Second Printing 1970

Copyright © 1966, University of Notre Dame Press, Notre Dame, Indiana. The introduction and first three chapters were first published in England in 1963 by Sheed and Ward Ltd., in a volume of the same title © Mary B. Hesse.

Library of Congress Catalog Card Number 66-14364

Manufactured in the United States of America by NAPCO Graphic Arts, Inc., Milwaukee, Wisconsin

# Contents

agree, fairly amicably, to differ, but during the course of the argument they have, I think, succeeded in clarifying and settling some of the issues which often befog this topic. The Campbellian has also made some sort of a case for greater attention to be paid in the philosophy of science to logical questions about the nature and validity of analogical argument from models. The subsequent chapters attempt to pursue some of these logical questions, albeit in a preliminary and elementary fashion.

I should like to express my grateful thanks to Professor R. B. Braithwaite and Mr. G. Buchdahl for discussions which have inspired some of the points made here, although probably neither of them will recognize the arguments put into the mouths of my disputants as being positions which they would ever have defended. To avoid a great bulk of footnotes, I have collected in the suggestions for further reading most of the references to published work that I have found valuable in thinking about models and the logic of analogy.

# The Function of Models: A Dialogue

*Campbellian:* I imagine that along with most contemporary philosophers of science, you would wish to say that the use of models or analogues is not essential to scientific theorizing and that theoretical explanation can be described in terms of a purely formal deductive system, some of whose consequences can be interpreted into observables, and hence empirically tested, but that the theory as a whole does not require to be interpreted by means of any model.

*Duhemist:* Yes. I do not deny of course that models may be useful guides in suggesting theories, but I do not think they are essential, even as psychological aids, and they are certainly not *logically* essential for a theory to be accepted as scientific. When we have found an acceptable theory, any model that may have led us to it can be thrown away. Kekulé is said to have arrived at the structure of the benzene ring after dreaming of a snake with its tail in its mouth, but no account of the snake appears in the textbooks of organic chemistry.

*Campbellian:* I, on the other hand, want to argue that models in some sense *are* essential to the logic of scientific theories. But first let us agree on the sense in which we are using the word "model" when we assert or deny that models are essential. I should like to explain my sense of the word by taking Campbell's well-worn example of the dynamical theory of gases. When we take a collection of billiard balls in random motion as a model for a gas, we are not asserting that billiard balls are in all respects like gas particles, for billiard balls are red or white, and hard and shiny, and we are not intending to suggest that gas molecules have these properties. We are in fact saying that gas molecules are *analogous* to billiard balls, and the relation of analogy means that there are some properties of billiard balls which are not found in molecules. Let us call those properties we know belong to billiard balls and not to molecules the *negative analogy* of the model. Motion and impact, on the other hand, are just the properties of billiard balls that we do want to ascribe to molecules in our model, and these we can call the *positive analogy*. Now the important thing about this kind of model-thinking in science is that there will generally be some properties of the model about which we do not yet know whether they are positive or negative analogies; these are the interesting properties, because, as I shall argue, they allow us to make new predictions. Let us call this third set of properties the *neutral analogy*. If gases are really like collections of billiard balls, except in regard to the known negative analogy, then from our knowledge of the mechanics of billiard balls we may be able to make new predictions about the expected behavior of



gases. Of course the predictions may be wrong, but then we shall be led to conclude that we have the wrong model.

*Duhemist:* Your terminology of positive, negative, and neutral analogies is useful but is there not still a possible ambiguity about the sense of "model"? You have mentioned gas molecules and billiard balls. When you speak of the model for gases, do you mean the billiard balls, positive and negative analogy and all, or do you mean what we imagine when we try to picture gas molecules as ghostly little objects having some but not all the properties of billiard balls? I should say that both senses are widely used (among many others), but it is important to distinguish them.

*Campbellian:* I agree they should be distinguished, and I think we can do so conveniently by means of my terminology. Let us agree that when we speak of a model in its primary sense (in this discussion let us call it model$_1$), we are not speaking of another object which can, as it were, be built or imagined alongside the phenomena we are investigating. The model$_1$ is the imperfect copy (the billiard balls) *minus the known negative analogy,* so that we are only considering the known positive analogy, and the (probably open) class of properties about which it is not yet known whether they are positive or negative analogies. When we consider a theory based on a model as an explanation for a set of phenomena, we are considering the positive and neutral analogies, not the negative analogy, which we already know we can discard.

*Duhemist:* Are you not confusing "model" with the theory itself? There is no difference between the theory and the model$_1$ as you now explain it, so why use the word "model" at all?

*Campbellian:* Partly because there is a tendency, particularly among people of your school of thought, to use the word "theory" to cover only what I would call the known positive analogy, neglecting the features of the model which are its growing points, namely its neutral analogy. My whole argument is going to depend on these features, and so I want to make it clear that I am not dealing with static and formalized theories, corresponding only to the known positive analogy, but with theories in the process of growth. Also, since you disagree with me that models are essential to theories, you will necessarily use the word "theory" in a wider sense than my "model"—to cover formal deductive systems which have only a partial interpretation into observables. My models$_1$, on the other hand, are total interpretations of a deductive system depending on the positive and neutral analogies with the "copy." Since I shall also want to talk about the object or copy that includes the negative analogy, let us agree as a shorthand expression to call this "model$_2$." If it is indifferent which sense is meant, I shall simply use "model."

Let us now try to produce a reconstruction of the



use of models and analogies in a familiar example —the wave models for sound and for light. At an elementary level we can set up the following correspondences:

| WATER WAVES | SOUND | LIGHT |
|---|---|---|
| Produced by motion of water particles | Produced by motion of gongs, strings, etc. | Produced by moving flame, etc. |
| Properties of reflection | Echoes, etc. | Reflection in mirrors, etc. |
| Properties of diffraction | Hearing round corners | Diffraction through small slits, etc. |
| Amplitude | Loudness | Brightness |
| Frequency | Pitch | Color |
| Medium: Water | Medium: Air | Medium: "Ether" |

The first three rows indicate some respects in which these three processes appear to be alike to fairly superficial observation. They are, for example, the kind of properties that would go in Bacon's Tables of Presence, or Mill's Agreements. In all three cases there are present motion, something transmitted indirectly from one place to another by hitting an obstacle, and a bending round obstacles. This suggests that the three processes are perhaps alike in more fundamental respects, and in order to investigate this possibility, we look more closely at the one of the three about which we know most, namely, water waves. We postulate, with Huygens, that a disturbance of one particle communicates itself to neighboring particles in such a way that ripples spread from the center of disturbance in concentric circles, and by means of the elementary mathematics of simple harmonic motion we are able to represent the amplitude and frequency of the waves and to derive the laws of reflection and diffraction. We have then a theory of water ripples consisting of equations of the type

$$y = a \sin 2\pi f x$$

where $y$ is the height of the water at the point $x$ measured horizontally, $a$ is the maximum height or amplitude of the ripples, and $f$ is their frequency. From this mathematical theory some laws of the process, such as the equality of the angles of incidence and reflection, can be deduced.

