To Leonidas Hegenberg
with kind regards
from Karl Popper

Reprinted from
Studies in the Philosophy of Biology
edited by F. J. Ayala and T. Dobzhansky
and published by the Macmillan Press Ltd. 1974

CLH, ?

16. Scientific Reduction and the Essential Incompleteness of All Science

K. R. POPPER

I

The thesis from which I start* is that, for a conference convened by biologists, the outstanding questions of reduction are three:

(1) Can we reduce, or hope to reduce, biology to physics, or to physics and chemistry?

(2) Can we reduce to biology, or hope to reduce to biology, those subjective conscious experiences which we may ascribe to animals and, if question (1) is answered in the affirmative, can we reduce them further to physics and chemistry?

(3) Can we reduce, or hope to reduce, the consciousness of self and the creativeness of the human mind to animal experience, and thus, if questions (1) and (2) are answered in the affirmative, to physics and chemistry?

It is obvious that the replies to these three questions (to which I shall turn later in the paper) will partly depend on the meaning of the word 'reduce'. But for reasons which I have given elsewhere (1945, vol. II, 9–21) I am opposed to the method of meaning analysis and to the attempt to solve serious problems by definitions. What I propose to do instead is this.

I will begin by discussing some examples of successful and unsuccessful reductions in the various sciences, and especially the reduction of chemistry to physics; and also the residues left by these reductions.

In the course of this discussion, I will defend three theses. First, I will suggest that scientists have to be reductionists in the sense that nothing is as great a success in science as a successful reduction (such as Newton's reduction

* I am greatly indebted to David Miller and Jeremy Shearmur for their comments on an earlier draft of this paper.

—or rather explanation[1]—of Kepler's and Galileo's laws to his theory of gravity, and his correction of them; see my (1957)). A successful reduction is, perhaps, the most successful form conceivable of all scientific explanations, since it achieves what Meyerson (1908, 1930) stressed: an identification of the unknown with the known. Let me mention however that by contrast with a reduction, an explanation with the help of a new theory explains the known—the known problem—by something unknown: a new conjecture (see my (1963), 63, 102, 174).

Secondly, I will suggest that scientists, whatever their philosophical attitude towards holism, *have* to welcome reductionism as a *method*: they have to be either naive or else more or less critical reductionists; indeed, somewhat desperate critical reductionists, I shall argue, because hardly any major reduction in science has ever been *completely* successful: there is almost always an unresolved residue left by even the most successful attempts at reduction.

Thirdly, I shall contend that there do not seem to be any good arguments in favour of *philosophical* reductionism, while, on the contrary, there are good arguments against essentialism, with which philosophical reductionism seems to be closely allied. But I shall also suggest that we should, nevertheless, on methodological grounds, continue to attempt reductions. The reason is that we can learn an immense amount even from unsuccessful or incomplete attempts at reduction, and that problems left open in this way belong to the most valuable intellectual possessions of science: I suggest that a greater emphasis upon what are often regarded as our scientific failures (or, in other words, upon the great open problems of science) can do us a lot of good.

II

Apart from Newton's, one of the very few of the reductions known to me which have been almost completely successful is the reduction of rational fractions to ordered pairs of natural numbers. (That is, to relations or ratios between them.) It was achieved by the Greeks, although one might say that even this reduction left a *residue* which was dealt with only in the twentieth century (with the successful reduction, by Wiener (1914) and Kuratowski (1920), of the ordered pair to an unordered pair of unordered pairs; moreover, one should be aware that the reduction is one to *sets* of equivalent pairs, rather than to pairs themselves). It encouraged the Pythagorean cosmological research programme of arithmetisation which, however, broke down with the

[Note added in proof] In the text of this paper I have disregarded—perhaps carelessly, or because I dislike terminological minutiae—the distinction that can well be made between explanation in general, and reduction in the sense of an explanation by way of an established or more 'fundamental' theory. A distinction of major interest, I suppose, would be that between an explanation of something known by a new (unknown) theory on the one hand, and a reduction to an old (known) theory on the other. I have added an allusion to this distinction here to the text, and also the footnotes and Postscript, in the hope of avoiding possible misunderstandings.

proof of the existence of irrationals such as the square roots of 2, 3, or 5 (*cf.* my (1950), vol. I, ch. 6, n. 9; and (1963), ch. 2, 75–92). As I have suggested (*loc. cit.*) Plato replaced the cosmological research programme of arithmetisation by one of geometrisation, and this programme was carried on successfully from Euclid to Einstein. However, the invention of the calculus by Newton and Leibniz (and the problem of excluding the paradoxical results which their own intuitive methods failed to exclude) created the need for a new arithmetisation—a new reduction to natural numbers. And in spite of the most spectacular successes of the nineteenth and early twentieth centuries, we can say now, I believe, that this reduction has not been fully successful.

To mention only one unresolved residue, a reduction to a sequence of natural numbers or to a set in the sense of modern set theory is not the same as, or even similar to, a reduction to a set of equivalent ordered pairs of natural numbers. As long as the idea of a set was used naively and purely intuitively (as by Cantor) this was perhaps not obvious. But the paradoxes of infinite sets (discussed by Bolzano, Cantor and Russell) and the need to axiomatise set theory showed, to say the very least, that the reduction achieved was not a straightforward arithmetisation—a reduction to natural numbers—but a reduction to axiomatic set theory; and this turned out to be a highly sophisticated and somewhat perilous enterprise.

To sum up this example, the programme of arithmetisation—that is, of the reduction of geometry and the irrationals to natural numbers—has partly failed. But the number of unexpected problems and the amount of unexpected knowledge brought about by this failure are overwhelming. This, I shall contend, may be generalised: even where we do not succeed as reductionists, the number of interesting and unexpected results we may acquire on the way to our failure can be of the greatest value.

III

I have briefly hinted at the failure of the attempted reduction of the irrationals to natural numbers, and I have also indicated that attempts at reduction are part of the activities of scientific and mathematical explanation, simplification and understanding.

I will now discuss in a little more detail the successes and failures of attempted reductions in physics, and in particular the partial successes of the reduction of macrophysics to microphysics and of chemistry to both microphysics and macrophysics.

IV

I have elsewhere (1956, 365–72; 1963, ch. 3, 103–7) given the name 'ultimate explanation' to the attempt to explain or reduce things by an appeal to something that is neither in need of, nor capable of, further explanation, more especially an 'essence' or a 'substance' (*ousia*).

A striking example is the Cartesian reduction of the whole of the physics of inanimate bodies to the idea of an *extended substance*; a substance (matter) with only one essential property; that is, spatial extension.

This attempt to reduce the whole of physics to the one apparently essential property of matter was highly successful in so far as it gave rise to an understandable picture of the physical universe. The Cartesian physical universe was a moving clockwork of vortices in which each 'body' or 'part of matter' pushed its neighbouring part along, and was pushed along by its neighbour on the other side. Matter alone was to be found in the physical world, and all space was filled by it. In fact, space too was reduced to matter, since there was no empty space but only the essential spatial extension of matter. And there was only one purely physical mode of causation: *all causation was push*, or action by contact.

This way of looking at the world was found satisfactory even by Newton, though he felt compelled to introduce by his theory of gravity a new kind of causation: *attraction*, or action at a distance.

