3 To Save the Phenomena[1]

> Physicists call a theory satisfactory if (1) it agrees with the experimental facts, (2) it is logically consistent, and (3) it is simple as compared to other explanations . . . In fact, the author's interest in hidden-variable theories was kindled only when recently he became aware of the possibility of such experimental tests.
> On the other hand, we do not want to ignore the metaphysical implications of the theory.
> F. J. Belinfante, Foreword, *A Survey of Hidden-Variable Theories* (1973)

The realist arguments discussed so far were developed mainly in a critique of logical positivism. Much of that critique was correct and successful: the positivist picture of science no longer seems tenable. Since that was essentially the only picture of science within philosophical ken, it is imperative to develop a new account of the structure of science. This account should especially provide a new answer to the question: what is the *empirical content* of a scientific theory?

§1. Models

Before turning to examples, let us distinguish the syntactic approach to theories from the semantic one which I favour. Modern axiomatics stems from the discussion of alternative geometric theories, which followed the development of non-Euclidean geometry in the nineteenth century. The first meta-mathematics was meta-geometry (a term already used in Bertrand Russell's *Essays on the Foundations of Geometry* in 1897). It will be easiest perhaps to introduce the relevant axiomatic concepts by way of some simple geometric theories. Consider the axioms:[2]

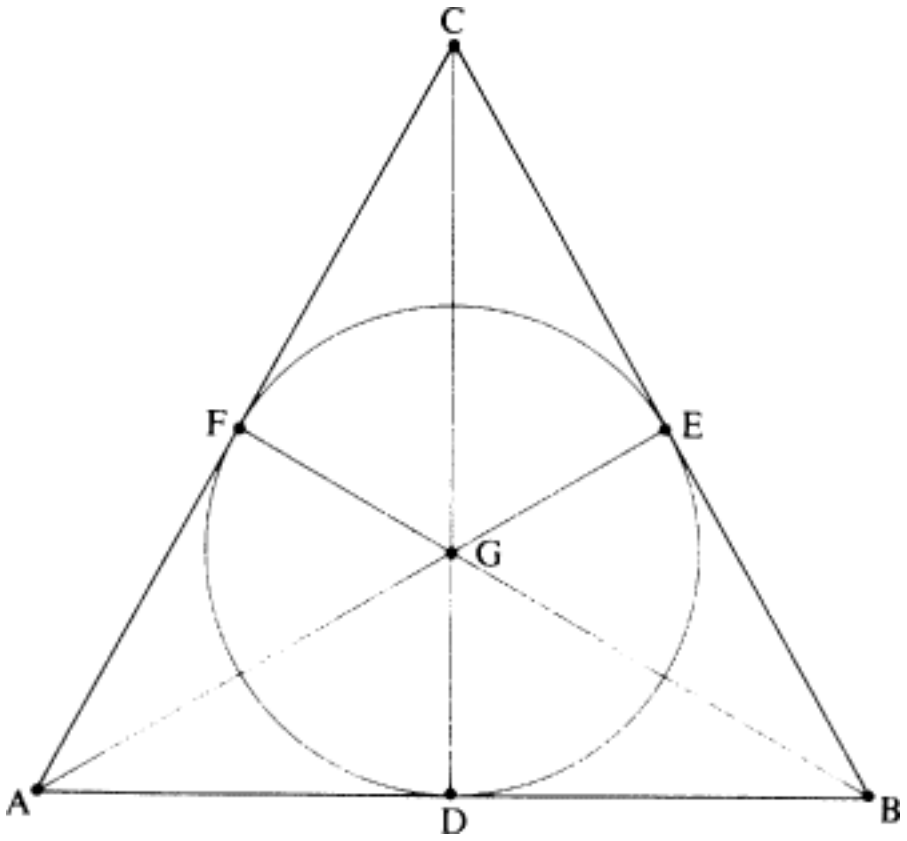
*A*0 There is at least one line.

*A*1 For any two lines, there is at most one point that lies on both.
*A*2 For any two points, there is exactly one line that lies on both.
*A*3 On every line there lie at least two points.
*A*4 There are only finitely many points.
*A*5 On any line there lie infinitely many points.

We have here the makings of three theories: T_0 has axioms *A*1–*A*3; T_1 is T_0 plus *A*4; and T_2 is T_0 plus *A*5.

There are some simple logical properties and relations easily observed here. Each of the three theories is *consistent*: no contradictions can be deduced. Secondly, T_1 and T_2 are *inconsistent with* each other: a contradiction can be deduced if we add *A*5 to T_1. Thirdly, T_1 and T_2 each *imply* T_0: all theorems of T_0 are clearly also theorems of the other two. The first achievement of modern symbolic logic was to give these logical properties and relations precise syntactic definitions, solely in terms of rules for manipulating symbols.

Yet, it will also be noticed that these logical notions have counter-parts in relations expressible in terms of what the theory says, what it is about, and what it could be interpreted as being about. For instance, the consistency of theory T_1 is easiest to show by exhibiting a simple finite geometric structure of which axioms *A*1–*A*4 are true. This is the so-called Seven Point Geometry:



In this structure, only seven things are called 'points', namely, A, B, C, D, E, F, G. And equally, there are only seven 'lines', namely the three sides of the triangle, the three perpendiculars, and the inscribed circle. The first four axioms are easily seen to be true of this structure: the line DEF (i.e. the inscribed circle) has exactly three points on it, namely, D, E, and F; the points F and E have exactly one line lying on both, namely DEF; lines DEF and BEC have exactly one point in common, namely E; and so forth.

Any structure which satisfies the axioms of a theory in this way is called a *model* of that theory. (At the end of this section I shall relate this to other uses of the word 'model'.) Hence, the structure just exhibited is a model of T_1, and also of T_0, but not of T_2. The existence of a model establishes consistency by a very simple straight-forward argument:

> all the axioms of the theory (suitably interpreted) are true of the model; hence all the theorems are similarly true of it; but no contradiction can be true of anything; therefore, no theorem is a contradiction.

Thus logical claims, formulated in purely syntactic terms, can nevertheless often be demonstrated more simply by a detour via a look at models—but the notions of *truth* and *model* belong to semantics.

Nor is semantics merely the handmaiden of logic. For look at the theories T_1 and T_2; logic tells us that these are inconsistent with each other, and there is an end to it. The axioms of T_1 can only be satisfied by finite structures; the axioms of T_2, however, are satisfied only by infinite ones such as the Euclidean plane.

Yet you will have noticed that I drew a Euclidean triangle to convey what the Seven Point Geometry looks like. For that seven-point structure can be *embedded* in a Euclidean structure. We say that one structure can be embedded in another, if the first is isomorphic to a part (substructure) of the second. Isomorphism is of course total identity of structure and is a limiting case of embeddability: if two structures are isomorphic then each can be embedded in the other. The seven-point geometry is isomorphic to a certain Euclidean plane figure, or in other words, it can be embedded in the Euclidean plane. This points to a much more interesting relationship between the theories T_1 and T_2 than inconsistency:

every model of T_1 can be embedded in (identified with a substructure of) a model of T_2.