So far we have two sources of information to aid our construction of theories for sound and for light, namely, their observed properties and their observed analogies with water waves, and it is important to notice that both of them appeal only to descriptions of "observable" events. We may define *observation statements* as those descriptive statements whose truth or falsity in the face of given empirical circumstances would be agreed upon by all users of English with or without scientific training. Let us also introduce the term *explicandum* for the set of observation statements connected with the phenomena we are attempting to explain by means of a theory—that is, in this case, the observed properties of



sound or of light. All users of English might not, of course, *notice* the analogies between the three processes until they are pointed out, and up to this point they may have no more significance than the fact that the class of fingers on a hand and petals on a buttercup are similar in that both have five members. But when the analogies have been pointed out, no esoteric insight and no specifically scientific knowledge are required to recognize that they exist. It is not quite the same with the mathematical theory of water waves, for here some knowledge of trigonometry is required, but there is no difficulty in understanding the *terms* "height of water," "frequency of waves," etc. into which the mathematical symbols are interpreted. In this sense the mathematical system is "about" (has its interpretation in terms of) observable events.

Now consider what happens when we make use of the known theory of water waves and the analogies between them and sound in order to construct a theory of sound. The analogies suggest that sound is produced by the motion of air particles propagated in concentric spherical waves from a center of disturbance. Since we know that the greater the disturbance of water the greater the amplitude of the waves, and the greater the disturbance of gongs, strings, hammers, etc., the greater the noise produced, it is easy to identify loudness of sound with amplitude of sound waves, and, similarly, experiences with strings of varying lengths persuade us that pitch or sound is to be identified with frequency of sound waves. In some such way we construct one-to-one correspondences between the observable properties of sound (the explicandum) and those of water waves (the model$_2$), and we are then in a position to test the mathematical wave theory as a theory of sound. Further tests of this kind, of course, may or may not show the theory to be satisfactory. I am not claiming that the use of analogy leads us to an infallible theory, only that it is used in this way to *suggest* a theory. I do not suppose you will want to dispute this so far.

*Duhemist:* No, I have no objection to your reconstruction of the way this particular model might be used. But I am unhappy about the sense in which you say that the initial analogies and the interpretations of the mathematical wave theory in terms of water can be said to be "observable," as contrasted, I suppose, with the air particles which are not observable. I cannot see that there is an important difference here. Surely, to "observe" a similarity between the behavior of ripples at the edge of the swimming bath and the behavior of sound in a mountain valley is a far from superficial observation. It requires a very sophisticated framework of physical ideas in which, for example, the phenomena of echoes are described in terms of a train of physical causes initiated by a shout, rather than in terms of an imitative spirit of the mountains.

*Campbellian:* Yes, I agree with this, and your example indicates that, contrary to what some empiricist philosophers seem to have held, "observation-descriptions" are not written on the face of events to be transferred directly into language but are already "interpretations" of events, and the kind of interpretation depends on the framework of assumptions of a language community. It can plausibly be argued that there is *no* descriptive statement, not even the "blue-here-now" beloved of sense-data theorists, which does not go beyond what is "given" in the act of observing. But I do not wish to pursue this argument here. Would you be prepared to agree that scientific theories bring something *new* into our descriptions of events, and that it is therefore possible to make a distinction between the observation statements of a given language community sharing a framework of assumptions, and the statements going beyond this shared framework which are introduced in scientific theories? It is in contrast with these novelties, which may be called *theoretical statements,* containing *theoretical terms,* that certain at present agreed kinds of description may be called observable. This is to make the distinction a pragmatic one, relative to the assumptions of a given language community, but it does not mean that the traditional empiricist problem of the relation between theory and observation disappears. To realize that every hen was once a chicken is not to absolve oneself from the task of finding out how a hen gives birth to a chicken.

*Duhemist:* Our dispute does not turn on the precise nature of the observation language, and I will accept your pragmatic description of it. But I have another objection to your account of the genesis of a theory of sound. You seem to imply that there are two sorts of theory-construction going on here. First there is the theory of water waves, which is arrived at by making a hypothesis about the propagation of disturbances, expressing this in mathematical language, and deducing from it the observed properties of water waves. There is no mention of any analogies or models here. But in the case of sound it is said that one-to-one correspondences between properties of water and properties of sound are set up first, and then the mathematical wave theory is transferred to sound. This may well be the way in which theories are often arrived at in practice, but you have said nothing to show that reference to the water model is essential or that there is any difference in principle between the relations of theory and observation in the two cases. Both theories consist of a deductive system together with an interpretation of the terms occurring in it into observables, and from both systems can be deduced relations which, when so interpreted, correspond to observed relations, such as the law of reflection. This is all that is required of an explanatory theory. You have implicitly acknowledged it to be sufficient in the case of water waves, and it is also sufficient in the case of sound waves. If



we had never heard of water waves, we should still be able to use the same information about sound to obtain the same result. The information consists of the observed production of sound by certain motions of solid bodies, the relations between the magnitudes of these motions and the loudness of the sound, and between lengths of strings and pitch of note, and the phenomena of echoes and bending.

All of these can be deduced from a mathematical wave theory with appropriate interpretation, without mentioning the water-wave model, and, what is more important, without supposing that there is anything connected with the transmission of sound or light which is analogous to water—that is, without supposing there are some "hidden" motions of particles having the same relation to these observed properties of sound or light that the motions of water particles have to the properties of water waves. In fact, it would be very misleading to suppose any such thing, because some of the further consequences derived from a theory of water waves turn out *not* to be true if transferred, by the one-to-one correspondence, to sound and light transmission.

*Campbellian:* No, but the reason for this in the case of sound at least is not that there is *no* wave model but that ripples are the *wrong* wave model. The oscillation of particles constituting sound waves takes place along the direction of transmission of the sound, like the motion of a piston, and not at right angles to that direction, as with ripples. But what I have just described is itself a model of the motions of air particles derived by analogy with observable events such as the action of buffers on the trucks of a train or (to take Huygens' example) the transmission of pressure along a line of billiard balls when the ball at one end is struck in the direction of the line, and the ball at the other end moves off by itself in the same direction.

*Duhemist:* I did not intend to say that in many cases an alternative model cannot be found when the first breaks down, but only that mention of a model is not part of the logical structure of an explanatory theory, and that it is not even always a useful device for finding such a theory, for it may positively suggest the wrong theory.

It is a question of logic I should like your reactions to. I am impressed, you see, by the situation throughout a large part of modern physics, where it is impossible to find any model like the model of air motions for sound, and where nevertheless the criteria for a deductive theory which I have outlined are still satisfied, and theory construction and testing go on much as before. It may be less satisfactory to the imagination to have no picturable model, and more difficult to construct theories without it, but the continuance of physics in the same logical shape as before shows that the model is not logically necessary to the process.

