Chapter Fourteen

Aim-Oriented Empiricism since 1984

Since the publication of the first edition of this book I have made a number of important improvements to aim-oriented empiricism (AOE), and developed further arguments intended to show that the doctrine solves fundamental philosophical problems about science – such as the problems of simplicity and induction – which standard empiricism cannot solve.

1 Improved Versions of Aim-Oriented Empiricism (AOE)

In the first edition – in a deliberately simplified way – AOE represented knowledge in physics (and thus in natural science to some extent) at five levels. These are: (1) the thesis that the universe is comprehensible in some way (physicalism being a special case), (2) physicalism (the thesis that the universe is physically comprehensible), (3) best available metaphysical blueprint, (4) fundamental physical theory, and (5) empirical phenomena.

There is an obvious objection to the doctrine formulated like this. What if thesis (1) is false, and the universe is only imperfectly or partially comprehensible (in some way or other)? In chapter 9, I considered this possibility, and argued that science should reject it. I argued that science is justified in accepting that the universe is perfectly physically comprehensible because, if it is only imperfectly or approximately physically comprehensible, the best way we can acquire knowledge of this is to assume perfect physical comprehensibility, search for it and fail in the attempt. Even in an imperfectly or approximately physically comprehensible universe, in other words, the assumption of perfect comprehensibility is the most fruitful, heuristically and methodologically, to make. But I now think this need not be the case. The universe might be so constituted that infinitely many theoretical revolutions are required before we can arrive at the true physical theory of everything, it being the case, however, that after each revolution one more force needs to be postulated. In such an ultimately incomprehensible universe, science might still be possible. Furthermore, we might well, after two or three revolutions, cotton on to the point that the universe is such that the number of forces go up after each revolution. In such a universe, this would be a more fruitful assumption to make than that of perfect physical comprehensibility. Considerations such as this add support to the point that AOE needs to take into account the possibility that the universe is not perfectly comprehensible.

And should we not, perhaps, consider much more extravagant possibilities? For all that we know for certain, ultimate reality may be very different from the way it is depicted by modern theoretical physics. Perhaps it is not physical at all. Perhaps physics is a sort of temporary illusion, and some time in the future physics will cease to apply to phenomena and we will find ourselves in a strange new world. Perhaps the universe is not even partially comprehensible. It might still be possible to live, and to acquire knowledge, even though of a kind very different from current scientific knowledge, nothing like modern science being possible in that its basic presupposition, imperfect or approximate physicalism, is false, and phenomena do not even occur as if it were true. Even in such bizarre and no doubt (in some sense) wildly improbably circumstances, there is one assumption about the world we would be entitled to make, namely that it is such that we can acquire some knowledge about our local environment sufficient to make life possible. If this assumption is false, we have had it, whatever we assume. In making this assumption we cannot, in any circumstances, endanger the pursuit of knowledge, and we may, quite possibly, aid it. This assumption of the *partial knowability* of the universe thus deserves to be a permanent item of scientific knowledge which is accepted, not because we have good grounds for holding it to be true, but because accepting it cannot obstruct, and can only aid, the pursuit of knowledge.

I have subsequently developed a number of more elaborate versions of AOE to take these possibilities into account. The first modification, the most elaborate, was spelled out in Maxwell (1998). This version of AOE has *ten* levels. The specific details of this version of AOE, low down in the hierarchy, might need to be modified if, for example, one is considering the history of science before Galileo, or if the future produces such dramatic changes in our conception of the universe that theses, low down in the hierarchy, need to be changed. In order to take such eventualities into account, I put forward *generalized aim-oriented empiricism* (GAOE): see Maxwell (1998, p. 101). GAOE holds that some kind of hierarchical view needs to be adopted, the top one, or more, levels being those of the ten-level version of AOE, but other, lower levels possibly being different. Subsequently I decided that the ten-level version of AOE was too elaborate, and I reduced the number of levels to *seven*: see Maxwell (2004b; 2005b). I have also suggested an alternative to physicalism – a sort of cosmological physicalism: see Maxwell (2004b, appendix, section 5, pp. 198-205). This amounts to a modification of the seven-level version of AOE. Then I complicated the picture again, and developed a version of AOE which takes, as the hierarchy of metaphysical theses, different versions of physicalism: see Maxwell (2004a). This version of AOE, which presupposes that the universe is *physical* in character, can be embedded in the earlier, more accommodating versions of AOE which allow for the possibility that the universe might turn out to be non-physical. I have also spelled out how endlessly many much more restrictive versions of AOE can be developed which are applicable to different scientific specialities, each with its own restrictive presuppositions: see Maxwell (2004b, pp. 41-47). This is discussed below in connection with the problem of induction.