It was the almost incredible explanatory and predictive success of Newton's theory which destroyed the Cartesian reduction programme. Newton himself, I have elsewhere conjectured (1969, 107, n. 21), attempted to carry out the Cartesian reduction programme by explaining gravitational attraction by the 'impulse' (radiation pressure combined with an umbrella effect) of a cosmic particle bombardment (the attempt is usually linked with the name of Le Sage). But I also conjectured that Newton became aware of the fatal objection to this theory. Admittedly it would reduce attraction and action at a distance to push and to action by contact; but it would also mean that all moving bodies would move in a resisting medium which would act as a brake on their movement (consider the excess push of rain on the windscreen of a car over that on the rear window) and which would thus invalidate Newton's use of the law of inertia.

Thus, in spite of its intuitive attractiveness, and in spite of Newton's own rejection as 'absurd' of the view that attraction at a distance could be an essential property of matter, the attempt at an ultimate reduction of attraction to push breaks down.

V

We have here our first and very simple example of a promising scientific reduction and its failure, and of how much one can learn by attempting a reduction and discovering that it fails.

(I conjecture that this failure was the immediate reason why Newton described space as the sensorium of God. Space was 'aware', so to speak, of the distribution of all bodies: it was, in a sense, omniscient. It was also omnipresent, for it transmitted this knowledge with infinite velocity to all locations at every moment of time. Thus space, sharing at least two characteristic properties of the divine essence, was itself part of the divine essence.

This, I suggest, was another attempt by Newton at an essentialist ultimate explanation.)

The Cartesian reduction may be taken as an illustration of my remark that for methodological reasons we have to attempt reductions. But it may also give an indication of the reason why I suggest that as reductionists we must not be sanguine but can be only somewhat despairing concerning the complete success of our attempted reductions.

VI

It is clear, I think, that the Cartesian attempt (which, if I am right in my historical conjecture, was also a Newtonian attempt) to reduce everything in the physical world to extension and push became a failure when it was judged against the success of Newton's theory of gravity. And the success was so great that Newtonians, beginning with Roger Cotes, began to look upon Newtonian theory itself as an ultimate explanation and thus at *gravitational attraction* as an essential property of matter, in spite of Newton's own views to the contrary. But Newton had seen no reason why *extension* (of his atoms) and *inertia* should not be essential properties of mass (*cf.* my (1956), 370, or (1963), 106*f.*). Thus we can say that Newton was clearly aware of the distinction, later stressed by Einstein, between inertial and gravitational mass, and of the problem opened by their proportionality (or equality); a problem which, because of the obscurantism of the essentialist approach, was almost lost sight of between Newton and Eötvös or even Einstein.

Einstein's Special Relativity theory destroyed the essentialist identity of inertial and gravitational mass, and this is the reason why he tried to explain it, somewhat *ad hoc*, by his principle of equivalence. But when it was discovered (first by Cornelius Lanczos) that Einstein's equations of gravitation led by themselves to the principle, previously separately assumed, that gravitating bodies move on a space–time geodesic, the principle of inertia was in fact reduced to the equations of gravitation and thus inertial mass to gravitational mass. (I believe that Einstein, though strongly impressed by the importance of this result, did not fully accept that it solved Mach's central problem—the explanation of inertia—in a more satisfactory way than the famous but far from unambiguous 'Mach principle': the principle that the inertia of each body is due to the combined effect of all the other bodies in the universe. To Einstein's disappointment, this principle was, at least in some of its interpretations, incompatible with General Relativity which, for a space empty of all bodies, yields Special Relativity, in which the law of inertia, contrary to Mach's suggestion, is still valid.)

Here we have what I regard as a most satisfying example of a successful reduction: the reduction of a generalised principle of inertia to a generalised principle of gravitation. But it has been rarely considered in this light; not even by Einstein, though he strongly felt the significance of a result which, from a purely mathematical point of view, could be regarded as elegant but not as particularly important. For the dependence or independence of an axiom within a system of axioms is in general not of more than formal interest. Why should it matter, therefore, whether the law of motion on a geodesic had to be assumed as a separate axiom or could be derived from the rest of gravitational theory? The answer is that by its derivation, the identity of inertial and gravitational mass was *explained*, and the former reduced to the latter.

In this way one might say that Newton's great problem of action at a distance (couched in the phraseology of essentialism) was solved not so much by the finite velocity of Einsteinian gravitational action as by the reduction of inertial matter to gravitational matter.

VII

Newton and the Newtonians knew, of course, about the existence of magnetic and electrical forces; and until at least the beginning of the twentieth century, attempts were made to reduce electromagnetic theory to Newtonian mechanics, or to a modified form of it.

The outstanding problem in this development was the reduction of *prima facie* non-central forces (Oersted forces) to central forces, the only ones which seemed to fit into even a modified Newtonian theory. The outstanding names in this development were Ampère and Weber.

Maxwell too began by trying to reduce Faraday's electromagnetic field of (lines of) forces to a Newtonian mechanism or model of the luminiferous ether. But he gave up the attempt (though not the luminiferous ether as the carrier of the electromagnetic field). Helmholtz also was attracted by a Newtonian and partly Cartesian reduction programme, and when he suggested to his pupil, Heinrich Hertz, that he should work on this problem, Helmholtz seems to have done so in the hope of saving the research programme of mechanics. But he accepted Hertz's confirmation of Maxwell's equations as crucial. After Hertz and J. J. Thomson, precisely the opposite research programme became more attractive—the programme of reducing mechanics to electromagnetic theory.

VIII

The electromagnetic theory of matter—that is, the reduction of both mechanics and chemistry to an electromagnetic theory of atomism—was strikingly successful from at least 1912, the year of Rutherford's planetary or nuclear atom model, until about 1932.

In fact, quantum mechanics (or 'the new quantum theory', as it was once called) was, until at least 1935, simply another name for what was then regarded as the final form of the reduction of mechanics to the new *electromagnetic theory of matter*.

In order to realise how important this reduction appeared to leading physicists even shortly before quantum mechanics, I may quote Einstein who wrote (1920; 1922, 24; see also my (1967) where I discuss the same point): '. . . according to our present conceptions the elementary particles [that is, electrons and protons] are . . . *nothing else* than condensations of the electromagnetic field . . . , our . . . view of the universe presents two realities . . . , namely, gravitational ether and electromagnetic field, or—as they might also be called—space and matter.'

Note the 'nothing else' which I have italicised because it is characteristic of reduction in the grand style. Indeed, to the end of his life, Einstein tried to unify the gravitational and the electromagnetic fields in a unified field theory, even after his view of 1920 had been superseded—or rather, had broken down (especially owing to the discovery of nuclear forces).

What amounts, essentially, to the same reductionist view was accepted at that time (1932) by almost all leading physicists: Eddington and Dirac in England and, besides Einstein, Bohr, de Broglie, Schrödinger, Heisenberg, Born and Pauli on the continent of Europe. And a very impressive statement of the view was given by Robert A. Millikan (1932, 46), then of the Californian Institute of Technology:

> Indeed, nothing more beautifully simplifying has ever happened in the history of science than the whole series of discoveries culminating about 1914 which finally brought practically universal acceptance to the theory that the material world contains but two fundamental entities, namely, positive and negative electrons, exactly alike in charge, but differing widely in mass, the positive electron—now usually called a proton—being 1850 times heavier than the negative, now usually called simply the electron.

This reductionist passage was written in the very nick of time: it was in the same year that Chadwick (1932) published his discovery of the neutron, and that Anderson (1933) first discovered the positron. Yet some of the greatest physicists, such as Eddington (1936), continued to believe, even after Yukawa's suggestion of the existence of what was to be called the meson (1935), that with the advent of quantum mechanics the electromagnetic theory of matter had entered into its final state and that all matter consisted of electrons and protons.