*Campbellian:* I am not convinced that there is



such an absence of models in modern physics as you suggest, and I may come back to that later. Also, it is a little misleading to speak of "pictures" as if they were synonymous with models, for I would say, for example, that a three-dimensional space curved in a fourth dimension is a perfectly good *model* in relativity theory, but it is certainly not *picturable*. A model, for me, is any system, whether buildable, picturable, imaginable, or none of these, which has the characteristic of making a theory *predictive* in a sense I shall describe later when I try to substantiate my claim that models are logically essential for theories.

But let us stick for the moment to our simple example, because it is easier to bring out the difference between us there. If I understand you, you are saying that in the case of sound waves there is no point even in speaking about motions of air particles, because these are not part of the observed data you list, and we can explain these data equally well by means of a mathematical theory, *some* of whose consequences can be interpreted to give relations between the observables. You will at least admit that here there is a difference between the two theories I described, those for ripples and for sound, in that the motions of water particles *are* observable in the pragmatic sense we have agreed on, and so all the symbols in the equations of the ripple theory are interpretable as observables. In the case of sound, however, we cannot "observe" in this sense the amplitude and frequency of the waves, indeed we cannot observe "waves" at all, we can only infer them from data such as impact of hammer on gong, and vibration of strings. Do you wish to say that a theory of sound need not mention "waves" at all, since these are not observable?

*Duhemist:* I am not suggesting that convenient and universal modes of speech such as this should necessarily be dropped, but let us see what exactly we mean by talking about sound waves. We do not mean just the same as we do when talking about water waves, because, as we have seen, sound waves are longitudinal and not transverse. The word persists because both theories use the same mathematical formalism, which we call the wave equation, differently applied in the two cases. What ripples and sound waves have in common is completely contained in the mathematical formalism, and it is this we point to by continuing to use the word "wave." Of course, I am not denying that it is legitimate to think of the propagation of sound in terms of pulsating spheres of air particles, so long as what we mean by this is controlled by what we know from observation about sound, and not by reference to some *other* process. I suppose this can be expressed in your terminology by saying that if the positive analogy between sound and a model of pulsating spheres is believed to be complete, then this model is identical with our theory of sound, and there is no harm in using the language of the model, as an interpre-



tation of the mathematics of the theory. But I am denying that we can always get this sort of model, and that when we can't we somehow have less of an explanation.

*Campbellian:* I am surprised you are prepared to allow as much as this for common modes of speech, and I am not sure you are consistent in doing so. If you had regarded all talk of "oscillations of air particles" as misleading and dispensable, I should have respected your consistency, but I should then have attacked you on the grounds that you do not give a plausible account of the meaning of theoretical terms. On what I take to be the consistent formalist view, the theory in this case consists *only* of a formal deductive system—marks on paper manipulated according to certain rules—together with the interpretations in terms of observables, so that the only meaning that can be given to, for instance, the parameter *a* in the wave equation is in terms of intensity of sound at the point where that is recorded. There is nothing to say about *a* during the time which elapses between the banging of the gong and the reception of the sound at some distant point. I can say, on the other hand, that *a* has an interpretation at all times during the passage of the sound—namely, it is the amplitude of oscillations of air particles, even though these are "unobservable." Thus I have a solution to the so-called problem of the "meaning of theoretical terms."

*Duhemist:* Well, of course all kinds of definitions of theoretical terms have been suggested to cover cases like this, and in the case of sound waves a conditional definition in terms of observables might be given in the form:

> For all $(x, y, t)$, the amplitude of a sound wave at $(x, y)$ is $a$ if a microphone placed at $(x, y)$ at time $t$ records sound of intensity proportional to $a^2$.

But it is not always possible to give definitions even of this conditional kind, and when it is not, I am content to say that the meaning of "amplitude of sound wave" is given indirectly by the position of $a$ in the deductive system, and the fact that some consequences of the system, when interpreted, have ordinary empirical meaning.

*Campbellian:* So when you spoke of "pulsating spheres of air particles," you were not smuggling in a reference to any $\text{model}_2$, but only intended these words to be a way of speaking about the mathematical symbols? According to you it would be wrong to look up their meaning in a dictionary on this occasion—what is required is to look up the position of the corresponding symbols in the deductive system. This is surely a very strange account of "meaning"? It implies that "indirect meaning" can be given to any word I like to coin by inserting it in a deductive system, for example in the syllogism:

All toves are white
My car is a tove
*therefore* My car is white,



the conclusion of which is observable. "Toves" now has indirect meaning in your sense.

*Duhemist:* This account of indirect meaning must be regarded as necessary but not sufficient. To make it sufficient I should have to add that for a theoretical term to have scientific meaning in this way, it must occur in a deductive system which is seriously considered in science, that is, one which has many observable consequences in different circumstances, all of which are confirmed by observation and none refuted. This is entirely a question for scientific research, an empirical, not a logical question, and so the conditions for a theoretical term to have scientific meaning cannot be logically formalized. But it is clear that your syllogism about toves would not qualify.

*Campbellian:* This still seems to me very strange, the more so because you have agreed to accept an account of observational and theoretical terms in which the distinction between them is not logical but pragmatic. If you accept this you must allow for the frontier between them to shift as science progresses. This is done in my account by saying that we *discover* that sound waves *are* pulsating spheres of air particles *in the ordinary sense of these words,* and if this is accepted by everybody in the language community (as I suppose it is in ours), it does not much matter where the line of "observability" is drawn. Admittedly it would be odd in ordinary speech to talk about "observing" air pulses, but a statement about them might well function as an observation statement in a particular scientific experiment, that is to say, everyone would accept its truth or falsity as the final court of appeal without deducing further "observable" consequences from it. On your account I do not see how "pulsating spheres of air particles" ever gets into ordinary language, because you have specifically denied that these words are used in their ordinary sense.

*Duhemist:* My account is not in the least inconsistent with what we have previously agreed—in fact, I have given an account of *how* the frontier of observability shifts while you have not. The essence of this shift is surely that ordinary language itself changes—when we talk about air pulses we are *not* using the words in exactly the sense they previously had, and what I have done is precisely to explain how ordinary language is extended to take in new senses of these words, depending on the structure of the scientific theory in which they occur. You, on the contrary, have not explained how the ordinary senses of words change. Moreover, I think you have smuggled in a quite different issue here, namely, the question of the "reality" of the air pulses. You seem to imply that I am committed to a nonrealistic view, to saying that they are fictional entities or heuristic devices or what-not, but this is not the case. For me,



to say that "air pulses exist" means just what I have explained—they are entities referred to by (the values of variables in) a deductive system having all characteristics of an accepted scientific theory that I have described. I do hold that *models* are heuristic devices, but I am not committed to holding that theoretical entities understood wholly as interpretations of an accepted mathematical theory are also. If you like, my theoretical entities are related to your models in having the *known positive analogy only.*

*Campbellian:* You have certainly made your position clearer, and I agree that we need not differ on the subject of existence of theoretical entities. We differ on what it is that is asserted to exist. I say that to assert a theory is to assert a model$_1$, positive *and* neutral analogy; you say it is to assert the positive analogy only, and according to you the neutral analogy is merely a heuristic device.

