfaction of (i), viz. that systems not satisfying (i) could have Archimedes' Property. I think a whole cluster of problems surrounding the issue of theoretical (property) identifications and reduction can be avoided by substituting the language of instantiation for the language of identity.[15]

It is now commonplace in the philosophy of mind to say that type-type identity theories fail because mental properties can be "realized" (i.e., instantiated) in more than one way. (The idea seems to have begun with Putnam, 1960.) The same point can be made about functional properties: adding is one sort of physical process in a mechanical calculator and another sort in a computer. And about some chemical properties: bonding is one sort of thing in H_2 and another sort of thing in NaCl. Hence, the advent of "token-identity" theories: bonding is identical with one thing in H_2—call it ψH_2—and with another in NaCl—ψNaCl. Is this reduction or not? It *seems* to be, for if bonding *is* ψH_2 (in H_2), then talk about bonds (in H_2) is just talk about ψH_2. But it can't be reduction, for there are truths about bonding that aren't truths about ψH_2 or any other physical kind. And isn't this a strange sort of identity? We have bonding = ψH_2 and bonding = ψNaCl but $\psi H_2 \neq \psi$NaCl. Even if bonding were physically homogeneous—just ψ every time—this would surely be a contingent fact, and hence our physics would license only a "contingent identity." There is, by now, a notorious controversy over whether this notion is coherent, and if it is, whether such an identity would license reductions.

We can begin to make progress in this matter by substituting the language of instantiation for the language of identity. Rather than saying that bonding is identical with one thing in H_2 and something else in NaCl, we can say that bonding is instantiated one way in H_2, another in NaCl. This avoids the logical muddle lately rehearsed while retaining the central point: multiple instantiation blocks reduction because there are truths about the instantiated property that are not truths about the instantiations.

But suppose there aren't multiple instantiations? Suppose bonding were always instantiated as, say, hook-and-eye connection? If we express this in the language of identity, we prejudice the issue of reduction: if bonding is everywhere identical with hook-and-eye connection, how could there be truths about bonding that weren't truths about hook-and-eye connection?

Well, perhaps the identity is "contingent." Then there will be modal truths about bonding that aren't truths about hook-and-eye connection—e.g., "Bonding *could be* something other than hook-and-eye connection." (I take it this is what saying the identity is "contingent" amounts to.) But, perhaps the notion of "contingent identity" doesn't make sense, as Kripke (1972) has argued. And if it doesn't, then it seems we cannot have theoretical identities without reduction.

Once again, identity is a red herring. The question is whether uniform instantiation amounts to reduction. Most of the familiar points carry over, but without the logical and metaphysical problems surrounding identity.

(i) Uniform instantiation is necessary for reduction, since diverse instantiation blocks it.

(ii) Uniform instantiation is not sufficient for reduction, since a variety of modal truths will hold of the target property that do not hold of the instantiation.

(iii) If uniform instantiation is not "accidental" but nomically necessary, then we do have genuine reduction.

A word of caution about (iii) is in order. We can have a law of the form:

(7) S has P iff S has analysis $[C_1 \ldots C_n, O]$,

and yet have no reduction, for something of the form (7) might be true and lawlike yet not give the instantiation of P. Consider (8) and (9):

(8) (E) (n) {E has a valence of n iff [(n is positive and E atoms have n free hooks) or (n is negative and E atoms have n free eyes)]}.

(9) (E) (n) {E has a valence of n iff [(n is positive and E atoms have a mean diameter of cn) or (E is negative and E atoms have a mean diameter of kn)]} (where c and k are distinct constants).

Assume for the sake of argument that (8) and (9) are both lawlike and true. (8) would license a reduction, but (9) would not, because (9) does not give a hint as to how valence is instantiated. To have a valence is to have a disposition to form chemical bonds, *which* disposition depending on *which* valence. Such a disposition could be instantiated as possession of free hooks or eyes but could not be instantiated as possession of a certain mean diameter. (9) would allow us to derive possession of P (a valence of +n) from the analysis (the element consists of atoms with a mean diameter of cn), but (9) itself would not (presumably) be derivable from fundamental nomic attributions, whereas (8) has a painfully obvious derivation (one hook to each eye, hence H_2O but not H_3O, etc.).

I know of no general criterion by which the reductive instances of (7) are to be distinguished from the nonreductive except this: the reductive instances specify how P is instantiated, the details being revealed by a derivation of the instantiation law from underlying nomic attributions. It seems obvious that (8) but not (9) is (or could be) reductive in this sense, because there is no route from attributions of diameter to bonding dispositions and hence (9), but there is a route from attributions of hooks and eyes to bonding dispositions and hence (8).