This sort of relationship, which is peculiarly semantic, is clearly very important for the comparison and evaluation of theories, and is not accessible to the syntactic approach.

The syntactic picture of a theory identifies it with a body of theorems, stated in one particular language chosen for the expression of that theory. This should be contrasted with the alternative of presenting a theory in the first instance by identifying a class of structures as its models. In this second, semantic, approach the language used to express the theory is neither basic nor unique; the same class of structures could well be described in radically different ways, each with its own limitations. The models occupy centre stage.

The use of the word 'model' in this discussion derives from logic and meta-mathematics. Scientists too speak of models, and even of models of a theory, and their usage is somewhat different. 'The Bohr model of the atom', for example, does not refer to a single structure. It refers rather to a type of structure, or class of structures, all sharing certain general characteristics. For in that usage, the Bohr model was intended to fit hydrogen atoms, helium atoms, and so forth. Thus in the scientists' use, 'model' denotes what I would call a model-type. Whenever certain parameters are left unspecified in the description of a structure, it would be more accurate to say (contrary of course to common usage and convenience) that we described a structure-type. Nevertheless, the usages of 'model' in meta-mathematics and in the sciences are not as far apart as has sometimes been said. I will continue to use the word 'model' to refer to specific structures, in which all relevant parameters have specific values.

Rather than pursue this general discussion I turn now to a concrete example of a physical theory, in order to introduce the crucially relevant notions by illustration.

§2. Apparent Motion and Absolute Space

When Newton wrote his *Mathematical Principles of Natural Philosophy* and *System of the World,* he carefully distinguished the phenomena to be saved from the reality to be postulated. He distinguished the 'absolute magnitudes' which appear in his axioms from their 'sensible measures' which are determined experimentally. He discussed carefully the ways in which, and extent to which, 'the true motions of particular bodies may be determined from the apparent', via the assertion that 'the apparent motions . . . are the differences of true motions'.[3]

We can illustrate these distinctions through the discussion of planetary motion preceding Newton. Ptolemy described these motions on the assumption that the earth was stationary. For him, there was no distinction between true and apparent motion: the true motion is exactly what is seen in the heavens. (What that motion is, may of course not be evident at once: it takes thought to realize that a planet's motion really does look like a circular motion around a moving centre.) In Copernicus's theory, the sun is stationary. Hence, what we see is only the planets' motion relative to the earth, which is not itself stationary. The apparent motion of the planets is identified as the difference between the earth's true motion and the planets' true motion—true motion being, in this case, motion relative to the sun. Finally, Newton, in his general mechanics, did not assume that either the earth or the sun is stationary. And he generalized the idea of apparent motion—which is motion relative to the earth—to that of motion of one body relative to another. We can speak of the planets' motion relative to the sun, or to the earth, or to the moon, or what have you. What is observed is always some relative motion: an apparent motion is a motion relative to the observer. And Newton held that relative motions are always identifiable as a difference between true motions, whatever those may be (an assertion which can be given precise content using vector representation of motion).

The 'apparent motions' form relational structures defined by measuring relative distances, time intervals, and angles of separation. For brevity, let us call these relational structures *appearances*. In the mathematical model provided by Newton's theory, bodies are located in Absolute Space, in which they have real or absolute motions. But within these models we can define structures that are meant to be exact reflections of those appearances, and are, as Newton says, identifiable as differences between true motions. These structures, defined in terms of the relevant relations between absolute locations and absolute times, which are the appropriate parts of Newton's models, I shall call *motions,* borrowing Simon's term.[4] (Later I shall use the more general term *empirical substructures*.)

When Newton claims empirical adequacy for his theory, he is claiming that his theory has some model such that *all actual appearances are identifiable with (isomorphic to) motions* in that model. (This refers of course to all actual appearances throughout the history of the universe, and whether in fact observed or not.)

Newton's theory does a great deal more than this. It is part of his theory that there is such a thing as Absolute Space, that absolute motion is motion relative to Absolute Space, that absolute acceleration causes certain stresses and strains and thereby deformations in the appearances, and so on. He offered in addition the *hypothesis* (his term) that the centre of gravity of the solar system is at rest in Absolute Space.[5] But as he himself noted, the appearances would be no different if that centre were in any other state of constant absolute motion. This is the case for two reasons: differences between true motions are not changed if we add a constant factor to all velocities; and force is related to changes in motion (accelerations) and not to motion directly.

Let us call Newton's theory (mechanics and gravitation) *TN,* and *TN(v)* the theory *TN* plus the postulate that the centre of gravity of the solar system has constant absolute velocity *v.* By Newton's own account, he claims empirical adequacy for *TN(0)*; and also that, if *TN(0)* is empirically adequate, then so are all the theories *TN(v).*

Recalling what it was to claim empirical adequacy, we see that all the theories *TN(v)* are empirically equivalent exactly *if all the motions in a model of TN(v) are isomorphic to motions in a model of TN*(*v*+*w*), for all constant velocities *v* and *w.* For now, let us agree that these theories are empirically equivalent, referring objections to a later section.

§3. Empirical Content of Newton's Theory

What exactly is the 'empirical import' of *TN(0)*? Let us focus on a fictitious and anachronistic philosopher, Leibniz*, whose only quarrel with Newton's theory is that he does not believe in the existence of Absolute Space. As a corollary, of course, he can attach no 'physical significance' to statements about absolute motion. Leibniz* believes, like Newton, that *TN(0)* is empirically adequate; but not that it is true. For the sake of brevity, let us say that Leibniz* *accepts* the theory but that he does not *believe* it; when confusion threatens we may expand that idiom to say that he *accepts the theory as empirically adequate,* but does not *believe it to be true.* What does Leibniz* believe, then?

Leibniz* believes that *TN(0)* is empirically adequate, and hence equivalently, that all the theories *TN(v)* are empirically adequate. Yet we cannot identify the theory which Leibniz* holds about the

world—call it *TNE*—with the common part of all the theories *TN(v)*. For each of the theories *TN(v)* has such consequences as that the earth has *some* absolute velocity, and that Absolute Space exists. In each model of each theory *TN(v)* there is to be found something other than motions, and there is the rub.

To believe a theory is to believe that one of its models correctly represents the world. You can think of the models as representing the possible worlds allowed by the theory; one of these possible worlds is meant to be the real one. To believe the theory is to believe that exactly one of its models correctly represents the world (not just to some extent, but in all respects). Therefore, if we believe of a family of theories that all are empirically adequate, but each goes beyond the phenomena, then we are still free to believe that each is false, and hence their common part is false. For that common part is phraseable as: one of the models of one of those theories correctly represents the world.

The theory which Leibniz* holds of the world, *TNE,* can nevertheless be stated, and I have already done so. Its single axiom can be the assertion that *TN(0)* is empirically adequate: that *TN(0)* has a model containing motions isomorphic to all the appearances. Since *TN(0)* can be stated in English, this completes the job.