*Duhemist:* Of course, the theory may not be describable in terms of models at all, in cases where I deny that there are models. Then in order to assert the existence of a theoretical entity, we must either coin new words or give old words a new significance by the method of indirect meaning in deductive systems I have described. To go back to your original examples, the word "ether," which you have put in quotes in the third column of your table, was surely a word adopted and given significance in just this way," that is to say, there were some theories seriously considered at one stage in physics in which the ether had a well-defined place in a deductive system and the observable consequences of its properties could be empirically tested.

*Campbellian:* I am not satisfied that this is sufficient. I want to say that the well-defined place it had was due to its being understood in terms of wave models and that its meaning was given by a series of analogies of the form:

$$\frac{\text{water waves}}{\text{water particles}} :: \frac{\text{sound waves}}{\text{air particles}} :: \frac{\text{light waves}}{\text{ether particles}}$$

*Duhemist:* I do not really understand how "meanings" are given by analogies in this way at all. Are you saying simply that *when* there is a model$_2$ for a theory, as in the case of this theory of light, then "air" and "ether" are interpretations of the same set of symbols in the theory, air in the case of sound and ether in the case of light? If so, I agree that we may acquire an intuitive understanding of "ether" in this indirect way, by analogy with the air model$_2$. But since I do not regard models as part of the logic of theories, I cannot regard this sense of "meaning" as interesting for the logician.

*Campbellian:* I do mean by my analogical relations what you suggest, but I also mean something more, which I hope to convince you *is* part of the logic of theories. Let us go back to the example and try to fill out my account of the way the theory of sound is arrived at. I am prepared to concede your objection that, given all the observational information I have allowed myself, I could have gone straight to a mathematical wave theory from which the observations could be deduced, without going through the process of finding one-to-one correspondences with water waves. There will generally be an indefinite number of such mathematical theories, but I agree with you that there is no guarantee that the water-wave model$_2$ will lead to the correct theory, so you rightly ask whether I can have any reasons for using this analogy except the comfortable feeling that I have seen the mathematics before. Well, I think I have a reason, and I can explain it by taking a slightly different situation.



Suppose we are now attempting to construct a theory of light. Your procedure will be to find, no matter how, a mathematical system from which the observed properties of the explicandum—say, reflection and refraction—can be deduced, and for this you will only demand interpretations of the formulae which yield the observable relations you wish to explain. Suppose, by whatever method or lack of method you use, you do choose the mathematical wave theory out of the indefinite number of possibilities. I shall arrive at the same theory by noticing the analogies between light and sound, and setting up a model$_1$ of light transmission in terms of oscillation of particles in a medium. Now, we must distinguish between the various results we can obtain. You will be able to deduce the simple laws of reflection and refraction by using space coordinate and intensity observables, but these will be the only terms in the theory which you will interpret as observables. So far you will have deduced geometrical optics from a mathematical wave theory. If you want to do more than this, you will have to interpret the symbol $f$ in your equation $y = a \sin 2\pi f x$, as well as the symbols $a$ and $x$.

Now you may have observational information that will allow you to do this directly. For example, you may derive from your theory equations relating to the passage of light through a prism, in which you notice that the angle of refraction depends on the value of $f$. If you also have experimental data on the production of a spectrum of colors by the prism, it will be reasonable to set up a one-to-one correspondence between values of $f$ and colors in the spectrum. The theory will then have shown itself capable of explaining the laws of dispersion as well as those of geometrical optics. But suppose you do not know the prism experiment or any other relating to colors. How is $f$ to be interpreted? You may of course make a guess, that since there are lights of different colors and there is an available parameter $f$ in the theory, it would be worth investigating whether the identification of values of $f$ with different colors will yield a correspondence between theory and experiment. Or you may decide that $f$ is uninterpretable; it is part of the machinery of the deductive theory but has no observable correlate. In this case you will not be able to include dispersion in your theory. Have I described your possible procedures correctly?



*Duhemist:* Yes, I will accept that in principle I should have these three possibilities in the case of a hitherto uninterpreted term in the theory. Of course the example you are using hardly brings out the points in a realistic way, because the wave equation was not introduced into optics until after the facts of color dispersion were already known, and so there was little difficulty about this particular identification. But I can see that in other cases there might be no obvious identification of a theoretical term; and then one might, as you suggest, decide to leave it uninterpreted, as in the case of a Schrödinger $\psi$-function in some schools of quantum physics; or one might make what you call a guess, but I should prefer to call a hypothesis, about its interpretation and investigate the experimental consequences of the hypothesis. What I cannot see is that you are any better off when it comes to interpreting a feature of your model. You will of course know that $f$ is what corresponds to the frequency of waves in the model, but in the absence of any observations connecting color with the laws of geometrical optics, which you have already explained by the theory, how does that help you to identify frequency of waves with color? You have the same choice that I have, either to leave $f$ uninterpreted, and hence "frequency of waves" uncorrelated with anything in your theory of light, or to resort to guesswork.

*Campbellian:* I used the word "guess" rather than "hypothesis" to bring out the fact that on your account of the nature of theories you *cannot give any reasons* for choosing to examine one interpretation rather than any other. And I notice that you did not give any actual example of a theoretical term being interpreted without the help of a model. It is no accident that it is difficult to think of an example, because I suggest there always in practice *are* reasons for examining a hypothetical interpretation, and these reasons are drawn from models.

*Duhemist:* Why should I give any reasons before having carried out experimental tests? I cannot give any reasons for choosing one theory rather than another until I have tested it, and the interpretation of a particular theoretical term is only an element in a theory, to be considered as part of the whole. But you have not answered my question about your own procedure. How does your model help you to give reasons for your interpretation?

*Campbellian:* This is where I appeal to the analogy between the model and the phenomena to be explained. Let us first see how I can interpret the parameter $a$ of the theory, which is already correlated in my model with the amplitude of the waves. I suggest that the model$_2$ immediately makes it reasonable to suppose that "magnitude" of the waves corresponds with "magnitude" of the light, and in the case of light, "magnitude" means brightness. Just as a greater wave disturbance means a louder



sound, so does a greater wave disturbance mean a brighter light, although this cannot be investigated directly since we cannot "make a greater wave disturbance" by moving a body as we can with sound. The hypothesis that this is the case comes from an analogy of the following kind:

$$\frac{\text{loudness}}{\text{properties of sound}} :: \frac{\text{brightness}}{\text{properties of light}}$$

I suggest that this analogy is found in the language before any wave theory is thought of. It is independent of the particular theory of light we are considering and so can be used to develop this theory.

*Duhemist:* One might, surely, just as plausibly suggest that brightness is correlated with shrillness, or loudness with purple or scarlet (called, be it noted, "loud" colors).

*Campbellian:* Admittedly there may be some ambiguities of this kind, but if we consider the points of similarity of loudness and brightness—the scale of intensities from absence of sound or light to indefinitely large degrees of it, the analogies between their effects on our sense organs ("deafening" and "blinding"), and so on, the suggested correspondence seems the most plausible.