What is one to make of these different versions of AOE? I am inclined to think, now, that which one you choose to adopt depends on what your purpose is. If you want to solve the problem of induction, it may be necessary to consider the ten-level version. If you are exclusively interested in theoretical physics, and you are happy to assume that *some* version of physicalism is true, the version of AOE expounded in Maxwell (2004a) may suffice. Philosophers, anthropologists and others exploring wild cosmological possibilities might find it useful to do so within the framework of GAOE. Those concerned with specific scientific specialities – whether scientists, or historians or philosophers of science, will need to consider an appropriate specific version of AOE. These diverse versions of AOE are not *rivals*: they are more or less detailed exemplifications of a single basic idea.

In what follows, I first expound the ten-level version of AOE, and then, briefly, the seven-level version. Then, in the next section I tackle the fundamental problem of what the *unity* of a physical theory is. The solution to this problem provides us with eight distinct versions of physicalism. In the section after, I expound that version of AOE which exploits these eight versions of physicalism. I then tackle the problem of verisimilitude and, to conclude this chapter, I argue that AOE solves the problem of induction.

The basic idea behind the ten level version of AOE – see figure 11 – and the other versions, can be put like this. For science to proceed, and for the enterprise of acquiring knowledge to proceed more generally, an untestable, metaphysical assumption must be made about the nature of the universe. In order to give ourselves the best chance of achieving success we need to make an assumption that is fruitful and true, but it is more than likely that the assumption we make will be false. Granted this, in order to give ourselves the best hope of making progress in acquiring knowledge, we need to make, not just *one*, but a *hierarchy* of assumptions, these assumptions becoming increasingly insubstantial, and so increasingly likely to be true, as we ascend the hierarchy. We make those assumptions which seem to be implicit in our apparently most successful ventures at improving knowledge, and which seem to be inherently fruitful for improving knowledge, if true. The hierarchy, initially, simply makes explicit what is implicit in what seem to be our most successful efforts at acquiring knowledge. We then *revise* metaphysical assumptions, and associated methodological rules, in the light of which seem to lead to the most empirically successful research programmes, but in such a way that we keep such revisions as low down in the hierarchy of assumptions as possible. Only when efforts at acquiring knowledge seem to be meeting with little success do we actively consider more radical revisions higher up the hierarchy. We assume, quite generally, that the top level 10 assumption, of figure 11, is true, and the bottom level 3 assumption is false. As we descend from 10 to 3, at some point we move from truth to falsity, and thus to an assumption which needs to be revised. Our hope is that as we proceed, and learn more about the nature of the universe, we progressively bring truth lower and lower down in the hierarchy. As our knowledge improves, assumptions and associated methods improve as well. There is positive feedback between improving knowledge and improving assumptions and methods – that is, knowledge-about-how-to-improve-knowledge. This positive feedback between improving knowledge, and improving knowledge-about-how-to-improve-knowledge is the *sine qua non* of scientific methodology and rationality. As science improves its knowledge and understanding of nature, it adapts its own nature to what it has discovered. The astonishing progressive success of science in improving our knowledge and understanding of nature owes much to the exploitation of this positive feedback, meta-methodological feature of AOE in scientific practice. (Even though the scientific community has officially upheld standard empiricism, fortunately its allegiance to this doctrine has been sufficiently hypocritical to make it possible to implement something close to AOE in scientific practice. Paying lip service to standard empiricism has nevertheless been damaging; freeing science of this hypocrisy, so that AOE becomes the official philosophy of science, would have beneficial consequences: see Maxwell, 1998 and 2004b, for further details.)

If we can be reasonably confident that the best available thesis at level 3 is false, we can be even more confident that accepted fundamental physical theories, at level 2, are false, despite their immense empirical success. This confidence comes partly from the vast empirical content of these theories, and partly from the historical record. The greater the content of a proposition the more likely it is to be false; the fundamental theories of physics, general relativity and the standard model have such vast empirical content that this in itself almost guarantees falsity. And the historical record backs this up; Kepler's laws of planetary motion, and Galileo's laws of terrestrial motion are corrected by Newtonian theory, which is in turn corrected by special and general relativity; classical physics is corrected by quantum theory, in turn corrected by relativistic quantum theory, quantum field theory and the standard model. Each new theory in physics reveals that predecessors are false. Indeed, if the level 4 assumption of AOE is correct, then all current fundamental physical theories are false, since this assumption asserts that the true physical theory of everything is unified, and the totality of current fundamental physical theory, general relativity plus the standard model, is notoriously disunified. AOE actually predicts that accepted fundamental physical theory, that is not both unified and (in principle) applicable to all phenomena, is false, whatever empirical success it may have.