It may be objected that so stated, *TNE* does not look like a physical theory. Indeed it looks metalinguistic. This is a poor objection. The theory is clearly stated in English, and that suffices. Whether or not it is axiomatizable in some more restricted vocabulary may be a question of logical interest, but philosophically it is irrelevant. Secondly, if the set of models of *TN(0)* can be described without metalinguistic resources, then the above statement of *TNE* is easily turned into a non-metalinguistic statement too. Not that this matters. The only important point here is that the empirical import of a family of empirically equivalent theories is not usually their common part, but can be characterized directly in the same terms in which empirical adequacy is claimed.

§4. Theories and Their Extensions

The objection may be raised that theories can seem empirically equivalent only as long as we do not consider their possible extensions. When we consider their application beyond the originally intended domain of application, or their combination with other

theories or hypotheses, we find that distinct theories do have different empirical import after all.[6] An imperfect example is furnished by Brownian motion, which established the superiority of the kinetic theory over phenomenological thermodynamics. This example is imperfect, for it was known that the two theories disagreed even on macroscopic phenomena over sufficiently long periods of time. Until the discovery of Brownian motion, it was thought that experiments would not yield 'fine' enough data to shorten sufficiently the period of time required to show the divergence of the theories.

A perfect example can be constructed as a piece of quite realistic science fiction: let us imagine that such experiments as Michelson and Morley's, which led to the rise of the theory of relativity, did not have their spectacular, actual, null-outcome, and that Maxwell's theory of electromagnetism was successfully combined with classical mechanics. In retrospect we realize that such a development would have upset even Newton's deepest convictions about the relativity of motion; but we can imagine it.

Electrified and magnetic bodies appear to set each other in motion although they are some distance apart. Early in the nineteenth century mathematical theories were developed treating these phenomena in analogy with gravitation, as cases of action at a distance, by means of forces which such bodies exert on each other. But the analogy could not be perfect: it was found necessary to postulate that the force between two charged particles depends on their velocity as well as on the distance.

Adapting the idea of a universal medium of the propagation of light and heat (the luminiferous medium, or ether) found elsewhere in physics, Maxwell developed his theory of the electromagnetic field, which pervades the whole of space:

> It appears therefore that certain phenomena in electricity and magnetism lead to the same conclusions as those of optics, namely, that there is an ethereal medium prevading all bodies, and modified only in degree by their presence . . .[7]

The force on an electrified body is a force 'exerted by' this medium, and depends on the body's position and on its velocity. Maxwell's Equations describe how this field develops in time.

The difficulties with Maxwell's theory concerned the mechanics of this medium; and his ideas about what this medium was like, were not successful. But this did not blind the nineteenth century to the power and adequacy of his equations describing the electromagnetic

field. The consensus is perhaps expressed by Hertz's famous statement 'Maxwell's Theory is Maxwell's Equations'. It would therefore not be appropriate to call Maxwell's theory a mechanical theory, but it did have mechanical models. The existence of such models follows from a mathematical result due to Koenig, as Poincaré detailed in the preface of his *Electricité et Optique* and Chapter XII of his *Science and Hypothesis*. There was this strange new feature, however; the forces depend on the velocities, and not merely on the accelerations. There was accordingly a spate of thought-experiments designed to measure absolute velocity. The very simplest was described by Poincaré:

> Consider two electrified bodies; though they seem to us at rest, they are both carried along by the motion of the earth; an electric charge in motion, Rowland has taught us, is equivalent to a current; these two charged bodies are, therefore, equivalent to two parallel currents of the same sense and these two currents should attract each other. In measuring this attraction, we shall measure the velocity of the earth; not its velocity in relation to the sun or the fixed stars, but its absolute velocity.[8]

The frustratingly uniform null-outcome of all such experiments led to the demise of classical physics and the advent of relativity theory. But let us imagine that the classical expectations were not disappointed. Imagine that values are found for absolute velocities; specifically for the centre of gravity of the solar system. In that case, it would seem, one of the theories *TN(v)* may be confirmed and all the others falsified. Hence those theories were not empirically equivalent after all.

But the reasoning is spurious. The definition of empirical equivalence did not rely on the *assumption* that only absolute acceleration can have discernible effects. Newton made the distinction between sensible measures and apparent motions on the one hand, and true motions on the other, without presupposing more than that basic mechanics within which there are models for Maxwell's equations. The assertion was that each motion in a model of *TN(v)* is isomorphic to a motion in a model of *TN*(v+w), for all constant velocities v and w. This assertion was the reason for the claim of empirical equivalence. The question before us is whether that assertion was controverted by those nineteenth-century reflections.

The answer is definitely *no*. The thought-experiment, we may imagine, confirmed the theory that adds to *TN* the hypotheses:

H0 The centre of gravity of the solar system is at absolute rest.

E0 Two electrified bodies moving in parallel, with absolute velocity v, attract each other with force $F(v)$.

This theory has a consequence strictly about appearances:

CON Two electrified bodies moving with velocity v relative to the centre of gravity of the solar system, attract each other with force $F(v)$.

However, that same consequence can be had by adding to *TN* the two alternative hypotheses:

Hw The centre of gravity of the solar system has absolute velocity w.
Ew Two electrified bodies moving with absolute velocity $v + w$ attract each other with force $F(v)$.

More generally, for each theory $TN(v)$ there is an electromagnetic theory $E(v)$ such that $E(0)$ is Maxwell's and all the combined theories $TN(v)$ plus $E(v)$ are empirically equivalent with each other.

There is no originality in this observation, of which Poincaré discusses the equivalent immediately after the passage I cited above. Only familiar examples, but rightly stated, are needed it seems, to show the feasibility of concepts of empirical adequacy and equivalence. In the remainder of this chapter I shall try to generalize these considerations, while showing that the attempts to explicate those concepts *syntactically* had to reduce them to absurdity.

§5. Extensions: Victory and Qualified Defeat

The idea that theories may have hidden virtues by allowing successful extensions to new kinds of phenomena, is too pretty to be left. Developed independently of the example in the last section it might yet trivialize empirical equivalence. Nor is it a very new idea. In the first lecture of his *Course de philosophie positive,* Comte referred to Fourier's theory of heat as showing the emptiness of the debate between partisans of calorific matter and kinetic theory. The illustrations of empirical equivalence have that regrettable tendency to date; calorifics lost. Federico Enriques seemed to place his finger on the exact reason when he wrote: 'The hypotheses which are indifferent in the limited sphere of the actual theories acquire significance from the point of view of their possible *extension*."[9] And this suggests that after all, distinct theories can never be *really*

empirically equivalent, because they may differ significantly in their extensions.

To evaluate this suggestion we must ask what exactly is an extension of a theory. Let us suppose as in the last section that experiments did indicate the combined theory *TN(0)* plus *E(0)*. In that case we would surely say that mechanics had been successfully extended to electromagnetism. What, then, is a successful extension?