*Dehumist:* All right, but what about the correspondence between pitch, frequency, and color which you must claim if your method is to work for interpreting the symbol $f$?

*Campbellian:* This is, admittedly, more difficult.

I do not, for example, see how the correspondence of frequency of waves with pitch could have been arrived at without observed correlations involving such things as vibrating strings. In the case of sound used as a model for light there is some plausibility in claiming a pre-scientific analogy:

$$\frac{\text{pitch}}{\text{properties of sound}} :: \frac{\text{color}}{\text{properties of light}}$$

if we think of the various metaphors from sound to light—Locke's blind man's "scarlet sound of trumpets," and the use of such terms as "harmony" and "clash," appealing to analogies of pleasure and pain in their effects on our sense organs.

*Duhemist:* I am not at all convinced that this roundabout way of recognizing analogies can be shown to be other than entirely arbitrary, but even if it can, you seem to me only to have given me one way of making my "guess" at an interpretation of a theoretical term. You have not shown that it constitutes any *reason* for expecting a guess made by this method to be a right or even fruitful one.

*Campbellian:* I hope you will waive for the moment the question of whether any objective analogies of the kind I describe actually exist, because I hope to go into this in more detail later. Meanwhile, I should like to examine the objection you have just made. There are two things I should like to say about it. First, I claim that to assert an analogy between amplitude of waves and loudness of sound or brightness of light, even before any experimental correlation is known, *is* to give a reason for the interpretation of the symbol *a* of a kind which can never be given on your account of the matter.



*Duhemist:* Let me interrupt you before you go any further. Of course, *if* it is possible to find a model$_2$, as it is in this case, an interpretation derived from the model$_2$ can be said to have the reason that it is derived from the model$_2$, and this distinguishes it from any interpretation I might decide to make. But this is pure evasion. I cannot accept a reason *in terms of* a model, for I claim that no model is required. I am asking for a reason for assuming that the model *is* required, or even that it is likely to lead to a better interpretation than one I may make.

*Campbellian:* Of course I cannot expect you to accept a reason appealing to a model, but what I want to point out is that as scientists use the word "reason" in this context, they *will* accept reasons appealing to models. This can be seen in the way they make predictions from models and use them as tests of theories. A prediction will be thought to be reasonable if it follows from an "obvious interpretation" given to a theoretical term by appeal to a model. If the prediction comes off, the theory and its model$_1$ will be regarded as strengthened, whereas if it fails to come off, this may be regarded as sufficiently serious to refute the theory and the model$_1$ together. For example, the corpuscular model of light was regarded as refuted when the obvious interpretation that two corpuscles falling on one spot would produce twice the intensity of light produced by one was shown to be contrary to diffraction experiments. That the model led to the wrong interpretation was in this instance a "reason" for abandoning the whole theory.

*Duhemist:* I am not clear why on your account it should be, for you have already allowed for the possibility that a $\text{model}_2$ may not correspond to the phenomena in *all* respects. Why cannot the feature which fails in this instance be removed to the negative analogy and the rest of the corpuscular $\text{model}_1$ retained?

*Campbellian:* To answer this would certainly require further analysis. Roughly, it would turn on the fact that some properties of $\text{models}_2$ are more "essential" than others, that is to say are causally more closely connected or tend to co-occur more frequently. For example, color is not an essential property of a billiard ball from the point of view of mechanics, but momentum is. If a prediction derived from color fails, this does not essentially affect a mechanical $\text{model}_1$, but if something derived from momentum fails, the $\text{model}_1$ is refuted.

*Duhemist:* But such refutation still depends on the assumption that a theory must have a model, which I am denying. And your example plays into my hands, for we know that the "essential property" you have appealed to in the case of the corpuscular



theory of light is *not* now allowed to refute that theory. The quantum theory of radiation accommodates both diffraction experiments and model-talk about light particles. But the way particles and other models are used in quantum theory is quite consistent with my account. The theory is regarded as satisfactory if it is possible to deduce observed results from the mathematical formalism plus interpretation of some of its terms, and $\text{models}_2$ are used as only mnemonic and heuristic devices when convenient. In this theory $\text{models}_2$ need not even be consistent with one another to be useful.

*Campbellian:* I want to come back to this question of models in quantum theory later, but before that, let us look at this question of prediction more carefully, for this is my second point in answer to your challenge to me to produce reasons for using models.

I have suggested that my model enables me to make predictions because it leads to new and obvious interpretations of some theoretical terms which may then be used to derive new relations between observables. You reply that *any* assignment of a new interpretation, with or without the use of a model, will enable you to make predictions, and that there is no reason to have more confidence in my predictions than in yours. I agree that I have not yet given any reason, but I still want for a moment to pursue my point that the *kind of prediction required* can only be obtained by using models.

I take it that we both agree that a criterion for a theory is that it should be falsifiable by empirical tests. Falsifiability is closely connected with predictive power, although they cannot quite be identified without further analysis. I want to point out that usage of the criterion of falsifiability covers at least three requirements on theories, only the strongest of which is sufficient to establish the superiority of my theory-plus-model over your formal theory. Let us consider three types of falsifiability and three corresponding types of theory, *G, A,* and *B*.

### Type G

In science a single observation statement hardly ever purports to describe only one unique event, but the set of events that would be observed under sufficiently similar circumstances at any time. Hence an observation statement may always be said to be falsifiable in the sense that the circumstances it describes, or sufficiently similar circumstances, may always, in principle, be repeated; hence it is conceivable that a statement which has been confirmed in the past may be falsified in the future. Questions about what would constitute "sufficiently similar circumstances," and what we should be disposed to say about an unexpected falsification of this kind need not detain us, because it is clear that such a sense of "falsifiable" is far too weak to satisfy those who wish to say that a condition for scientific theories is that they are falsifiable. A theory must do more than predict that the same observation statements that



have been confirmed in the past will, in sufficiently similar circumstances, be confirmed in the future.

A scientific theory is required to be falsifiable in the sense that it leads to new observation statements which can be tested, that is, that it leads to new and perhaps unexpected and interesting predictions. But here there is an ambiguity. The weaker sense of such a requirement is that new correlations can be found between the *same* observation predicates; the stronger sense is that new correlations can also be found which involve new observation predicates. It will be convenient to introduce some notation here. I want to argue on the basis of your own account, because I think it does provide some *necessary* conditions that theories must satisfy; what I deny is that they are *sufficient*. Let us consider an observation language containing observation predicates $O_1, \ldots O_j$, $P_1, \ldots P_k$. Suppose there is a set of observation statements each of which is *accepted,* that is to say, each member of the set expresses an empirical correlation between some of the *O*'s and *P*'s which, at a given stage of use of the language, is accepted as true. If the set also exhausts all such accepted statements it will be called the *accepted* set. It represents then a science of these particular observables at the stage of empirical generalizations, before explanatory theories have been introduced. It may not, of course, exhaust *all* the true statements containing *O*'s and *P*'s, because there may be some correlations which remain unnoticed at this stage.