Figure 11: Aim-Oriented Empiricism (AOE)

In more detail, the ten-level version of AOE amounts to the following[1]: see figure 11.
Level 1: P_1. Empirical data (low level observational and experimental laws).[2]

---

[1] The following scheme deliberately ignores vast tracts of scientific knowledge, such as: all of phenomenological physics, including such areas as solid state physics, thermodynamics and statistical mechanics; observational science carried on for its own sake, in astronomy, geology and elsewhere, and not in order to test fundamental physical theories; chemistry; biology; all of social science. For a justification of this neglect here, see remarks above and below, and Maxwell (1998, chapter 2, section 5). For my views about biology and social inquiry, all that which physics seems to miss out, see earlier chapters of the present work, and Maxwell (2001 and 2004b).

[2] These are in the form of laws appropriately restricted in terms of range of application and accuracy, so as to stand a good chance of being true, and of being derivable, in principle, from appropriate theory.

Level 2: P_2. All accepted fundamental dynamical theories, or accepted laws governing the way physical phenomena occur if no dynamical theory has been developed that applies to the phenomena in question. In terms of current scientific knowledge, this level consists of the so-called standard model (SM) – the quantum field theory of fundamental particles and the forces between them – plus general relativity (GR).

Level 3: P_3. *Best Blueprint*. The best available more or less specific metaphysical view as to how the universe is physically comprehensible, a view which asserts that everything is composed of some more or less specific kind of physical entity, all change and diversity being, in principle, explicable in terms of this kind of entity. Examples, taken from the history of physics are: the corpuscular hypothesis of the 17th century, according to which the universe consists of minute, infinitely rigid corpuscles that interact only by contact; the view, associated with Newton and Boscovich, according to which the universe consists of point-atoms that possess mass and interact at a distance by means of rigid, spherically symmetrical, centrally directed forces; the unified field view, associated with Faraday and Einstein, according to which everything is made up of one self-interacting field, particles of matter being especially intense regions of the field. Some might argue that the best available blueprint available today is the basic metaphysical idea of superstring theory, or M-theory as it is now called: the universe consists of minute quantum strings that move in 10 or 11 dimensions of space-time, all but four of which are curled up into a minute size, thus escaping detection. In Maxwell (1998, chapter 3) I argue, however, that the best available blueprint is a somewhat more general thesis that I call *Lagrangianism*.

Level 4: P_4. *Physical Comprehensibility*. The more imprecise thesis that the universe is physically comprehensible in some way or other, everything being made up of just one kind of physical entity (or perhaps just one entity), all change and diversity being in principle explicable in terms of this one kind of entity. This thesis asserts that the universe is such that some as-yet-to-be-discovered unified physical 'theory of everything' (in the current jargon of theoretical physicists) is true. This is the thesis I have been calling 'physicalism'.

As we have seen, there are a number of ways in which the universe might be comprehensible even though physicalism is false. It might be that God exists, all natural phenomena being explicable in terms of the will of God. It might be that a society of gods exist, natural phenomena being the outcome of (and being explicable in terms of) the diverse, and sometimes conflicting, desires of the gods. It might be that, even though there is no God, there is some sort of overall cosmic goal, everything being explicable in terms of this cosmic goal (being required to fulfil the goal). Or it might be that there is some kind of cosmic programme, somewhat like a computer programme, which determines how events unfold; in this case events would be explicable in terms of the basic cosmic programme.

These conflicting views as to how the universe is comprehensible, together with physicalism, despite their diversity, all have something in common. They all hold that the universe is such that there is *something* (kind of physical entity, God, tribe of gods, cosmic goal, cosmic programme or whatever) which does not itself change but which, in some sense, determines or is responsible for everything that does change (all change and diversity in the world in principle being explicable and understandable in terms of the underlying unchanging *something*). This is the thesis at the next level.

Level 5: P_5. *Comprehensibility*. The thesis (even more imprecise than physicalism) that the universe is comprehensible in some way or other, there being *something*, or an aspect of something (kind of physical entity, God, society of gods, cosmic purpose, cosmic programme or whatever) that runs through all phenomena, and in terms of which all phenomena can, in principle, be explained and understood. The thesis that the universe is comprehensible pushes to the limit the thesis that the universe is such that some phenomena can be explained and understood, to some extent at least: it asserts that the universe is such that *all* phenomena can, in principle, be fully explained and understood (insofar as this is logically possible), all phenomena being explicable in terms of the one, unchanging *something*, present everywhere, at all times and places, throughout all phenomena, in an invariant form.[3]

Level 6: P_6. *Near Comprehensibility*. The even more imprecise thesis that the universe is 'nearly comprehensible'. This means that the universe is sufficiently nearly comprehensible for the hypothesis that it is perfectly comprehensible to be more fruitful to adopt than any comparable assumption from the standpoint of the growth of knowledge.