There were mechanical models of electromagnetic phenomena; and also of the phenomena more traditionally subject to mechanics. What we have supposed is that all these appearances could *jointly* find a home among the motions in a single model of *TN(0)*. Certainly, we have here an extension of *TN(0),* but first and foremost we have a *victory*. We have an extension, for the class of models that may represent the phenomena has been narrowed to those which satisfy the equations of electromagnetism. But it is a victory for *TN(0)* because it simply bears out the claim that *TN(0)* is empirically adequate: all appearances can be identified with motions in one of its models.

Such victorious extensions can never distinguish between empirically equivalent theories in the sense in which that relation was described above, for such theories have exactly the same resources for modelling appearances. It follows logically from the italicized description in Section 2 that if one theory enjoys such a victory, then so will all those empirically equivalent to it.

So if Enriques's idea is to be correct at all, there must be other sorts of extensions, which are not victories. Let us suppose that a theory is confronted with new phenomena, and these are not even piece-wise identifiable with motions in the models of that theory. Must that old theory then suffer an unqualified defeat, and hope for nothing more than a survival as 'correct for a limited range', as approximating some fragment of some victorious new theory? There seems to be one possibility intermediate between victory and total defeat. The class of substructures called *motions* might, for example, be widened to a larger class; let us say, *pseudo-motions*. And the theory might be weakened, so that it would claim only that every appearance can be identified with a pseudo-motion.

This would be a defeat, for the claim that the old theory is empirically adequate has been rescinded. But still it may be called an extension rather than a replacement, for the class of models (the over-all structures within which motions and pseudo-motions are

defined) has no new members added. It is therefore an extension, which is not a victory but anyway a qualified defeat.

It is not so easy to find an example of this kind of extension within the sphere of mechanics, but the following may be one. Brian Ellis constructed a theory in which no forces are postulated, but the available motions are the same as in Newton's mechanics plus the postulate of universal gravitation.[10] The effect of gravitational attraction is cunningly woven into the basic equations of motion of Ellis's theory. But Ellis has pointed out that Newton's theory has a certain kind of superiority in that, if the effect of gravitation is just slightly different, then Newton's theory is much more easily amended than his. In other words, if Newton's theory turned out wrong in its astronomical predictions, there is an obvious way to try and repair it, without touching his basic laws of motion.

It is possible to construe this as follows: the two theories are empirically equivalent, but Newton's allows of certain obvious extensions of the second sort. To see it this way, one has to take the law *G* of universal gravitation as defining the *motions* (described in terms of relative distances) in Newton's models: a motion is a set of trajectories for which masses and forces can be found such that Newton's laws of motion and *G* are satisfied. Then if evidence accrued in favour of an alternative postulate *G'* about gravity, the extension could proceed on the idea that the gravitational force is itself a function of some other factor, and by defining *pseudo-motions* as trajectories satisfying the suitably generalized law.

It will, however, be clear that the second sort of extension is a defeat. There is a certain kind of superiority perhaps in the ability to sustain qualified rather than total defeat. But it is a pragmatic superiority. It cannot serve to upset the conclusion that two theories are empirically equivalent, for it does not show that they differ in any way (not even conditionally, not even counterfactually) in their empirical import.

Let me close this section with an example of another sort of pragmatic superiority, which strikes me as quite similar.

Suppose that two proposed theories have different axioms, but turn out to have the same theorems (and the same models, and the same specification of empirical substructures). I do not suppose that anyone would think that these two theories say different things. Even so, there may be a recognizable superiority, which appears when we attempt to generalize them. An interesting example of this is given

by Belinfante in his discussion of von Neumann's 'proof' that there can be no hidden variables in quantum-mechanical phenomena.[11] The observable quantities are represented by operators $A, B, \ldots$ each of which has associated with it an infinite matrix $(A)_{ij}$ and also a function $\langle A \rangle$ *which gives its expectation value* $\langle A \rangle_\phi$ in any state ϕ.

When he wrote his own theory, von Neumann could have chosen either of the following principles concerning combination of observable quantities to serve as an axiom:

1. $\langle aA + bB \rangle_\phi = a \langle A \rangle_\phi + b \langle B \rangle_\phi$
2. $(aA + bB)_{ij} = a(A)_{ij} + b(B)_{ij}$

With a suitable choice of other axioms and definitions, the one not chosen as an axiom would have been derivable as a theorem. In fact, von Neumann chose 1. When he then came to the question of hidden variables, he showed that their existence would contradict the generalization of his basic axioms to states supplemented with hidden variables. However, it can easily be shown that any reasonable hidden variable theory must reject the generalization of 1, although it can accept 2. Had von Neumann chosen his axioms differently, he might well have reached the conclusion that 1 can be demonstrated for all quantum-mechanical states, but does not hold for the postulable underlying microstates—and hence, that there could be hidden variables after all.

Such pragmatic superiorities of one theory over another are of course very important for the progress of science. But since they can appear even between different formulations of the same theory, and also may only show up in actual defeat, they are no reflection on what the theory itself says about what is observable.

§6. Failure of the Syntactic Approach

Specific examples of empirical adequacy and equivalence should suffice to establish the correctness and non-triviality of these concepts; but we need an account of them in general. It is here that the syntactic approach has most conspicuously been tried, and has most conspicuously failed.

The syntactic explication of these concepts is familiar for it is the backbone of the account of science developed by the logical positivists. A theory is to be conceived as what logicians call a deductive theory, hence, a set of sentences (the theorems), in a specified

language. The vocabulary is divided into two classes, the observational terms and the theoretical terms. Let us call the observational sub-vocabulary E. The empirical import of a theory T is identified as its set of testable, or observational, consequences; the set of sentences T/E, which are the theorems of T expressed in sub-vocabulary E. Theories T and T' are declared empirically equivalent exactly if T/E is the same as T'/E. An extension of a theory is just an axiomatic extension.

Obvious questions were raised and settled. A theory would seem not to be usable by scientists if it is not axiomatizable. Is T/E axiomatizable if T is? William Craig showed that, if the sub-vocabulary E is suitably specified, and T is recursively axiomatizable in its total vocabulary, then so is T/E in the vocabulary E.[12] Note that the question is interesting to logicians only if it is construed as being about axiomatizability in a *restricted* vocabulary. Of course, if T is axiomatizable and E suitably specifiable in English, then T/E is too. But logicians attached importance to questions about restricted vocabularies, and that was seemingly enough for philosophers to think them important too.