This seems obvious to me. But to many scientists brought up on a steady diet of hypothetical-deductivism, and to some philosophers of like history, this will seem question-begging at best: If (9) were true, well confirmed, and lawlike, why wouldn't it explain bonding?

(9) itself cries out for explanation: Why/how do different diameters lead to different bonding capacities? Perhaps when atoms bond, they fit into a kind of "frame"—a local property of the electromagnetic field, say—and "fit" is determined by diameter. An explanation of this sort would give (9) a status on a par with (8). Failure to explain (9) [or a nonreductive law used in "explaining" (9)] leaves us with a brute correlation—i.e., a correlation that holds for no reason at all (so far as our theory goes). This might be the best we could do, but to call it "explanation" (rather than "discovery of a brute correlation") is to make a virtue of necessity at the price of obscuring the very real difference in understanding achieved by deriving correlations from noncorrelational nomic attributions that are not themselves candidate explananda (though they require justification, of course).

Once it is clear that we can explain how a chemical property is instantiated in a physical system without identifying chemical and physical properties, the pressure to reduce chemistry to physics evaporates. Chemical properties need to be explained—presumably they don't figure in fundamental nomic attributions. If we mistakenly suppose that the only way to explain a property that doesn't figure in a fundamental nomic attribution is to identify it with a property that does, we will be committed to identifying chemical properties with physical properties. We have reduction just in case we have genuine property identification. Hence reduction will appear the only alternative to leaving chemical properties unexplained. But since we *can* explain how chemical properties are instantiated in a physical system without identifying chemical properties with physical properties, a physical explanation of chemical properties need not be reductionist.

I.4. CONCLUSION

In this chapter I have presented a characterization of two distinct explanatory strategies. And, although that characterization is rough and incomplete at many points, and probably just wrong in places, enough has been said, I think, to establish that the methodology of analysis is bound to be quite different from the methodology of causal subsumption. Indeed, we are now in a position to discern some rough analogues to the illustrative methodological canons mentioned in connection with causal subsumption. Corresponding to the basic requirement that subsumptive laws be causal, we have the requirement that instantiation laws should be derivable from nomic attributions specifying properties of components. Just as causal subsumption fails to get off the ground if the laws appealed to are not causal, so analysis fails to get off the ground if analyzing properties are not derivable from properties of the elements of the analyzed system, for in such a case we have no reason to think we have analyzed the target property as it is instantiated in the target system.

Corresponding to the requirement that transition theories must not countenance uncaused or idle events, we have the requirement (emphasized by Dennett, 1978, 123-124) that the analyzed property should not reappear in the analysis. Appealing to the analyzed property, or something comparable, in the analysis defeats the explanatory point of a property theory in the same way that uncaused or idle events defeat the explanatory point of a transition theory: in each case, the offending theory reintroduces the very thing it is supposed to explain.

Finally, corresponding to the principle that causes and effects must be observable or measurable independently of one another, we have the requirement that attributions of analyzing properties should be justifiable independently of the analysis that features them. If, for example, we analyze the capacity of a child to solve division problems into the capacity to copy numerals, to multiply, and to subtract, we must know, or be able to find out, that the child can copy numerals, multiply, and subtract without simply inferring this from the capacity to divide, and we must know, or be able to find out, that these capacities are in fact organized as the analysis specifies.

The trick to providing a good property theory is generally to manage to satisfy this requirement and the first one simultaneously. The hook-and-eye theory of chemical bonding satisfies the first requirement, but there is no other reason to believe that elementary parts (i.e., atoms, for the theory of bonding) have hooks and eyes. There is plenty of independent evidence that they have mean diameters, but no hint as to how diameters could produce bonds.

Sometimes the problem is making the first requirement mesh with the requirement that the analyzed property not reappear in the analysis. This is (notoriously) *the* problem in cognitive psychology: how to explain intelligence without recourse to equally intelligent components.

A full exposition and defense of the claims I have made here concerning analysis and subsumption would require a book by itself. My topic, however, is psychological explanation. This chapter has been included because it seems to me that most psychological explanation makes no sense when construed as causal subsumption but makes a great deal of sense construed as analysis. Hence, an understanding of the analytic strategy is essential to an understanding of psychological explanation. Equally important, however, is the realization that analysis is an important and generally applicable explanatory strategy, a strategy that is both common and of fundamental importance outside psychology and the life sciences. To appreciate this point is to see that psychological explanation is not anomolous–a special case–but continuous with the rest of science. Forcing psychological explanation into the subsumptivist mold made it continuous with the rest of science only at the price of making it appear trivial or senseless.