Level 7: P_7. *Rough Comprehensibility*. The even more imprecise thesis that the universe is 'roughly comprehensible' in the sense that the universe is such that there is some assumption of approximate comprehensibility (including the possibility of perfect comprehensibility as a special case) which is the most fruitful rationally discoverable[4] assumption to adopt from the standpoint of the growth of knowledge.

Level 8: P_8. *Meta-Knowability*. The still more imprecise thesis that the universe is 'meta-knowable', which means that the universe is such that there is some rationally discoverable assumption about it which leads to improved methods for the improvement of knowledge.

Level 9: P_9. *Epistemological Non-Maliciousness*. The universe is such that it does not exhibit comprehensibility, meta-knowability, or even mere partial knowability more generally, in our immediate environment only. However drastically phenomena at other times and places may differ from local phenomena, nevertheless the general nature of all such phenomena is such that it can in principle be discovered by us by developing knowledge acquired in our immediate environment. If inexplicable, arbitrary phenomena occur (phenomena specifiable only by some grossly *ad hoc* theory of the kind indicated in footnote 4 above), their occurrence is discoverable by us in our immediate environment.

Level 10: P_{10}. *Partial Knowability*. The universe is such that we possess and can acquire some knowledge of our immediate environment as a basis for action.

Corresponding to each metaphysical assumption, at level r, where r runs from 3 to 9, there is a methodological rule (represented by sloping dotted lines in figure 11) which asserts: accept that level r-1 assumption (or collection of fundamental dynamical theories if r = 3) which best exemplifies the level r assumption, and which best promotes the growth of empirical knowledge (at levels 1 and 2), or at least holds out the greatest hope of doing this.

A few words of clarification concerning the principles at levels 3 to 10. They all assert, in different degrees, that the cosmos is more or less comprehensible or knowable. As we ascend, from level 3 to level 10, the theses become increasingly unspecific and contentless and thus, other things being equal, increasingly likely to be true. Theories at level 2 are burdened with massive precision and content; AOE predicts that, however empirically successful they may be, if, taken together, they clash with physicalism (as at present), then they are false. They are, in this case, fragmentary imperfect glimpses of an underlying unity. The best blueprint at level 3 is the best current attempt to do justice both (a) to theoretical knowledge at level 2, and (b) to physicalism at level 4. Ideally, it exemplifies physicalism in the sense that, what the blueprint postulates to exist

[3] In order to explain, in science, it is not sufficient to predict; it is necessary, in addition, to show that ostensibly diverse phenomena are diverse aspects of *one* phenomenon (or one kind of phenomenon), as when the diverse motions of terrestrial projectiles, the moon round the earth, the earth and other planets round the sun, double stars round each other, and stars round our galaxy are all aspects of the *one* kind of phenomenon of objects moving and interacting in accordance with Newton's laws of motion and law of gravity. The thesis that this is the proper way to understand scientific explanation will be developed in the next section: see also Maxwell (1998, chapter 4, and 2004b).

[4] The notion of 'rationally discoverable' is problematic. As I am using the phrase, no thesis about the universe is rationally discoverable if it is grossly *ad hoc*, and the *ad hoc* phenomena, postulated by the thesis, lie beyond our experience. (A thesis is grossly *ad hoc* if it is like the theories discussed in chapter 9 in that part devoted to the refutation of standard empiricism – points 2 and 3 – or like the most severely disunified theories considered in the next section of the present chapter.) Any thesis *ad hoc* in this way is one of infinitely many rivals, all equally arbitrary, there being no rationale to prefer the given thesis.

that determines the way events occur must be (like what physicalism postulates) invariant throughout all phenomena. (If a blueprint is to exemplify physicalism perfectly, in other words, it must not add to physicalism in a patchwork way, for some, but not for all possible phenomena.) Level 3 blueprints have vastly less precision and content than current level 2 theory (SM plus GR); it is nevertheless reasonable to hold that all blueprints proposed so far are false, even if physicalism is true.