Now consider a set of theoretical predicates (the *T*'s) and a theory containing them which has as consequences all those observation statements of the accepted set which contain *O*'s and only *O*'s. That is to say, the theory is, in your sense, an explanation of the accepted statements containing only *O*'s. This theory may or may not, in addition, contain statements with observation predicates other than the *O*'s, namely the *P*'s. Falsifiability in senses *A* and *B* can now be explained as follows.

### Type A

Suppose the theory does *not* contain any *P*'s. Then it can have no consequences relating to predicates other than the *O*'s. Thus it cannot be used to explain the remaining statements of the accepted set containing any of the *P*'s, nor can it be used to predict correlations between them which are true but not yet accepted. That is to say, it is not falsifiable in the stronger sense. It may, however, be possible to use it to predict correlations between the *O*'s which are true but not yet accepted. Such a theory will be said to be *weakly falsifiable* or *weakly predictive* and will be called a *formal theory*. Many of the so-called "mathematical models" of modern cosmological, economic, and psychological theory are of this kind; they are mathematical hypotheses designed to fit experimental data, in which either there are no theoretical terms or if there are such terms, they are not further interpreted in a model$_2$.



### Type B

Suppose, however, the theory does contain some of the *P*'s. We may dismiss the case in which it contains them only in statements which contain no *T*'s, for then these statements cannot properly be said to be part of the theory, although they may be part of a scheme of empirical generalizations which remain wholly within the observation language. The theory may, however, contain some of the *P*'s in some statements which also contain some *T*'s. Such a theory may then yield as consequences observation statements containing any of these *P*'s, and hence may explain members of the accepted set containing them and may predict new correlations between them. It will then be said to be *strongly falsifiable* or *strongly predictive*.

Now consider how statements containing *T*'s and *P*'s (call them *P*-statements) could come to be introduced into the theory. They are not introduced by considering the observable relations between the *P*'s, because we have supposed that the theory was designed in the first place as an explanation of the *O*'s not the *P*'s. They are not introduced arbitrarily, because if they were there would be no reason why any particular statement should be introduced rather than any other, and such a theory could not be taken seriously as a predictive theory. Also, it would not, as a whole, be falsifiable, because falsification of one arbitrarily introduced *P*-statement could be dealt with by replacing it with another, leaving the rest of the theory unaffected. The only other possibility is that *P*-statements are introduced for reasons *internal to the theory*. These reasons, moreover, cannot be concerned merely with the formal properties of the theory—for example, its formal symmetry or simplicity—because they must be reasons for asserting particular things in the theoretical language about particular observation predicates (the *P*'s), and though the theoretical predicates may be seen from the formalist point of view as uninterpreted symbols, even from this point of view the observation predicates may not. Hence the set of *P*-statements must be interpreted in *terms of the theory*. It is this interpretation which, I maintain, is given by the model, and which requires the whole theory to have a model-interpretation.

*Duhemist:* I am not sure I have followed your symbolism. Surely the *P*'s are already interpreted, since they are observation predicates?

*Campbellian:* Yes, but I am concerned with how they get into the theory. By the conditions of my problem they are not introduced in virtue of their correlations with other observation predicates, hence they must have an interpretation in a model$_2$ which also provides an interpretation of the *theoretical* predicates. Consider my example of sound and light waves, where sound waves are a model$_2$ for the theory of light. Here the *O*'s might be position coordinates and intensities of light, and the *P*'s color predicates.



The theory of reflection and refraction explains the accepted *O*-statements but says nothing about the *P*'s. The question is, how do the *P*'s get into the theory to enable it to make predictions about color? They have to get in in the form of *P*-statements correlating the *P*'s with values of the parameter *f*, and *f* is a *theoretical predicate*. Now the model$_2$ comes in as an interpretation of all the *T*'s into predicates referring to sound waves. *f* is the frequency of sound waves, or pitch. This model$_2$, together with my suggested analogy "pitch corresponds to color," gives the interpretation "*f* in the theory of light corresponds to color," and the theory now yields predictions about color. This can be represented schematically:

| THEORY (containing *a*, *f*, etc. as theoretical predicates) | | INTERPRETATIONS *in sound model$_2$* | | INTERPRETATIONS *in light observables* |
|---|---|---|---|---|
| ↓ | *a* | loudness | $\longleftrightarrow$ | brightness |
| ↓ | *f* | pitch | $\longleftrightarrow$ | color |
| ↓ | | ↓ | | ↓ |
| *O-statements* (geometrical optics, etc.) | | *Observation statements for sound* | | *P-statements* (color dispersion, etc.) |

Here signs of equality indicate interpretations, within a theory, of theoretical predicates into observation predicates; double arrows indicate the direction of deduction; and double arrows indicate observable relations of analogy.

*Duhemist:* I can see that you are asking for a double interpretation of the *P*'s, once into observables and once into the $model_2$, and this is because you want to predict the observable *P*-statements by using the $model_2$.

But let me return to your argument in favor of the step involving the analogies between light and sound. I can see from your diagram that the analogies as well as the interpretations are of two kinds. The first kind are the one-to-one correspondences between theoretical predicates and predicates of the $model_2$, on the one hand, and between theoretical predicates and light observables on the other, giving one-to-one correspondences between predicates referring to light and to sound in virtue of the same formal theory. This, I take it, is the conventional use of "analogy" in mathematical physics, as when Kelvin exhibited analogies between fluid flow, heat flow, electric induction, electric current, and magnetic field, by showing that all are describable by the same equations with appropriate interpretations in each case. But you are asking for something in addition to this, namely, a sense of analogy in terms of which you can make these one-to-one correspondences *before you have got the theory,* by some kind of prescientific



recognition of analogies such as pitch: color. Is this correct?

*Campbellian:* Certainly. My whole point is that it is necessary to have these correspondences before the theory, otherwise the theory is not predictive or falsifiable in the strong sense.

*Duhemist:* I think the weakest part of your argument is where you distinguish your senses *A* and *B* of falsifiability. Even if I admit for a moment that strong falsifiability is required, I cannot accept that *P*-statements can only be introduced into theories by means of your dubious analogies. There are other ways of extending theories which do not deserve your epithet "arbitrary." They give, of course, no guarantee of success, but neither does your model method.

*Campbellian:* If you think there are other methods which will do all that my models do, I think it is up to you to exhibit them. I have already said I do not think merely formal considerations of simplicity and so on are sufficient, because they do not by themselves supply an *interpretation* of the theory as extended, and hence do not supply predictions in a new field of observables. If "simplicity" were extended to apply also to interpretation, then I think you would find you were after all using a model.