IX

Indeed, the reduction of mechanics and of chemistry to the electromagnetic theory of matter seemed almost perfect. What had appeared to Descartes and Newton as the space-filling essence of matter, and as Cartesian push had been reduced (as Leibniz had demanded long ago) to *repulsive forces*—the forces exerted by negative electrons upon negative electrons. The electrical neutrality of matter was explained by the equal number of positive protons and negative electrons; and the electrification (ionisation) of matter was explained by a loss of electrons from (or excess of electrons in) the planetary electron shell of the atom.

Chemistry had been reduced to physics (or so it seemed) by Bohr's quantum theory of the periodic system of elements, a theory which was ingeniously perfected by the use of Pauli's exclusion principle; and the theory of chemical composition, and of the nature of covalent chemical bonds, was reduced by Heitler and London (1927) to a theory of (homeopolar) valency which also made use of Pauli's principle.

Although matter was revealed to be a complex structure rather than an irreducible substance, there had never before been such unity in the universe of physics, or such a degree of reduction.

Nor has it ever been achieved again since.

True, we still believe in the reduction of Cartesian push to electromagnetic forces; and Bohr's theory of the periodic system of elements, though considerably changed by the introduction of isotopes, has largely survived. But everything else in this beautiful reduction of the universe to an electromagnetic universe with two particles as stable building blocks has by now disintegrated. Emphatically, we have learned an immense number of new facts in the process of this disintegration: this is one of my main theses. But the simplicity of the reduction has disappeared.

This process, which started with the discovery of neutrons and of positrons, has continued with the discovery of new elementary particles ever since. But particle theory is not even the main difficulty. The real disruption is due to the discovery of new kinds of forces, especially of short-range nuclear forces, irreducible to electromagnetic and gravitational forces.

Gravitational forces did not trouble the physicists very much in those days, because they had just been explained away by General Relativity, and it was hoped that the dualism of gravitational and electromagnetic forces would be superseded by a unified field theory. But now we have at least four very different and still irreducible kinds of forces in physics: gravitation, weak decay interaction, electromagnetic forces and nuclear forces.

X

Thus Cartesian mechanics—once regarded by Descartes and Newton as the basis to which all else was to be reduced—was, and still is, successfully reduced to electromagnetism. But what about the admittedly most impressive reduction of chemistry to quantum physics?

Let us assume for argument's sake that we have a fully satisfactory reduction to quantum theory of chemical bonds (both of covalent or twin electron bonds and of non-covalent, for example plug-and-hole, bonds), in spite of the telling remark of Pauling (1959), author of *The Nature of the Chemical Bond*, that he was unable to 'define' (or state precisely) what the nature of the chemical bond was. Let us further assume for argument's sake that we have a fully satisfactory theory of nuclear forces, of the periodic system of the elements and their isotopes, and especially of the stability and instability of the heavier nuclei. Does this constitute a fully satisfactory reduction of chemistry to quantum mechanics?

I do not think it does. An entirely new idea has to be brought in, an idea which is somewhat foreign to physical theory: the idea of evolution, of the history of our universe, of cosmogony.

This is so because the periodic table of the elements and the (reformulated) Bohr theory of the periodic system explain the heavier nuclei as being composed of lighter ones; ultimately as being composed of hydrogen nuclei (protons) and neutrons (which in turn *might* be regarded as a kind of composition of protons and electrons). And this theory assumes that the heavier elements have a history—that the properties of their nuclei actually result from a rare process which makes several hydrogen nuclei fuse into heavier nuclei, under conditions which are only rarely encountered in the cosmos.

We have much evidence in favour of the view that this really happened and still happens; that the heavier elements have an evolutionary history and that the fusion process by which heavy hydrogen is transformed into helium is the main source of the energy of our own sun and also of the hydrogen bomb. Thus helium and all the heavier elements are the result of cosmological evolution. Their history, and especially the history of the heavier elements, is, according to present cosmological views, a strange one. The heavier elements are at present regarded as the products of supernovae explosions. Since helium, according to some recent estimates, forms twenty-five per cent of all matter by mass and hydrogen two-thirds or three-quarters of all matter by mass, all the heavier nuclei appear to be extremely rare (together perhaps one or two per cent by mass). Thus the earth and presumably the other planets of our solar system are made mainly of very rare (and I should say very precious) materials.

At present the most widely accepted theory of the origin of the universe[2]—that of the hot big bang—claims that most of the helium is the product of the big bang itself: that it was produced within the very first minute of the existence of the expanding universe. The precariousness of the scientific status of this speculation (originally due to Gamow) need not be stressed. And since we have to appeal to theories of this kind in our attempts to reduce chemistry to quantum mechanics, it can hardly be claimed that this reduction has been carried out without residue.

The truth is that we have reduced chemistry, at least in part, to cosmology rather than to physical theory. Admittedly, modern classical relativistic cosmology started as an applied physical theory; but, as Bondi has stressed, these times seem now to be over and we must face the fact that some of our ideas (for example, those that started with Dirac and Jordan) could almost be described as attempts to reduce physical theory to cosmogony. And both cosmology and cosmogony, though immensely fascinating parts of physics, and though they are becoming better testable, are still almost borderline

2 [Added after the conference]: This theory may now be threatened by the new theory of redshifts proposed by J. C. Pecker, A. P. Roberts and J. P. Vigier, Non-velocity redshifts and photon-photon interactions, *Nature*, 237 (1972), 227–9.

cases of physical science, and hardly yet mature enough to serve as the bases of the reduction of chemistry to physics. This is one reason why I regard the so-called reduction of chemistry to physics as incomplete and somewhat problematic; but of course I welcome all these new problems.

XI

But there is a second residue of the reduction of chemistry to physics. Our present view is that hydrogen alone, and especially its nucleus, is the building material of all the other nuclei. We believe that the positive nuclei strongly repel each other electrically down to very short distances, but that for still shorter distances (achievable only if the repulsion is overcome by tremendous velocities) they attract each other by nuclear forces.

But this means that we attribute to the hydrogen nucleus relational properties which are inoperative in the overwhelming majority of the conditions in which hydrogen nuclei exist in our universe. That is to say, these nuclear forces are potentialities that become operative only under conditions which are extremely rare: under tremendous temperatures and pressures. But this means that the theory of the evolution of the periodic table looks very much like a theory of essential properties which have the character of *predestination, or of a preestablished harmony.*[3] At any rate, a solar system like ours depends, according to present theories, on the preexistence of these properties, or rather, potentialities.

Moreover, the theory of the origin of the heavier elements in explosions of supernovae introduces *a second kind of predestination or preestablished harmony*. For it amounts to the assertion that gravitational forces (apparently the weakest of all, and so far unconnected with nuclear or electromagnetic forces) can, in big accumulations of hydrogen, become so powerful as to overcome the tremendous electrical repulsion between the nuclei, and to make them fuse due to the action of the nuclear forces. Here the harmony is between the inherent potentialities of nuclear forces and of gravitation. I do not want to assert the untruth of any philosophy of preestablished harmony. But I do not think that a preestablished harmony can be regarded as a satisfactory reduction; and I suggest that the appeal to it is an admission of the failure of the method of reducing one thing to another.

Thus the reduction of chemistry to physics is far from complete, even if we admit somewhat unrealistically favourable assumptions. Rather, this reduction assumes a theory of cosmic evolution or cosmogony, and in addition two kinds of preestablished harmony, in order to allow sleeping potentialities, or relative propensities of low probability built into the hydrogen atom, to become activated. It appears, I suggest, that we should

[3] I have used the term 'preestablished harmony' here to stress that our explanation is not in terms of the manifest physical properties of the hydrogen atom. Rather, a hitherto unknown and unsuspected property of the hydrogen nucleus was postulated, and used as an explanation.

recognise that we are operating with the ideas of *emergence* and of *emergent properties*.[4] In this way we see that this very interesting reduction has left us with a strange picture of the universe—strange, at any rate to the reductionist; which is the point I wanted to make in this section.