A more philosophical problem was apparently posed by the very distinction between observational and theoretical terms. Certainly in some way every scientific term is more or less directly linked with observation. When the distinction began to seem untenable, those who wished still to work with the syntactic scheme began to divide the vocabulary into 'old' and 'new' (or 'newly introduced') terms.[13]

But all this is mistaken. The empirical import of a theory cannot be isolated in this syntactical fashion, by drawing a distinction among theorems in terms of vocabulary. If that could be done, T/E would say exactly what T says about what is observable and what it is like, and nothing more. But any unobservable entity will differ from the observable ones in the way it systematically lacks observable characteristics. As long as we do not abjure negation, therefore, we shall be able to state in the observational vocabulary (however conceived) that there are unobservable entities, and, to some extent, what they are like. The quantum theory, Copenhagen version, implies that there are things which sometimes have a position in space, and sometimes have not. This consequence I have just stated without using a single theoretical term. Newton's theory implies that there is something (to wit, Absolute Space) which neither has a position nor occupies a volume. Such consequences

are by no stretch of the imagination about what there is in the observable world, nor about what any observable thing is like. The reduced theory T/E is not a description of part of the world described by T; rather, T/E is, in a hobbled and hamstrung fashion, the description by T of everything.

Thus on the syntactic approach, the distinction between truth and empirical adequacy reduces to triviality or absurdity, it is hard to say which. Similarly for empirical equivalence. Recalling Section 2, we see that *TN(0)* and *TNE* must be empirically equivalent, for the latter stated that *TN(0)* is empirically adequate. But the former states that there is something (to wit, Absolute Space) which is different from every appearance by lacking even those minimal characteristics which all appearances share. Hence, $TN(0)/E$ is not the same as TNE/E; and so, on the syntactic approach, these theories are not empirically equivalent after all.

Philosophers seem to have been bothered more by ways in which the syntactic definition of empirical equivalence might be too broad. It was noted that many theories T are such that T/E is tautological, or wellnigh so. Such theories presumably derive their empirical import from the consequences they have when conjoined with other theories or empirical hypotheses. But in that case, T/E and T'/E might be the same even if T and T' are about totally different subjects.

To eliminate this embarrassment, extensions of theories were considered.[14] With a bow to Enriques, one may newly stipulate T and T' are empirically equivalent if and only if all their axiomatic extensions are, that is, if for every theory T'', (T plus T''/E is the same as (T'plusT'')$/E$.

While this manœuvre removes the second embarrassment, it runs afoul of the first. *TN(0)* and *TNE* are again declared non-equivalent. Worse yet. *TN(0)* is no longer empirically equivalent to the other theories $TN(v)$. This is shown by the examples of spurious reasoning in Section 4 above: *TN(0)* plus *E(0)* is not equivalent to $TN(v)$ plus *E(0)* for non-zero values of v. But all the theories $TN(v)$ *are* empirically equivalent. Nor is it easy to see how we could restrict the class of axiomatic extensions to be considered so as to repair this deficiency.

These criticisms should suffice to show that the flaws in the linguistic explication of the empirical import of a theory are not minor or superficial. They do not, of course, constitute an *a priori*

proof of the impossibility of a pure observation language. But such a project loses all interest when it appears so clearly that, even if such a language could exist, it would not help us to separate out the information which a theory gives us about what is observable. It seems in addition highly unlikely that such a language could exist. For at the very least, if it existed it might not be translatable into natural language. An observation language would be theoretically neutral at all levels. So if *A* and *B* are two of its simplest sentences, they would be logically independent. This shows at once that they could not have the English translations 'there is red-here-now' and 'there is green-here-now', which are mutually incompatible. Pursuing such questions further does not seem likely to shed any light on the nature or structure of science.

The syntactically defined relationships are simply the wrong ones. Perhaps the worst consequence of the syntactic approach was the way it focused attention on philosophically irrelevant technical questions. It is hard not to conclude that those discussions of axiomatizability in restricted vocabularies, 'theoretical terms', Craig's theorem, 'reduction sentences', 'empirical languages', Ramsey and Carnap sentences, were one and all off the mark—solutions to purely self-generated problems, and philosophically irrelevant. The main lesson of twentieth-century philosophy of science may well be this: no concept which is essentially language-dependent has any philosophical importance at all.

§7. The Hermeneutic Circle

We have seen that we cannot interpret science, and isolate its empirical content, by saying that our language is divided into two parts. Nor should that conclusion surprise us. The phenomena are saved when they are exhibited as fragments of a larger unity. For that very reason it would be strange if scientific theories described the phenomena, the observable part, in different terms from the rest of the world they describe. And so an attempt to draw the conceptual line between phenomena and the trans-phenomenal by means of a distinction of vocabulary, must always have looked too simple to be good.

Not all philosophers who have discussed the observable/unobservable distinction, by any means, have done so in terms of vocabulary. But there has been a further assumption common also to critics of that distinction: that the distinction is a philosophical one. To draw

it, they seem to assume, is in principle anyway the task of the philosophy of perception. To draw it, in principle anyway, philosophy must mobilize theories of sensing and perceiving, sense data and experiences, *Erlebnisse* and *Protokolsaetze*. If the distinction is a philosophical one, then it is to be drawn, if at all, by philosophical analysis, and to be attacked, if at all, by philosophical arguments.

This attitude needs a Grand Reversal. If there are limits to observation, these are a subject for empirical science, and not for philosophical analysis. Nor can the limits be described once and for all, just as measurement cannot be described once and for all. What goes on in a measurement process is differently described by classical physics and by quantum theory. To find the limits of what is observable in the world described by theory *T* we must inquire into *T* itself, and the theories used as auxiliaries in the testing and application of *T*.

We have now come to the 'hermeneutic circle' in the interpretation of science. I want to spell this out in detail, because one might too easily get a feeling of vicious circularity. And I want to give specific details on how science exhibits clear limits on observability.

Recall the main difference between the realist and anti-realist pictures of scientific activity. When a scientist advances a new theory, the realist sees him as asserting the (truth of the) postulates. But the anti-realist sees him as displaying this theory, holding it up to view, as it were, and claiming certain virtues for it.

This theory draws a picture of the world. But science itself designates certain areas in this picture as observable. The scientist, in accepting the theory, is asserting the picture to be accurate in those areas. This is, according to the anti-realist, the only virtue claimed which concerns the relation of theory to world alone. Any other virtues to be claimed will either concern the internal structure of the theory (such as logical consistency) or be pragmatic, that is, relate specifically to human concerns.

To accept the theory involves no more belief, therefore, than that what it says about observable phenomena is correct. To delineate what is observable, however, we must look to science—and possibly to that same theory—for that is also an empirical question. This might produce a vicious circle if what is observable were itself not simply a fact disclosed by theory, but rather theory-relative or theory-dependent. It will already be quite clear that I deny this; I regard what is observable as a theory-independent question. It is a

function of facts about us *qua* organisms in the world, and these facts may include facts about the psychological states that involve contemplation of theories—but there is not the sort of theory-dependence or relativity that could cause a logical catastrophe here.

Let us consider two concrete examples which have been found puzzling. The first, already mentioned by Grover Maxwell, concerns molecules. Certain crystals, modern science tells us, are single molecules; these crystals are large enough to be seen—so, some molecules are observable. The second was mentioned to me by David Lewis: astronauts reported seeing flashes, and NASA scientists came to the conclusion that what they saw were high-energy electrons.