*Duhemist:* I think I can do better than to appeal to a vaguely defined sense of "simplicity." We might realize strong falsifiability in the following way. Suppose we are given a number of accepted statements correlating some of the *P*'s with some of the *O*'s. If the consequences of the formal theory are developed, it may be the case that the structure of some of them appear formally similar to that of the accepted statements, in the sense that a one-to-one correspondence between symbols of the theory and terms of the observation statements can be found. It will then be possible to identify some of the *P*'s with symbols of the theory. The theory can then be said to explain the accepted correlations containing these *P*-predicates, and it may also be capable of generating new and as yet unaccepted sentences containing the *P*'s, and therefore of making genuinely new predictions. It was surely in some such way that Maxwell's equations, developed for explanation of electromagnetic phenomena, were seen to explain also the transmission of light, because their solutions were wave equations formally similar to equations of the wave theory of light.

*Campbellian:* I can see some objections to this. First, it is not clear what is meant by "the structure of some of the consequences of the theory being formally similar to that of the accepted statements." In a case such as that of Maxwell's equations it was clear that there was such a similarity or isomorphism, and what the isomorphism consisted in. But it is not easy to say in general how one would recognize a situation of isomorphism, for example, how much formal manipulation of the theory would be admitted before the identifications were found to be



possible? It might even be possible to show that the occurrence of isomorphism is trivial in the sense that *any* sufficiently rich theory could be made isomorphic with any given accepted statements, especially if these were simple and few in number.

You might, of course, be able to evade this objection by tightening up the formal criteria of isomorphism in some way, but even then it is not clear that success in finding an isomorphism would be sufficient in itself to confirm the wider applicability of the theory. Mere formal appearance of the wave equation in two different systems would not suffice to show a correlation in one theory, unless, as with Maxwell's equations, there were some interpretation which made it plausible to assume that one set of phenomena, the optical, was produced by the other, the electromagnetic—the interpretation in this case being that of wave propagation in the material ether. Whittaker gives the example of Mathieu's Equation, which appears in both the theory of elliptic membranes and the theory of equilibrium of an acrobat in a balancing act. It would not be suggested that any unification of theory is accomplished by noticing this fact.

Again, for your program to work significantly, there must already be a fairly well-developed system of relations in the observation language. The less developed this is, the more difficult it will be to ensure that an apparent isomorphism is not accidental or arbitrary. This means that the program will not be universally applicable and not applicable at all to observation predicates not already part of such a system in the observation language. It almost seems as though, for the formalist program to work at all, a previous stage of science making use of theories with models is necessary, in order that a sufficiently complex observation language shall have been built up. That this is the case would be admitted by those who regard classical physics as an observation language for which no further theoretical models are possible, even though classical physics itself consists of theories with models from the point of view of the observation language of common discourse.

The description you now give of the formalist program does not in any case provide necessary criteria for a theory, for, on the formalist view, it can never be more than a lucky accident that a satisfactory isomorphism is found. Whenever it is found, there is a spectacular unification of two or more previously disconnected fields, as in optics and electromagnetism, but such theoretical developments are exceptions which cannot be systematically sought for.

*Duhemist:* But of course we all know that the progress of science is not a mechanically systematic affair, but depends partly on hunches, intuitions, and guesswork, "lucky accidents" if you like, and I do not think my account involves a greater proportion of these than anybody else's. I am in fact prepared to accept that much of the progress of science does depend on these things and to say that the requirement



of falsifiability in sense $B$ is too strong if it is taken to mean that theories of this kind can be sought for systematically. After all, spectacular predictions in observational domains outside the original range of a theory are in fact rare in science and cannot be regarded as necessary logical conditions for a theory. I suggest that whether a theory is required to be falsifiable in this strong sense will depend on the initial complexity of the correlations in the observation language. If this contains only the predicates of ordinary language, and prescientific correlations between them, it is likely that weak falsifiability will not be sufficient for a genuine theory. For if correlations between only a few $O$'s are known, no theory of type $A$ will be able to predict any more, and a theory explaining the correlations between the $O$'s remains imprisoned within the same limited observational situations. If, however, the observation language is already complex—if it is, for example, the language of classical physics—then it is possible that the formal theory may go on for a long time providing interesting correlations between new observational situations which are still described by the same predicates, between, for example, various kinds of particles described by the classical predicates "mass," "charge," and "spin." Parts of the theory of quantum mechanics may well be purely formal, and yet falsifiable in this sense.

*Campbellian:* This is an interesting suggestion, and it would need a far more detailed investigation than we can undertake now. But I should like to introduce some examples from quantum physics to indicate that there may be more elements of model-thinking in it than are recognized by your school of thought. It is usually claimed that, at least on the so-called Copenhagen view, quantum theory is an example of an accepted and useful theory in which models have been abandoned and which, therefore, proves that models are not essential to the progress of theories. And it is certainly true that the Copenhagen view can be regarded as a formalist view of quantum theory in that it refrains from making any interpretations of the formalism of the theory except such as can be made directly in terms of classical physics. It need not trouble us that what stands in place of the observation language here is not ordinary descriptive language but the language of classical physics, which is from another point of view highly theoretical, for we have already agreed that what counts as an observation language is pragmatically relative. But it does not follow that because the adherents of the Copenhagen view refrain from making interpretations when talking *about* quantum theory, they also avoid implicit interpretations when actually using it in the process of research. Many examples could be given from technical papers to show that they do not in fact avoid interpretations. Let me describe a comparatively simple one, which is typical of the kind of argument that cannot be avoided when developments of the theory are suggested.



In terms of classical physics, acting here as the observation language, it is sometimes possible to describe certain phenomena as effects of charged particles, for example, electrons. It is never possible, however, to speak in classical terms of identifying an individual electron on different occasions or, in particular, of distinguishing the state of a system containing two electrons in given positions from that in which the electrons have changed places. According to the Copenhagen view, then, we must not make any interpretation implying anything about the identity of individual electrons. If, however, we do not adhere to this view, there are two possible interpretations of a situation in which an object cannot be re-identified, one exemplified by the model$_2$ of identical billiard balls, and the other by the model$_2$ of pounds, shillings, and pence in a bank balance. In the case of identical billiard balls, if we are not in a position to observe them continuously, we cannot in practice distinguish a situation in which two balls are in two given pockets from a situation at a later time in which they have changed places. But the two situations are in fact different, and if we were concerned with the number of arrangements of two balls in the two pockets, we should have to count them as two different arrangements. With pounds, shillings, and pence in a bank balance, however, it is not merely the case that we cannot in practice re-identify a given pound appearing in the credit column, but that there is no sense in speaking of the self-identity of this pound, and of asking where it reappears in another column or whether it is the pound paid over the counter yesterday. In this case the number of ways of arranging two unit pounds in different places in a column is just one, and there is no sense in speaking of another arrangement in which they have changed places. Units which behave in this way conform to the so-called Fermi-Dirac statistics, and not to the statistics of objects having self-identity.