XII

To sum up what has been said so far: I have tried to make the problem of reduction clear with the help of examples, and I have tried to show that some of the most impressive reductions in the history of the physical sciences are far from completely successful, and leave a residue. One might claim (but see footnote 1 above) that Newton's theory was a complete successful reduction of Kepler's and Galileo's. But even if we assume that we know much more physics than we do, and that we have a unified field theory which yields with high approximation General Relativity, quantum theory and the four kinds of forces as special cases (this is perhaps a claim implicit in Mendel Sachs's unified field theory), even then we can say that chemistry has not been reduced without residue to physics. In fact the so-called reduction of chemistry is to a physics that assumes evolution, cosmology and cosmogony, and the existence of emergent properties.

On the other hand, in our not fully successful attempts at reduction, especially of chemistry to physics, we have learned an incredible amount. New problems have given rise to new conjectural theories, and some of these, such as nuclear fusion, have not only led to corroborating experiments, but to a new technology. Thus from the point of view of method, our attempted reductions have led to tremendous successes, even though it can be said that the attempted reductions have, as such, usually failed.

XIII

The story here told and the lesson here drawn from it will hardly strike a biologist as unexpected. In biology too, reductionism (in the form of physicalism or materialism) has been extremely successful, though not fully successful. But even where it has not succeeded, it has led to new problems and to new solutions.

I might perhaps express my view as follows. As a philosophy, reductionism is a failure. From the point of view of method, the attempts at detailed reductions have led to one staggering success after another, and its failures have also been most fruitful for science.

It is perhaps understandable that some of those who have achieved these scientific successes have not been struck by the failure of the philosophy. Perhaps my analysis of the success and of the failure of the attempt to reduce chemistry completely to quantum physics may give them pause, and may make them look at the problem again.

[4] I use here the term 'emergent' to indicate an apparently unforeseeable evolutionary step.

XIV

The main points made so far may be regarded as an elaboration of a brief remark made by Jacques Monod in the Preface to his *Chance and Necessity* (1970; 1971, xii): 'Nor can everything in chemistry be predicted or resolved by means of the quantum theory [or reduced to quantum theory] which, beyond any question, underlies all chemistry.' In the same book Monod also puts forward a suggestion (not an assertion, to be sure) concerning the origin of life, which is very striking, and which we may consider from the point of view reached here. Monod's suggestion is that life emerged from inanimate matter by an extremely improbable combination of chance circumstances, and that this may not merely have been an event of low probability but of zero probability—in fact, a *unique* event.

This suggestion is experimentally testable (as Monod pointed out in a recent discussion with Eccles). Should we succeed in producing life under certain well-defined experimental conditions, then the hypothesis of the uniqueness of the origin of life would be refuted. Thus the suggestion is a testable scientific hypothesis, even though it may not look like one at first sight.

What, besides, makes Monod's suggestion plausible? There is the fact of the uniqueness of the genetic code, but this could be, as Monod points out, the result of natural selection. What makes the origin of life and of the genetic code a disturbing riddle is this: the genetic code is without any biological function unless it is translated; that is, unless it leads to the synthesis of the proteins whose structure is laid down by the code. But, as Monod points out, the machinery by which the cell (at least the non-primitive cell which is the only one we know) translates the code 'consists of at least fifty macromolecular components *which are themselves coded in DNA*' (Monod, 1970; 1971, 143). Thus the code cannot be translated except by using certain products of its translation. This constitutes a really baffling circle: a vicious circle, it seems, for any attempt to form a model, or a theory, of the genesis of the genetic code.

Thus we may be faced with the possibility that the origin of life (like the origin of the universe) becomes an impenetrable barrier to science, and a residue to all attempts to reduce biology to chemistry and physics. For even though Monod's suggestion of the uniqueness of life's origin is refutable—by attempts at reduction, to be sure—if true, it would amount to a denial of any fully successful reduction. With this suggestion Monod, who is a reductionist for reasons of method, arrives at the position which, I believe, is the one forced upon us all in the light of our earlier discussion of the reduction of chemistry to physics. It is the position of a critical reductionist who continues with attempted reductions even if he despairs of any ultimate success. Yet it is in going forward with attempted reductions, as Monod stresses elsewhere in his book, rather than in any replacement of reductionist methods by 'holistic' ones, that our main hope lies—our hope of learning more about old problems and of discovering new problems, which in turn may lead to new solutions, to new discoveries.

I do not want to discuss holism in any detail here, but a few words may be needed. The use of holistic experimental methods (such as cell transplantation in embryos), though inspired by holistic thought, may well be claimed to be methodologically reductionist. Holistic theories are, on the other hand, trivially needed in the description of even an atom or a molecule, not to speak of an organism or of a gene population. There is no limit to the variety of possibly fruitful conjectures, whether holistic or not.[5] In view of my main thesis, doubt arises only about the character of experimental methods in biology: whether they are not all, more or less, of a reductionist character. (A similar situation arises, incidentally, as David Miller reminds me, with regard to deterministic and indeterministic theories. Though we must, I think, be metaphysical *indeterminists*, methodologically we should still search for deterministic or causal laws—except where the problems to be solved are themselves of a probabilistic character.)

XV

I should like to point out that even if Monod's suggestion of the uniqueness of the origin of life should be refuted by the production of life from inanimate matter under definite experimental conditions, this would not amount to a complete reduction. I do not wish to argue *a priori* that a reduction is impossible; but we have produced life from life for a long time without understanding what we have been doing, and before we had even an inkling of molecular biology or the genetic code. Thus it is certainly possible that we may produce life from inanimate matter without a full physicochemical understanding of what we are actually doing; for example, how we managed to break the vicious circle inherent in the translation of the code.

At any rate we can say that the undreamt-of breakthrough of molecular biology has made the problem of the origin of life a greater riddle than it was before: we have acquired new and deeper problems.

XVI

As I have tried to show, the attempt to reduce chemistry to physics demands the introduction of a theory of evolution into physics; that is, a recourse to the history of our cosmos. A theory of evolution is, it appears, even more indispensable in biology. And so is, in addition, the idea of purpose or teleology or (to use Monod's term) of teleonomy, or the very similar idea of problem solving; an idea which is quite foreign to the subject matter of the non-biological sciences (even though the role played in these sciences by maxima and minima and by the calculus of variations has been regarded as remotely analogous).

[5] This is now stressed in the second point of the Postscript to the present paper.

It was of course the great achievement of Darwin to show that there is a possibility of explaining teleology in non-teleological or ordinary causal terms. Darwinism is the best explanation we have. There are not, at the moment, any seriously competing hypotheses (*cf.* my (1961) and (1966*a*)).

XVII

Problems and problem solving seem to emerge together with life (see my (1966*a*)). Even though there is something like natural selection at work prior to the origin of life—for example, a selection of the more stable elements owing to the radioactive destruction of the less stable ones—we cannot say that for atomic nuclei, survival is a 'problem' in any sense of this term. And the close analogy between crystals and microorganisms and their molecular parts (organelles) breaks down here too. Crystals have no problems of growth or of propagation or of survival. But life is faced with the problems of survival from the very beginning. Indeed, we can describe life, if we like, as problem solving, and living organisms as the only problem solving complexes in the universe. (Computers are *instrumental in* problem solving but not, in this sense, problem solvers.)