Is there anything puzzling about these examples? Only to those who think there is an intimate link between theoretical terms and unobservable entities or events. Compare the examples with Eddington's famous table: that table is an aggregate of interacting electrons, protons, and neutrons, he said; but that table is easily seen. If a crystal or table is classified by a theory as a theoretically described entity, does the presence of this observable object become evidence for the reality of other, different but similarly classified entities? Everything in the world has a proper classification within the conceptual framework of modern science. And it is this conceptual framework which we bring to bear when we describe any event, including an observation. This does not obliterate the distinction between what is observable and what is not—for that is an empirical distinction—and it does *not* mean that a theory could not be right about the observable without being right about everything.

We should also note here the intertranslatability of statements about objects, events, and quantities. There is a molecule in this place; the event of there-being-a-molecule occurs in this place (this is, roughly, Reichenbach's event language); a certain quantity, which takes value *one* if there is a molecule here and value *zero* if there is not, has value *one*. There is little difference between saying that a human being is a good detector of molecules and saying that he is a good detector of the presence of molecules. Any such classification of what happens may be correct, relative to a given, accepted theory. If we follow the principles of the general theory of measurement used in discussions of the foundations of quantum mechanics, we call system Y a measurement apparatus for quantity A exactly if Y has a certain possible state (the ground-state) such that if Y is in that state and coupled with another system X in *any* of its possible

states, the evolution of the combined system (*X* plus *Y*) is subject to a law of interaction which has the effect of correlating the values of *A* in *X* with distinct values of a certain quantity *B* (often called the 'pointer reading observable') in system *Y*. Since observation is a special subspecies of measurement, this is a good picture to keep in mind as a partial guide.

Science presents a picture of the world which is much richer in content than what the unaided eye discerns. But science itself teaches us also that it is richer than the unaided eye *can* discern. For science itself delineates, at least to some extent, the observable parts of the world it describes. Measurement interactions are a special subclass of physical interactions in general. The structures definable from measurement data are a subclass of the physical structures described. It is in this way that science itself distinguishes the observable which it postulates from the whole it postulates. The distinction, being in part a function of the limits science discloses on human observation, is an anthropocentric one. But since science places human observers among the physical systems it means to describe, it also gives itself the task of describing anthropocentric distinctions. It is in this way that even the scientific realist must observe a distinction between the phenomena and the trans-phenomenal in the scientific world-picture.

§8. Limits to Empirical Description

Are there limits to observation? While the arguments of Grover Maxwell aim to establish that in principle there are not (so as to undercut the very possibility of the statement of an empiricist philosophy of science), other arguments aim to establish the inadequacy of empiricism because of these limits. Since physical theory cannot be translated, without remainder, into a body of statements that describe only what the observable phenomena are like, such arguments run, empiricism cannot do justice to science. I grant the premiss, of course, and wish here to reinforce it by giving a more precise statement of the limits of empirical description, and some examples.

Before attempting precision, let us examine the standard example of 'underdetermination' to be drawn from foundational studies in classical mechanics. In the context of that theory, and arguably in all of classical physics, all measurements are reducible to series of measurements of time and position. Hence let us designate as basic

observable all quantities which are functions of time and position alone. These include velocity and acceleration, relative distances and angles of separation—all the quantities used, for example, in reporting the data astronomy provides for celestial mechanics. They do not include mass, force, momentum, kinetic energy.

To some extent, and in many cases, these other quantities can be calculated from the basic observables. Hence the many proposed 'definitions' of force and mass in the nineteenth century, and the availability of axiomatic theories of mechanics today in which mass is not a primitive quantity.[15] But, as Patrick Suppes has emphasized, if we postulate with Newton that every body has a mass, then mass is not definable in terms of the basic observables (not even if we add force).[16] For, consider, as simplest example, a (model of mechanics in which a) given particle has constant velocity throughout its existence. We deduce, within the theory, that the total force on it equals zero throughout. But every value for its mass is compatible with this information.

What, then, of those 'definitions' of mass? The core of truth behind them is that mass is experimentally accessible, that is, there are situations in which the data about the basic observables, plus hypotheses about forces and Newton's laws, allow us to calculate the mass. We have here a *counterfactual*: if two bodies have different masses and if they *were* brought near a third body in turn, they *would* exhibit different acceleration. But as the example shows, there are models of mechanics—that is, worlds allowed as possible by this theory—in which a complete specification of the basic observable quantities does not suffice to determine the values of all the other quantities. Thus the same observable phenomena equally fit more than one distinct model of the theory. (Remember that empirical adequacy concerns actual phenomena: what does happen, and not, what would happen under different circumstances.)

I mentioned briefly the axiomatic theories of mechanics developed in this century. We see in them many different treatments of mass. In the theory of McKinsey, Sugar, and Suppes, as I think in Newton's own, each body has a mass. But in Hermes's theory, the mass ratio is so defined that if a given body never collides with another one, there is no number which is the ratio of its mass to that of any other given body. In Simon's, if a body X is never accelerated, the term 'the mass of X' is not defined. In Mackey's any two bodies which are never accelerated, are arbitrarily assigned the same mass.[17]

What explains this divergence, and the conviction of the authors that they have axiomatized classical mechanics? Well, the theories they developed are demonstrably empirically equivalent in exactly the sense I have given that term. Therefore, from the point of view of empirical adequacy, they are indeed equal. The thesis of constructive empiricism, that what matters in science is empirical adequacy, and not questions of truth going beyond that, explains this chapter in foundational studies.

In quantum mechanics we can find a similarly simple, telling example. First, I must make some preliminary remarks. The states are represented by vectors in a Hilbert space, and simple mathematical operations can be performed on these vectors. To calculate the probability of a measurement outcome, the theory tells us to proceed as follows. First we represent the state of the system by means of such a vector in a Hilbert space. Then we multiply that vector by a positive scalar, so that the result is a new vector just like the first, except that it has unit length. Next, we express this unit vector ψ in terms of a family of vectors (eigen-vectors) specially associated with the physical magnitude we are measuring, in this form:

$$\psi = c_1\psi_1 + \ldots + c_i\psi_i + \ldots$$

Each vector ψ_i corresponds to one possible measurement result r_i. The probability that the result will be r_k equals the square of co-efficient c_k (or what corresponds to the square for complex numbers, if that coefficient is complex).

In view of this, it is often said that all positive multiples of ψ represent the same state. For if you begin with $k\psi$ or with $m\psi$, your first step will be to 'normalize', that is, multiply by a scalar so as to arrive at unit vector ψ, There is 'no physical difference', 'the phase has no physical significance' people say; and the reason they give is that the probabilities for measurement outcomes are the same.

Now consider a simple physical operation, rotation. Rotating a system changes its state. There are corresponding operations on vectors, to change the vector that represented the system before, to the one that represents it after the rotation. Let us call the vector operation that corresponds to a rotation through angle x by the name R_x. If the old state was ψ, the new state, after this rotation, is $R_x\psi$. In general, the probabilities for measurement outcomes are

very different in this new state, so there is here in general a genuine physical difference.