If the Copenhagen view with regard to electrons were adhered to, we should be unable to say which of these two models of indistinguishability was appropriate, because we should not be in a position to use any models at all. But in fact we find the following argument very frequently used. We are unable to identify individual electrons, hence it is meaningless to speak of the self-identity of electrons, hence electrons are like pounds, shillings, and pence in a balance and not like indistinguishable billiard balls, and hence they conform to Fermi-Dirac statistics. The last step of this argument can be made to yield observable predictions, since there are various ways in which the behavior of entities satisfying Fermi-Dirac statistics is different in classically observable ways from those satisfying the statistics of ordinary objects. But the argument, in spite of its agnosticism about what cannot be observed, does in fact involve an interpretation and a choice between two different models, and without this choice the observable predictions cannot be derived. The crucial step from



formalism to interpretation in the argument occurs when what the observer cannot do—namely, make certain distinction—is taken to be a property of the interpreted system, namely, that there is no such distinction. Such arguments are very commonly used in quantum theory to derive observable results and are sufficient to show that the theory is not as a whole a counter-example to the view that interpretations are essential for predictions.

Another example can be given to indicate the inadequacy of the Copenhagen view, which was developed to deal with the paradoxes of elementary quantum theory and has never been consistently adhered to in the later developments of quantum field theory. In the case of Dirac's prediction of the positron, not only was an interpretative theory successful, but also the same theory treated formally would have been refuted and discarded. The successful prediction arose as follows. The equations of motions of both classical and quantum charged particles admit of solutions representing particles with either positive or negative energy. In classical physics, however, the occurrence of negative energy solutions can be ignored, since in classical physics energy values change continuously, and if a particle is once taken to have positive energy it can never reach a negative-energy state. In quantum physics, however, energy changes take place discontinuously. Thus an electron may jump from one energy state to another, and negative states are as accessible as positive. Now, if the theory of these equations of motion is taken in a formal sense, the nonappearance of negative energy particles in any known experiment would count as a refutation of the theory. Dirac, however, made an interpretation of the theory which depended on the idea that each of the possible negative states is already filled by an electron which is not observable as long as it remains in this state, but which becomes observable if it is knocked out of the state, leaving a "hole" in the negative states which is also observable. By combining the two negatives provided by negative energy and the notion of "hole," the hole can be expected to behave like a particle of *positive* energy, and it will also have positive charge. This predicted particle, the positron, was in fact observed, and hence the interpreted theory both made a successful prediction and explained the previous nonappearance of negative energy particles which threatened to refute the theory regarded formally.

*Duhemist:* It may be true that there are still some preformalist arguments used in quantum theory, but you cannot maintain that in general quantum theory supports your case that models are essential. The fact that here the mathematical formalism may sometimes be usefully interpreted in terms of waves and sometimes in terms of particles, and that these models contradict each other although the formalism is self-consistent, shows that the models cannot be essential to the logic of the theory. The theory



is here the formalism, not the partial interpretations, such as those in your examples, although these may be useful for special and limited problems.

*Campbellian:* I have to agree that the situation in quantum theory is peculiar from my point of view. Perhaps I can put it this way in the terminology I introduced earlier. The particle model (model$_2$) has some positive analogy with atomic phenomena and some negative analogy, and the same applies to the wave model$_2$. Much of the particle model's positive analogy is the wave model's negative analogy, and vice versa, and this is why the two models appear to be contradictory. If that were all there were to say, we could simply extract the two sets of positive analogies and drop all talk about particles and waves, but that is not all there is to say, because in both cases there are still features about which we do not know whether they are positive or negative analogies. And it is in arguing in terms of these features that the particle and wave models are still essential, supplemented by the hunches physicists have acquired about when to argue in terms of one and when the other. And, as you have suggested earlier, developments in quantum theory which appear to be novel (in the sense of falsifiability *B*) may actually be results of novel deductions within parts of the theory already interpreted, and hence be only what I have called extensions of type *A*. These are surely going to yield diminishing returns, and any quantum theorist who adopts my point of view on models will presumably be dissatisfied with the state of the theory until a new model is found incorporating the positive analogies of both particles and waves, but not involving their contradictions. But I don't suppose either of us wishes to rest his arguments on current disputes in quantum theory or on speculations about its future.

*Duhemist:* It sometimes seems that our whole dispute reduces to a difference of opinion about what kind of theory will predominate in the future, and this is rather unprofitable to speculate upon. I think, however, that you have been forced to admit that important extensions of theory *may* take place without the use of models, and so you have effectively admitted that models are not logically essential. You could only continue to maintain that they are by showing that all my examples of formal methods are either unacceptable or not purely formal, and this you have not done. For my part, I can see that it may be possible and useful to analyze in more detail what is involved in using models when they are used and to enquire whether there is any justification for expecting more systematic theory-construction with their aid than without. This would be an extension of inductive logic in application to the hypothetico-deductive structure of theories. I must confess that in view of the inconclusive results of inductive logic in the simpler case of empirical generalizations, I am not very optimistic about the success of such an investigation.



*Campbellian:* I think two sorts of problems have to be distinguished here. There is the general problem of the justification of induction, of which the problem of justifying the inference to hypotheses by means of models would be a special case, and I agree that the history of inductive logic does not make the prospects for this very bright. But there are subsidiary problems to this, namely, to find the conditions for the assertion of an analogy, to elucidate the nature of arguments using models and analogies, and to compare these arguments with those usually called inductive in a more general sense. These problems arise on your view of the nature of theories as well as on mine, because even if models are merely dispensable aids to discovery it is still profitable to ask how they work, and if this is to be called a "psychological" investigation, it may be none the worse for that. Certainly the use of models is not psychological in the sense of being wholly an individual and subjective matter, since communication and argument often go on between scientists in terms of models, and if this shows no more than a uniformity in the scientific temperament, it is still worth investigating.

It does not, of course, follow that such an investigation will provide anything like an infallible method for the construction of theories, any more than it is the intention of accounts of methods of induction to provide infallible induction machines. All that is being attempted is an analysis of what assumptions are made when analogies are used in science, and how it is that certain hypotheses rather than others suggest themselves "by analogy." Whether the hypotheses thus suggested turn out to be *true* is, as always, a matter for empirical investigation. The logic of analogy, like the logic of induction, may be descriptive without being justificatory.

# Material Analogy

Two questions raised in our dialogue now require more detached investigation:

1. What is an analogy?
2. When is an argument from analogy valid?

It is characteristic of modern, as opposed to classical and medieval logic, that the answer to the first question is taken to be either obvious or unanalyzable, while the second is taken to be a question involving induction, and therefore highly problematic. In classical and medieval logic, on the other hand, there is a certain amount of analysis of types of analogy, but practically no attempt at justification of the validity of analogical arguments, although such arguments are frequently used. And since neither the classical types of analogy nor the sketchily defined analogies of modern logic bear much resemblance to analogy as used in reasoning from scientific models,[1] we need to examine the relation of this problem to the traditional discussions. I shall, then, put forward a definition of the analogy *relation* in this chapter, and go on to consider the justification of analogical *argument* in the next.

It is as well to begin by considering very briefly

1. In this chapter, the sense of "model" will always be $\text{model}_2$ of the first chapter unless otherwise stated.