This does not mean that we have to ascribe to all life a *consciousness* of the problems to be solved: even on the human level we constantly solve many problems, such as keeping our balance, without becoming aware of them.

XVIII

There can be little doubt that animals possess consciousness and that, at times, they can even be conscious of a problem. But the emergence of consciousness in the animal kingdom is perhaps as great a mystery as is the origin of life itself.

I do not want to say more about this than that panpsychism, or hylozoism, or the thesis that matter is, generally, endowed with consciousness (of a low degree), does not seem to me to help in the least. It is, if taken at all seriously, another theory of predestination or of a preestablished harmony. (It was of course part of Leibniz's original form of his theory of preestablished harmony.) For in nonliving matter, consciousness has no function at all; and if (with Leibniz, Diderot, Buffon, Haeckel and many others) we attribute consciousness to nonliving particles (monads, atoms) then we do so in the vain hope that it will help to explain the presence of those forms of consciousness which have some function in animals.

For there can be little doubt that consciousness in animals has some function, and can be looked at as if it were a bodily organ. We have to assume, difficult as this may be, that it is a product of evolution, of natural selection.

Although this might constitute a programme for a reduction, it is not itself a reduction, and the situation for the reductionist looks somewhat desperate; which explains why reductionists have either adopted the hypothesis of panpsychism or why, more recently, they have denied the existence of consciousness (the consciousness say, of a toothache) altogether.

Though this behaviourist philosophy is quite fashionable at present, a theory of the nonexistence of consciousness cannot be taken any more seriously, I suggest, than a theory of the nonexistence of matter. Both theories 'solve' the problem of the relationship between body and mind. The solution is in both cases a radical simplification: it is the denial either of body or of mind. But in my opinion it is too cheap (see my (1970), 7–9). I shall say a little more about this second 'outstanding question' and especially about panpsychism in Section XXI where I criticise psychophysical parallelism.

XIX

Of the three 'outstanding questions of reduction' listed at the beginning of this paper I have briefly touched upon two. I am now coming to the third one, the question of the reduction of the human consciousness of self and the creativeness of the human mind.

As Sir John Eccles has often stressed, this third question is the problem of the 'mind–brain liaison'; and Jacques Monod calls the problem of the human central nervous system the 'second frontier', comparing its difficulty with the 'first frontier', the problem of the origin of life.

No doubt this second frontier is a dangerous region to dwell in, especially for a lay biologist; nevertheless I may say that the attempts at a partial reduction seem to me more hopeful in this region than in that of the second question. As in the region of the first question, it seems to me that more new problems can be discovered here with reductionist methods, and perhaps even solved, than in the region of the second question—a region which looks to me comparatively sterile. I hardly need to stress that a completely successful reduction in any of the three regions seems to me most unlikely, if not impossible.

With this, it may perhaps be said, I have fulfilled my promise to discuss, or at any rate mention, those three outstanding questions of reduction listed at the beginning of this paper. But I wish to say a little more about the third of them—about the body–mind problem, or mind–body problem—before proceeding to my thesis of the incompletability of all science.

XX

I regard the problem of the emergence of consciousness in animals (question 2), of understanding it and, perhaps, of reducing it to physiology, as most likely insoluble; and I feel similarly about the further problem of the emergence of the specifically human consciousness of self (question 3)—that is, the body–mind problem. But I do think that we can throw at least some light upon the problem of the human self.

I am, in many ways, a Cartesian dualist (see my (1953)), even though I should prefer to describe myself as a pluralist; and of course I do not subscribe to either of Descartes's two substances. Matter, we have seen, is no ultimate substance with the essential property of extension, but consists of complex structures about whose constitution we know a great deal, including an explanation of its 'extension': that it takes up space by electrical repulsion.

My first thesis is that the human consciousness of self, with its apparently irreducible unity, is highly complex, and that it may perhaps be, in part, explicable.

In a course of lectures given at Emory University in May 1969 I suggested (as I had done some years before in lectures at the London School of Economics) that the higher human consciousness, or consciousness of selfhood, is absent in animals. I also suggested that Descartes's conjecture that locates the human soul in the pineal gland may not be as absurd as it has often been represented, and that, in view of Sperry's results with divided brain hemispheres (1964; see also Eccles (1970), 73–9), the location is to be looked for in the speech centre, in the left hemisphere of the brain. As Eccles has more recently informed me (1972), Sperry's later experiments (not known to me at the time) support this guess to a degree: the right brain may be described as that of a very clever animal while only the left brain appears to be human, and aware of selfhood.

I had based my guess upon the role which I ascribe to the development of a specifically human language.

All animal language—indeed, almost all animal behaviour—has an *expressive* (or symptomatic) and a *communicative* (or signalling) function, as Karl Bühler has pointed out. But human language has, besides, some further functions, which are characteristic of it and make it a 'language' in a narrower and more important sense of the word. Bühler drew attention to the basic *descriptive* function of human language, and I pointed out later (1949, 1953) that there are further functions (such as prescriptive, advisory and so on) of which the most important and characteristic one for human beings is the *argumentative* function. (Professor Alf Ross (1972) points out that many other functions may be added, for example, those of giving orders or making requests or promises.)

I do not think (and I never did think) that any of these functions are reducible to any of the others, least of all the two higher functions (description and argument) to the two lower ones (expression and communication). These, incidentally, are always present, which may perhaps be the reason why many philosophers mistake them for properties which are characteristic of human language.

My thesis is that, with the higher functions of the human language a new world emerges: the world of the products of the human mind. I have called it 'world 3' (following a suggestion of Sir John Eccles: originally I called it the 'third world'). I call the world of physical matter, fields of force, and so on, 'world 1'; the world of conscious and perhaps also subconscious experience 'world 2'; and 'world 3' especially the world of spoken (written or printed) language, like story telling, myth making, theories, theoretical problems, mistakes and arguments. (The worlds of artistic products and of social institutions may either be subsumed under world 3 or be called 'world 4' and 'world 5': this is just a matter of taste.)

I introduce the terms 'world 1', 'world 2' and 'world 3' in order to emphasise the (limited) *autonomy* of these regions. Most materialists or physicalists or reductionists assert that, of these three worlds, only world 1 really exists, and that it is therefore autonomous. They replace world 2 by behaviour, and world 3, more particularly, by verbal behaviour. (This, as indicated above, is just one of those all too easy ways of solving the body–mind problem: the way of denying the existence of the human mind and of a human consciousness of self—that is, of those things which I regard as some of the most remarkable and astonishing in the universe; the other equally easy way out is Berkeley's and Mach's immaterialism: the thesis that only sensations exist, and that matter is just a 'construct' out of sensations.)

XXI

There are in the main four positions with respect to the interrelationship between the body, or the brain, and the mind.

(1) A denial of the existence of the world 1 of physical states; that is, immaterialism, as held by Berkeley and Mach.

(2) A denial of the existence of the world 2 of mental states or events, a view common to certain materialists, physicalists and philosophical behaviourists, or philosophers upholding the identity of brain and mind.

(3) An assertion of a thoroughgoing parallelism between mental states and states of the brain; a position that is called 'psychophysical parallelism'. This was first introduced in the Cartesian school by Geulincx, Spinoza, Malebranche and Leibniz, mainly in order to avoid certain difficulties in the Cartesian view. (Like epiphenomenalism, it robs consciousness of any biological function.)

(4) An assertion that mental states can interact with physical states. This was the view of Descartes which, it is widely believed, was superseded by (3).