One special case is the rotation through 2π radians, a complete circle. Physically, it brings the system back to its original position, in the classical, macroscopic examples we know so well. In the quantum analogue, the operation $R_{2\pi}$ is also quite simple: multiplication by the scalar -1. Hence $R_{2\pi}\psi = -\psi$. If we now expand this new vector in terms of the eigen-vectors ψ_i, we get the coefficients $-c_i$. But if we then calculate the probabilities of measurement results, we square these, so the minuses disappear. Those probabilities are therefore exactly the same for the new state as for the old.

Following the same reasoning as before, we should now say that $R_{2\pi}$, like positive scalar multiplication, simply produces a vector representing the same physical state as the original vector. But there has been a good deal of discussion of this case in the literature, and that apparently easy way out is not available to us.[18] To explain this, we must look at another operation on vectors, namely superposition. If ϕ and ψ are two vectors, then $(k\phi + m\psi)$ is a superposition of them, which is again a vector in the same space, and also represents a physical state. We can sum up the argument in the literature as follows: if ψ and $R_{2\pi}\psi$ really represented exactly the same physical state, then the superposition $(k\phi + m\psi)$ would represent the same state as $(k\phi + mR_{2\pi}\psi)$. That the latter is not so, is easily seen by calculating probabilities for various observables. Klein and Opat designed an experiment on a beam of neutrons in which the observable differences between the two sorts of superpositions were verified: a Fresnel diffraction experiment in which the diffracting object was the boundary between two regions carrying opposite magnetic fields.

What should we conclude from this? The case is quite similar to that of classical mass. If in one possible world, an isolated system is in state ψ and in another it is in state $R_{2\pi}\psi$, no amount of empirical information actually available can tell the observer which of these two worlds he is in. But there is a *counterfactual* statement we are inclined to make about the case: if the system had interacted with another one in such and such a way, the results would have been different in the two cases. The observable phenomena which are actual, however, are the same.

The literature on the measurement problem in quantum mechanics

contains a great deal more tantalizing discussion of the extent to which the observable macroscopic phenomena 'underdetermine' the underlying microscopic state. I refer specifically to Nancy Cartwright's conclusion, based on the quantum thermodynamic approach of Daneri, Loinger, and Prosperi, that a certain superposition of states is indistinguishable in measurement with respect to all macroscopic observables, from a corresponding mixture.[19] Again it is impossible to say, really there is no underdetermination because the two are states between which there is no physical difference. For if systems in these two states *were* subject to interaction with a special third sort of system, the results *would be* different. (This is analogous to the similar point about the masses of actually unaccelerated bodies—the physical difference comes out in the counterfactual assertions we base on what the theory says about what would happen in other, non-actual conditions.) But I am here touching on large and complex issues, and it is hard to say something which is at once simple and uncontroversial.

For the theory of general relativity, we have two studies by Clark Glymour that clearly bring out limits of observation. The first assumes reasonably that measurement divulges only the values of local quantities, and then shows that measurement cannot uniquely determine the global structure of space-time.[20] The second arrives at the same conclusion from the assumption that any observed structure must lie in the absolute past cone of some space-time point.[21] But it is the theory of relativity itself, surely, that forces these assumptions on us, for it forces us to locate observers in space-time, and restricts the information that can reach them.

In this section I have tried to give examples of a very basic and general sort of how, in the description of the world by a physical theory, we can see a division between that description taken as a whole, and the part that pertains to what is observationally determined. The limitations exhibited reach very deeply into the theories in question, and do not relate merely to such 'accidental' limitations as perceptual thresholds and humanly available energy. Realists are generally a bit ambiguous in their feelings toward such limitations. On the one hand they want to emphasize them and say that as a consequence, there is much more to the world described by physics than is dreamt of in the empiricist's philosophy. On the other, they wish to play down the underdetermination, arguing that any precise definition of empirical adequacy and empirical equivalence will lead

to the conclusion that a physical theory is completely adequate only if it is true. My view is that physical theories do indeed describe much more than what is observable, but that what matters is empirical adequacy, and not the truth or falsity of how they go beyond the observable phenomena. And the precise definition of empirical adequacy, because it relates the theory to the *actual* phenomena (and not to anything which *would* happen if the world *were* different, assertions about which have, to my mind, no basis in fact but reflect only the background theories with which we operate) does not collapse into the notion of truth.

§9. A New Picture of Theories

Impressed by the achievements of logic and foundational studies in mathematics at the beginning of this century, philosophers began to think of scientific theories in a language-oriented way. To present a theory, you specified an exact language, some set of axioms, and a partial dictionary that related the theoretical jargon to the observed phenomena which are reported. Everyone knew that this was not a very faithful picture of how scientists do present theories, but held that it was a 'logical snapshot', idealized in just the way that point-masses and frictionless planes idealize mechanical phenomena. There is no doubt that this logical snapshot was very useful to philosophical discussion of science, that there was something to it, that it threw light on some central problems. But it also managed to mislead us.

A picture is only a picture—something to guide the imagination as we go along. I have proposed a new picture, still quite shallow, to guide the discussion of the most general features of scientific theories. To present a theory is to specify a family of structures, its *models*; and secondly, to specify certain parts of those models (the *empirical substructures*) as candidates for the direct representation of observable phenomena. The structures which can be described in experimental and measurement reports we can call *appearances*: the theory is empirically adequate if it has some model such that all appearances are isomorphic to empirical substructures of that model. I am certainly not the first to propose this picture: you can see it at work for example in the writings of Wojcicki and Przelewski in Poland, Dalla Chiara and Toraldo di Francia in Italy, Suppes and Suppe in America.[22] (For instance, what Patrick Suppes calls *empirical algebras* are instances of what I call *appearances,* and

he relates them to parts of the models, and thus describes the relation of theory to data, in much the way I have outlined.)

The form in which theories are actually presented in the technical literature is of course not a sure guide to the form we should conceive them to have. Yet I would still claim support for this proposed way of looking at theories from the actual form of presentation, and indeed from those presentations of theories that are most likely to support the opposing view: the axiomatic ones. In many texts and treatises on quantum mechanics, for example, we find a set of propositions called the 'axioms of quantum theory'. They do not look very much like what a logician expects axioms to look like; on the contrary, they form, in my opinion, a fairly straightforward description of a family of models, plus an indication of what are to be taken as empirical substructures of these models:

Axiom I To every pure state corresponds a vector, and to all the pure states of a system, a Hilbert space of vectors.
Axiom II To every observable (physical magnitude) there corresponds a Hermitean operator on this Hilbert space.
Axiom III The possible values of an observable are the eigenvalues of its corresponding operator.
Axiom IV The expectation value of observable A in state W equals the trace $Tr(AW)$.

To think that this theory is here presented axiomatically in the sense that Hilbert presented Euclidean geometry, or Peano arithmetic, in axiomatic form, seems to me simply a mistake.

Those axioms are instead a description of the models of the theory plus a specification of what the empirical substructures are. The ones I have given are only the beginning for quantum theory of course. For example, as next step there will be principles laid down that say which operator represents energy, or momentum, or how two operators representing two given observable quantities (such as position and momentum) are related to each other. In this further development there is no *a priori* right and wrong; the theory is successfully continued if we can find *some* Hermitean operator to represent energy; and so forth.

When Patrick Suppes first advocated this sort of picture of theories in his studies of mechanics (with the slogan that *philosophy of science should use mathematics, and not meta-mathematics*) he proposed a canonical form for the formulation of theories. This used

set theory. To present classical mechanics, for instance, he would give the definition: 'A system of classical mechanics is a mathematical structure of the following sort . . . ' where the dots are replaced by a set-theoretic predicate. Although I do not wish to favour any mathematical presentation as the canonical one, I am clearly following here his general conception of how, say, the theory of classical mechanics is to be identified.

Looking at Suppes's formulation, it is easy to discuss two points that might otherwise be puzzling. How could, for example, classical mechanics have a model into which all phenomena can be embedded, when that theory does not even mention electricity? The answer is that a mathematical structure might be a system of this or that and also have much structure that does not enter the description of that sort of system. To be a system of mechanics, for example, it must have a set of entities plus a function which assigns each of those a velocity at each instant in time. Well, it could also have in it a function that assigns each of those entities an electric charge; it would still be a system of mechanics, as well as, say, a system of electrodynamics. The second question concerns unintended realizations. Could it not be that some system of mechanics happens to be also a system of optics if we re-label its constituents in a certain way? Well, not in that example perhaps, but there can be examples of that sort. The same formula may govern diffusion of gases and of heat. So perhaps a theory could fail to be empirically adequate as intended, but be so when the phenomena are embedded in its models in an unexpected way? Certainly that is possible.

This suggests that the intention, of which sorts of phenomena are to be embedded in what kinds of empirical substructures, be made part of the theory. I do not think that is necessary. Unintended realizations disappear when we look at larger observable parts of the world; say, the optics and mechanics of moving light sources together. If for a little while some fairly weak theory is empirically adequate, but in a way its advocates are not in a position to notice, that seems hardly an important or frequent enough occurrence to guard against by more complex definitions.

Let me also mention, to complete this part of the discussion, that while I consider Suppes's account of the structure of scientific theories an excellent vehicle for the elucidation of these general distinctions, I do regard it still as relatively shallow. In this book I

am mainly concerned with the relation between physical theories and the world rather than with that other main topic, the structure of physical theory. With respect to the latter I see two main lines of approach: one deriving from Tarski and brought to maturity by Suppes and his collaborators (the *set-theoretic structure approach*) and the other initiated by Weyl and developed by Evert Beth (the *state-space approach*). The first is adamantly extensionalist, the second gives a central role to modality. Each was somewhat language-oriented to begin with, buth both shed these linguistic trappings as they were developed. My own inclination in that subject area has been toward the state-space approach. The general concepts used in the discussion of empirical adequacy, in this chapter, pertain to scientific theories conceived in either way.

Having insisted that the new picture of theories constitutes a radical break with the old, I wish to conclude by outlining some of its peculiar features. Of course, it too provides an idealization: only in foundational studies in physics do we see the family of models carefully described, and only when paradox threatens (as in the measurement problem for quantum mechanics) does anyone try to be very precise about the relation between theory and experiment. That is what it is to be healthy; philosophy is professionally morbid. Still, it is reasonable to draw distinctions and define theoretical relations in terms of the idealization.

If for every model M of T there is a model M' of T' such that all empirical substructures of M are isomorphic to empirical substructures of M', then T is *empirically at least as strong as* T'. Let us abbreviate this to: $T >_e T'$.

We may put this as follows: empirical adequacy, like truth, is 'preserved under watering-down'. I can water a theory down *logically* by disjoining it with some further hypothesis: thus Ptolemy watered down Aristotle's theory of the heavens by asserting that planets did certainly move along circles, but those circles need not have stationary centres. If A is true, so is (A or B). Similarly we can water down a theory *empirically,* either by admitting some new models, or by designating some new parts as empirical substructures in the old models, or both.

Logical strength is determined by the class of models (inversely: the fewer the models the (logically) stronger the theory!) and empirical strength is similarly determined by the classes of empirical substructures. If $T >_e T'$ and $T' >_e T$, then they are *empirically*

equivalent. We may call a theory *empirically minimal* if it is empirically non-equivalent to all logically stronger theories—that is, exactly if we cannot keep its empirical strength the same while discarding some of the models of this theory.

The notions of empirical adequacy and empirical strength, added to those of truth and logical strength, constitute the basic concepts for the semantics of physical theories. Of course, this addition makes the semantics only one degree less shallow than the one we had before. The semantic analysis of physical theory needs to be elaborated further, preferably in response to specific, concrete problems in the foundations of the special sciences. Especially pressing is the need for more finely delineated concepts pertaining to probability for theories in which that is a basic item. I shall return to this subject in another chapter below.

Empirical minimality is emphatically *not* to be advocated as a virtue, it seems to me. The reasons for this point are pragmatic. Theories with some degree of sophistication always carry some 'metaphysical baggage'. Sophistication lies in the introduction of detours via theoretical variables to arrive at useful, adequate, manageable descriptions of the phenomena. The term 'metaphysical baggage' will, of course, not be used when the detour pays off; it is reserved for those detours which yield no practical gain. Even the useless metaphysical baggage may be intriguing, however, because of its potentialities for future use. An example may yet be offered by hidden variable theories in quantum mechanics.[23] The 'no hidden variables' proofs, as I have already mentioned, rest on various assumptions which may be denied. Mathematically speaking there exist hidden variable theories equivalent to orthodox quantum theory in the following sense: the algebra of observables, reduced *modulo* statistical equivalence, in a model of the one is isomorphic to that in a model of the other. It appears to be generally agreed that such theories confront the phenomena exactly by way of these algebras of statistical quantities. On that assumption, theories equivalent in this sense are therefore empirically equivalent. Such hidden variable models have much extra structure, now looked upon as 'metaphysical baggage', but capable of being mobilized should radically new phenomena come to light.

With this new picture of theories in mind, we can distinguish between two epistemic attitudes we can take up toward a theory. We can assert it to be true (i.e. to have a model which is a faithful

replica, in all detail, of our world), and call for belief; or we can simply assert its empirical adequacy, calling for acceptance as such. In either case we stick our necks out: empirical adequacy goes far beyond what we can know at any given time. (All the results of measurement are not in; they will never all be in; and in any case, we won't measure everything that can be measured.) Nevertheless there is a difference: the assertion of empirical adequacy is a great deal weaker than the assertion of truth, and the restraint to acceptance delivers us from metaphysics.