My own position is that a brain–mind parallelism is almost bound to exist *up to a point*. Certain reflexes, such as blinking when seeing a suddenly approaching object, are to all appearances of a more or less parallelistic character: the muscular reaction (in which no doubt the central nervous system is involved) repeats itself with regularity when the visual impression is repeated. If our attention is drawn to it we may be conscious of its happening, and so with some (but of course not all) other reflexes.

Nevertheless, I believe that the thesis of a *complete* psychophysical parallelism—position (3)—is a mistake, probably even in some cases where mere reflexes are involved. *I thus propose a form of psychophysical interactionism.* This involves (as was seen by Descartes) *the thesis that the physical* *world 1 is not causally closed*, but open to the world 2 of mental states and events; a somewhat unattractive thesis for the physicist, but I think one that is supported by the fact that world 3 (including its autonomous regions) acts upon world 1 *via* world 2.

I am quite willing to accept the view that whenever anything goes on in world 2, something connected with it goes on in world 1 (in the brain). But in order to speak of a complete or thoroughgoing parallelism, we would have to be able to assert that 'the same' mental state or event is always accompanied by an exactly corresponding physiological state, and *vice versa*.

As indicated, I am prepared to admit that there is something correct in this assertion, and that for example the electrical stimulation of certain brain regions may regularly give rise to certain characteristic movements or sensations. But I ask whether, as a universal rule about all mental states, the assertion has any content; whether it is not an empty assertion. For we can have a parallelism between world 2 elements and brain processes, or between world 2 *Gestalten* and brain processes, but we can hardly speak of a parallelism between a highly complex, unique and unanalysable world 2 process and some brain process. And there are many world 2 events in our lives which are unique. Even if we disregard creative novelty, hearing a melody twice and recognising that it is the same melody is not a repetition of the same world 2 event, just because the second hearing of the melody is connected with an act of *recognising* the melody, which was absent the first time. It is the world 1 object (in this case the melody) which is repeated, but not the world 2 event. Only if we could accept a kind of world 2 theory which, like associationist psychology, looks upon world 2 events as composed of atom-like elements could we make a clear distinction between the repeated part of the world 2 experience—the *hearing* of the same melody—and the non-repeated part, the *recognition* that it is the same melody (where the recognition experience in its turn is capable of recurrence in other contexts). But I think that it is clear that such an atomistic or analytical psychology is quite incapable of carrying us far.

World 2 is highly complex. While if we attend only to such fields as sense perception (that is, perception of world 1 objects) we may think that we can analyse world 2 by atomic or molecular methods, for instance *Gestalt* methods (methods which, I think, are all unrewarding as compared with the biological or functional methods of Egon Brunswik or Richard Gregory), the application of such methods turns out to be quite inadequate if we consider our unique attempts to invent, and to understand, a world 3 object, such as a problem or a theory.

The way in which our thinking or our understanding interacts with attempts at linguistic formulation and is influenced by it; the way in which we have first a vague feeling for a problem or a theory which becomes clearer when we try to formulate it, and still clearer when we write it down and criticise our attempts to solve it; the way in which a problem may change and still be in a sense the old problem; the way in which a train of thought is on the one hand interconnected and on the other hand, articulated: all this seems to me to be beyond analytical or atomistic methods, including the interesting molecular methods of *Gestalt* psychology. There is a unique history of unique world 2 events involved in all of these attempts, and as a consequence, the talk about (strictly) *parallel* physiological processes loses all content.

Besides, we have reason to believe that often, if one region of the brain is destroyed, another region can 'take over', with very little or perhaps no interference with world 2—another argument against parallelism, and this time based on experiments in world 1 rather than on the necessarily vague consideration of the more complex world 2 experiences.

All this sounds, of course, very antireductionist; and as a philosopher who looks at this world of ours, with us in it, I indeed despair of any ultimate reduction. But as a methodologist this does not lead me to an antireductionist research programme. It only leads to the prediction that with the growth of our attempted reductions, our knowledge, and our universe of unsolved problems, will expand.

XXII

Let us return now to the problem of the specifically human consciousness of self; my suggestion was that it emerges in interaction (feedback, if you like) between world 2 and the worlds 1 and 3. My arguments for the role played by world 3 are as follows.

The human consciousness of self is based, among other things, upon a number of highly abstract *theories*. Animals and even plants have, no doubt, a sense of time, and temporal expectations. But it needs an almost explicit *theory* of time (*pace* Benjamin Lee Whorf) to look upon oneself as possessing a past, a present and a future; as having a personal history; and as being aware of one's personal identity (linked to the identity of one's body) throughout this history. Thus it is a *theory* that, during the period of sleep, when we lose the continuity of consciousness, we—our bodies—remain essentially the same; and it is on the basis of this theory that we can consciously recall past events (instead of merely being influenced by them in our expectations and reactions which, I suggest, is the more primitive form which the memory of animals takes).

Some animals, no doubt, have personalities; they have something closely analogous to pride and ambition, and they learn to respond to a name. But the human consciousness of self is anchored in language and (both explicitly and implicitly) in formulated theories. A child learns to use his name of himself, and ultimately a word like 'ego' or 'I', and he learns to use it with the consciousness of the continuity of his body, and of himself; he also combines it with the knowledge that consciousness is not always unbroken. The great complexity and nonsubstantial character of the human soul, or the human self, become particularly clear if we remember that there are cases where men have forgotten who they are; they have forgotten part or the whole of their past history, but they have retained, or perhaps recovered, at least part of their selfhood. In a sense, their memory has not been lost, for they *remember how* to walk, to eat, and even to speak. But they do not *remember that* they come from, say, Bristol, or what their names and addresses are. In so far as they do not find their way home (which animals normally do) their consciousness of self is affected even beyond the normal level of animal memory. But if they have not lost the power of speech, some human consciousness is left that goes beyond animal memory.

I am not a great friend of psychoanalysis, but its findings seem to support the view of the complexity of the human self, in contrast to any Cartesian appeal to a thinking substance. My main point is that the consciousness of the human self involves, at the very least, an awareness of the (highly theoretical) temporal or historical continuity of one's body; an awareness of the connection between one's conscious memory and the single, unique body which is one's own; and the consciousness of the normal and periodical interruption of one's consciousness by sleep (which, again, involves a theory of time and temporal periodicity). Moreover, it involves the consciousness of belonging locally and socially to a certain place and circle of people. No doubt much of this has an instinctive basis and is shared by animals. My thesis is that in raising it even to the level of unspoken human consciousness, human language or interaction between worlds 2 and 3 plays an important role.

It is clear that the unity of the human self is largely due to memory, and that memory can be ascribed not only to animals but also to plants (and even perhaps, in some sense, to non-organic structures such as magnets). It is therefore most important to see that the appeal to memory as such is not enough to explain the unity of the human self. What is needed is not so much the 'ordinary' memory (of past events), but a memory of theories that link the consciousness of having a body to world 3 theories about bodies (that is, to physics); a memory which is of the character of a 'grasp' of world 3 theories. It comprises the dispositions which enable us to fall back on explicit world 3 theories if we need to, with the feeling that we possess such dispositions and that we can make use of them in order to articulate those theories if we need to. (This would, of course, explain to a certain extent the difference between the human consciousness of self with its dependence on human language, and animal consciousness.)

XXIII

These facts seem to me to establish the impossibility of any reduction of the human world 2, the world of human consciousness, to the human world 1, that is, essentially, to brain physiology. For world 3 is, at least in part, autonomous of the two other worlds. If the autonomous part of world 3 can interact with world 2, then world 2, or so it seems to me, cannot be reducible to world 1.

My standard examples of the partial autonomy of world 3 are taken from arithmetic.

I suggest that the infinite series of natural numbers is an invention, a product, of the human mind, and a part of developed human language. (There are, it appears, primitive languages in which one can count only 'one, two, many' and others in which one can count only to 'five'.) But once a method of counting without end has been invented, distinctions and problems arise autonomously: even and odd numbers are not *invented* but *discovered* in the series of natural numbers, and so are prime numbers, and the many solved and unsolved problems connected with them.

These problems, and the theorems which solve them (such as Euclid's theorem that there does not exist a greatest prime) arise autonomously; they arise as part of the internal structure of the man-created series of natural numbers, and independently of what we think or fail to think. But we can *grasp* or *understand* or *discover* these problems, and solve some of them. Thus our thinking, which belongs to world 2, depends in part on the autonomous problems and on the objective truth of theorems which belong to world 3: world 2 not only creates world 3, it is partly created by world 3 in a kind of feedback process.

My argument now runs as follows: world 3, and especially its autonomous part, are clearly irreducible to the physical world 1. But since world 2 depends, in part, upon world 3, it is also irreducible to world 1.

Physicalists, or philosophical reductionists as I called them (1970), are thus reduced to denying the existence of worlds 2 and 3. But with this, the whole of human technology (especially the existence of computers), which makes so much use of world 3 theorems, becomes incomprehensible; and we must assume that such violent changes in world 1 as are produced by the builders of airports or skyscrapers are ultimately produced, without the invention of world 3 theories or world 2 plans based on them, by the physical world 1 itself: they are predestined; they are part of a preestablished harmony built, ultimately, into hydrogen nuclei.

These results seem to me absurd; and philosophical behaviourism or physicalism (or the philosophy of the identity of mind and body) appears to me to be reduced to this absurdity. It seems to me to stray too far from common sense.

XXIV

Philosophical reductionism is, I believe, a mistake. It is due to the wish to reduce everything to an ultimate explanation in terms of essences and substances, that is, to an explanation which is neither capable of, nor in need of, any further explanation.

Once we give up the theory of ultimate explanation we realise that we can always continue to ask: 'Why?'. Why-questions never lead to an ultimate answer. Intelligent children seem to know this, but give way to the adults who, indeed, cannot possibly have time enough to answer what is in principle an endless series of questions.

XXV

The worlds 1, 2 and 3, though partly autonomous, belong to the same universe: they interact. But it can easily be shown that knowledge of the universe, if this knowledge itself forms part of the universe, as it does, must be incompletable.

Take a man who draws a detailed map of the room in which he is working. Let him try to include in his drawing the map which he is drawing. It is clear that he cannot complete the task, which includes an infinity of smaller and smaller maps within each map: every time he adds a new line to the map, he creates a new object to be drawn, but not yet drawn. The map which is supposed to contain a map of itself is incompletable.

The story of the map shows the incompleteness and openness of a universe that contains world 3 objects of knowledge. Incidentally, it can also be used as an argument for the view that our universe is indeterministic. For while, admittedly, each of the different 'last' strokes actually entered into the map determines, within the infinite sequences of maps to be drawn, a dependent stroke, the determinacy of the strokes holds only if we do not consider the fallibility of all human knowledge (a fallibility which plays a considerable role in the problems, theories and mistakes of world 3). Taking this into account, each of these 'last' strokes entered into the map constitutes a *problem* for the draughtsman, a problem of entering a further stroke which depicts the last stroke *precisely*. Because of the fallibility that characterises all human knowledge, this problem cannot possibly be solved by the draughtsman with absolute precision; and the smaller the strokes to which the draughtsman proceeds, the greater will be the relative imprecision, which in principle will be unpredictable and indeterminate and will constantly increase. In this way, the story of the map shows how the fallibility which affects objective human knowledge contributes also to the essential indeterminism of our universe, apart from showing the openness and unknowability of a universe that contains human knowledge as a part of itself.

This example can help us to see why all explanatory science is incompletable; for to be complete it would have to give an explanatory account of itself.

An even stronger result is implicit in Gödel's famous theorem of the incompletability of formalised arithmetics (though it has to be admitted that to use Gödel's theorem and other metamathematical incompleteness theorems in this context is to use heavy armament against a comparatively weak position). Since all physical science uses arithmetic (and since for a reductionist only science formulated in physical symbols has any reality), Gödel's incompleteness theorem renders all physical science incomplete; which to the reductionist should show that all science is incomplete. For the nonreductionist, who does not believe in the reducibility of all science to physically formulated science, science is incomplete anyway.

Not only is philosophical reductionism a mistake, but the belief that the method of reduction can achieve complete reductions is, it seems, mistaken too. We live, it appears, in a world of emergent evolution; of problems whose solutions, if they are solved, beget new and deeper problems. Thus we live in a universe of emergent novelty; of a novelty which, as a rule, is not completely reducible to any of the preceding stages.

Nevertheless, the method of attempting reductions is most fruitful, not only because we learn a great deal by its partial successes, by partial reductions, but also because we learn from our partial failures, from the new problems which our failures reveal. Open problems are almost as interesting as their solutions; indeed they would be just as interesting but for the fact that almost every solution opens up in its turn a whole new world of open problems.

References

Anderson, C. D. (1933). Cosmic ray bursts. *Physical Review*, **43**, 368–9.

Anderson, C. D. (1933). The positive electron. *Physical Review*, **43**, 491–4.

Chadwick, J. (1932). Possible existence of a neutron. *Nature*, **129**, 132.

Eccles, J. C. (1970). *Facing Reality*. Springer-Verlag, Berlin, Heidelberg, New York.

Eccles, J. C. (1972). Unconscious actions emanating from the human cerebral cortex (unpublished).

Eddington, A. (1936). *Relativity Theory of Protons and Electrons*. Cambridge University Press.

Einstein, A. (1920). *Äther und Relativitätstheorie*. Springer, Berlin. (Translated as (Einstein, 1922).)

Einstein, A. (1922). *Sidelights on Relativity*. Methuen, London.

Heitler, W. and London, F. (1927). Wechselwirkung neutraler Atome und homöopolare Bindung nach der Quantenmechanik. *Zeitschrift für Physik*, **44**, 455–72.

Kuratowski, C. (1920). Sur la notion de l'ordre dans la théorie des ensembles. *Fundamenta mathematica*, **2**, 154–66.

Meyerson, É. (1908). *Identité et Réalité*. F. Alcan, Paris. (Translated as (Meyerson, 1930).)

Meyerson, É. (1930). *Identity and Reality*. Allen and Unwin, London.

Millikan, R. A. (1932). *Time, Matter and Values*. University of North Carolina Press, Chapel Hill.

Monod, J. (1970). *Le Hasard et la Nécessité*. Editions du Seuil, Paris. (Translated as (Monod, 1971).)

Monod, J. (1971). *Chance and Necessity*. Knopf, New York.

Pauling, L. (1959). [Discussion remark.] In *The Origin of Life on the Earth* (proceedings of the First International Symposium on The Origin of Life on the Earth, Moscow, 19–24 August 1957, ed. A. I. Oparin and others). Ed. F. Clark and R. L. M. Synge. Pergamon Press, London, 119.

Popper, K. R. (1945). *The Open Society and its Enemies*, 2 volumes. Routledge and Kegan Paul, London. (See now (Popper, 1966*b*).

Popper, K. R. (1949). Towards a rational theory of tradition. *The Rationalist Annual*, London, 36ff. (Now ch. 4 of (Popper, 1963).)

Popper, K. R. (1950). *The Open Society and its Enemies*, revised edition. Princeton University Press. (See now (Popper, 1966*b*).)

Popper, K. R. (1953). Language and the body-mind problem. In *Proceedings of the XIth International Congress of Philosophy*, 7, 101–7. (Now ch. 12 of (Popper, 1963).)

Popper, K. R. (1956). Three views concerning human knowledge. In *Contemporary British Philosophy* (ed. H. D. Lewis). Allen and Unwin, London, 355–88. (Now ch. 3 of (Popper, 1963).)

Popper, K. R. (1957). The aim of science. *Ratio* [Oxford], 1, 24–35. (Now ch. 5 of (Popper, 1972).)

Popper, K. R. (1961). Evolution and the tree of knowledge (Herbert Spencer Lecture). Published for the first time as ch. 7 of (Popper, 1972).

Popper, K. R. (1963). *Conjectures and Refutations*. Routledge and Kegan Paul, London; Basic Books, New York. (See now also the third and fourth editions (Popper, 1969, 1972).)

Popper, K. R. (1966*a*). Of clouds and clocks (the second Compton Memorial Lecture, delivered 1965). Washington University Press, St Louis, Missouri. (Now ch. 6 of (Popper, 1972).)

Popper, K. R. (1966*b*). *The Open Society and its Enemies*, fifth edition, two volumes. Routledge and Kegan Paul, London.

Popper, K. R. (1967). Quantum mechanics without 'the observer'. In *Quantum Theory and Reality* (*Studies in the Foundations, Methodology, and Philosophy of Science*, volume 2) (ed. Mario Bunge). Springer-Verlag, Berlin, Heidelberg, New York, 7–44.

Popper, K. R. (1969). *Conjectures and Refutations*, third (revised) edition, Routledge and Kegan Paul, London. (Fourth edition, 1972.)

Popper, K. R. (1970). A realist view of logic, physics, and history. In *Physics, Logic, and History* (ed. W. Yourgrau and A. D. Breck). Plenum Press, New York and London, 1–30. (Now ch. 8 of (Popper, 1972).)

Popper, K. R. (1972). *Objective Knowledge: an evolutionary approach*. Clarendon Press, Oxford.

Ross, A. (1972). The rise and fall of the doctrine of performatives. In *Contemporary Philosophy in Scandinavia* (ed. R. E. Olsen and A. M. Paul). Johns Hopkins Press, Baltimore, 197–212.

Sperry, R. W. (1964). The great cerebral commissure. *Scientific American*, **210**, 42–52.

Wiener, N. (1914). A simplification of the logic of relations. *Proceedings of the Cambridge Philosophical Society*, 17, 387–90.

Yukawa, H. (1935). On the interaction of elementary particles, I. *Proceedings of the Physico-Mathematical Society of Japan*, 3rd series, **17**, 48–57.

Postscript

Except for minor revisions and a reference or two to this postscript, I have left the paper as originally prepared. But before it was discussed at the conference, I criticised it there myself, pointing out the first two of the following four important omissions.

(1) The first of these omissions is that in the paper there is no mention of the attempts to reduce thermodynamics to mechanics. This is an important example of a reduction, and an interesting one from the point of view of my thesis. For while the results of the attempted reduction have been important, there has not been anything like a complete reduction without remainder.

(2) There is a second and more important omission—a point which in the paper I took more or less for granted (I mentioned it only briefly in Section XIV; see text to footnote 5, above). It is this. Before we can even attempt a reduction, we need as great and as detailed a knowledge as possible of whatever it may be that we are trying to reduce. Thus before we can attempt

a reduction, we need to work on the level of the thing to be reduced (that is, the level of 'wholes'). I had pointed this out previously.[6]

(3) A third omission (not mentioned at the conference) is connected with the distinction (indicated at the beginning of the paper: see text to footnote 1) between a *reduction* which explains some theory by an existing theory and an *explanation with the help of a new theory*: though I will not quarrel about words I should now be disinclined to call an explanation with the help of a new theory a 'reduction'. Yet if this terminology is adopted the explanation of the wave theory of the propagation of light by Maxwell's theory of electromagnetism could be claimed as an example of a completely successful reduction (perhaps the only example of a completely successful reduction). However, it may be better not to describe this as a reduction of one theory to another, or one part of physics to another, but rather as a radically new theory which succeeded in unifying two parts of physics.

(4) Without wishing to advocate what one might call an antireductionistic research programme for biology, the following seems to be a reasonable comment on the situation.

The Newtonian mechanistic programme for physics broke down over the attempt to include electricity and magnetism, or, more precisely, over Faraday's introduction of non-central forces. (Maxwell's attempt to reduce these non-central forces to Newtonian theory by constructing a mechanical model of the ether proved extremely fruitful in suggesting to him his field equations, but nevertheless was unsuccessful and had to be dropped.) Einstein's realisation that Newton's and Maxwell's theories are incompatible led to Special Relativity. So physicists had to accept a radically new theory rather than a reduction. A similar fate befell physics when both mechanics and electromagnetic theory in the unified form due to Lorentz and Einstein were applied to new and largely statistical problems of the microstructure of matter. This led to quantum mechanics. We cannot rule out the possibility that the inclusion of biological problems may lead to a further expansion and revision of physics.

Discussion

Rensch
I still have some difficulties in understanding your world 3. In my opinion the human culture is the practical effect of your world 2, that is to say, of our mental abilities, or capability to use a language, to think in generalised symbols and to draw conclusions in conformity with universal logical laws. Being an identist I would regard your worlds 1, 2 and 3 as three evolutionary levels. If your world 3 is also characterised by the fact that culture originated through the influence of or the adaptation to causal and logical universal

[6] See my (1972), 285–318, esp. 297.

laws, then it would contain also elements of world 1. I put this question because you mentioned in your paper that you did not invent your classification but only discovered its existence.

Popper

What Professor Rensch calls 'your world 3' is indeed, as I always emphasise, the product of world 2. This agrees with Professor Rensch's remark that it is 'the practical effect' of world 2. I also think, like Professor Rensch, that we can 'regard . . . worlds 1, 2 and 3 as three evolutionary levels'. But all this does not mean that world 3 is part of world 2, or world 2 part of world 1, or anything of the sort. If I have a very bad toothache then, no doubt, it may be regarded as the product of my very bad tooth. Even though I feel the pain 'in' the tooth, few of us doubt that a nervous signal has to be transmitted from the tooth to the brain before I can feel it. Similarly, a book which I have written is the product of my thought processes, of my world 2; but the book, the product of my world 2, can be bought and read, while no part of my thought processes, of my world 2, can be bought and read. Thus the products of world 2 do not necessarily belong to world 2.

Look at the sentence 'This book is for sale in the bookshops, and it can be read'. Here the word 'book' is used for a world 1 object, located in physical space, and possessing mass and weight; but it is here also used for a world 3 object. For reading a book involves more than merely observing black marks on white paper: it involves the grasping (a world 2 process) of an *objective thought content*, a typical world 3 object.

Only thought contents can stand in the objective logical relationship of compatibility or incompatibility, or in the relationship of premise and conclusion. We can distinguish and, I suggest, we ought to distinguish, between Euclid's *discovery*—a world 2 object—of the theorem that there is no greatest prime number, and *that theorem itself*, which is a world 3 object.

This distinction (I do not think I spoke of classification) is very clearly made in the works of Bolzano and Frege. However, neither Bolzano nor Frege attributed to what I call world 3 objects the power of interacting with world 1 (*via* world 2).