Each assumption, from level 3 to 6, asserts that the universe is comprehensible (to some degree at least), but with decreasing precision and content as we ascend from level 3 to 6. P_7, at level 7 asserts, still more modestly, that the universe is such that some assumption of partial comprehensibility is more fruitful than any rival, comparable assumption. It might be the case, for example, that the universe is such that there are *three* fundamental forces, theoretical revolutions involving the development of theories that progressively specify the nature of these three forces more and more precisely. In this case, the assumption that there are three distinct forces would be more helpful than that there is just *one* fundamental force (required if the universe is to be perfectly comprehensible physically). Alternatively, it might be the case that the universe is such that progress in theoretical physics requires there to be a series of theoretical revolutions, there being, after each revolution, one more force: in this case, the assumption that the universe is such that the number of distinct forces goes up by one after each revolution would be more helpful for the growth of knowledge than the assumption that there is just one fundamental force. P_8, even more modestly, asserts merely that the universe is such that existing methods for improving knowledge can be improved. These methods might involve consulting oracles, prophets or dreams; they need not involve developing explanatory theories and testing them against experience. P_9 asserts, still more modestly, that the universe is such that local knowledge can be developed so that it applies non-locally;[5] and P_{10} asserts, even more modestly, that the universe is such that some factual knowledge of our immediate environment exists and can be acquired.

It is important to appreciate that these assumptions are to be understood in such a way that they presuppose some existing body of empirical knowledge (at levels 1 and 2), and existing methods for improving knowledge implicit in current practice. What is being asserted is that the universe is comprehensible, or meta-knowable, to *us*, with our current factual knowledge and implicit methods for improving knowledge.

The logical relationship between the propositions at the various levels is as follows. Let us suppose, initially, that the universe really is physically comprehensible, and the true theory of everything, T, at level 2, has been discovered. In this case, ideally, P_2 would entail P_1, and P_r would entail P_{r+1} for r = 2,...8. (P_9 does not entail P_{10} as we shall see below.) For $2 \leq r \leq 8$, we may think of P_{r+1} as consisting of a statement of the form 'P_r or P_r* or P_r** or ...', where P_r*, P_r**, etc., are rival cosmological theses to P_r. In moving down from level r+1 to level r we adopt the factual conjecture that P_r*, P_r**, etc., are all false, and P_r is true.

For $5 \leq r \leq 8$, the above does not represent an idealization; in our present state of knowledge, P_r entails P_{r+1}. But for $r < 5$ the above is an idealization in many ways.

To begin with, even if we had discovered the true, unified theory of everything, T, this P_2 proposition at most entails P_1 propositions insofar as they are couched in the form: if such and such a state of affairs, S_1, exists at time t_1, then such and such a state of affairs, S_2, exists at time

---

[5] P_9 is a kind of 'principle of the uniformity of nature'. P_9 is, however, intended to be very much weaker than uniformity principles as these are usually formulated and understood. It does not assert that all phenomena are governed by the same laws everywhere, since the possibility of (some) arbitrary, '*ad hoc*' phenomena is conceded. Instead, P_9 asserts that if such phenomena occur anywhere they occur in our immediate environment. P_9 does not even assert that approximately lawful phenomena occur everywhere, but merely that whatever it is that makes our immediate environment partially knowable extends throughout the universe. We might live in a partially knowable world even though no laws strictly obtain, as the notion of law is understood in natural science.

t_2. If T is comprehensive and true then it entails all true conditional statements of this type. However, our ability to extract detailed implications from T is bound to be severely restricted: the equations of T are likely to be solvable only for a few, extremely simple states of affairs; they may, indeed, not be solvable precisely at all, it being necessary to use approximation methods to extract predictions from T. This may involve making dubious additional assumptions, or simplifying assumptions known to be false. As theoretical physics has advanced, from Newtonian theory to general relativity and the standard model, so equations have become immensely more difficult to solve; it is reasonable to suppose that this trend will continue into the future.

Granted that we have discovered T (the true, unified theory of everything) no problem should arise in connection with P_2 implying P_3, P_3 in turn implying P_4, and P_4 implying P_5. But of course we have not discovered T (and may never do so, physicalism, perhaps, being false). Instead, we have at present at least two very different, even clashing, fundamental physical theories – the so-called standard model (SM) and general relativity (GR). This means P_2 conflicts with P_3. Even taken individually, currently accepted theories belonging to P_2 may clash with P_3 (as when Newtonian theory clashes with the corpuscular blueprint, or Maxwellian electrodynamics clashes with the Boscovichean blueprint). Furthermore, in trying to formulate P_3 in such a way that it does as much justice as possible to the theories of P_2, P_3 may conflict with P_4.

Although a theory, T, at level 2, may clash with a blueprint, B, it may also be a B-type theory, in the sense that it is a more or less disunified exemplification of B. Thus B might assert that the universe is made up of one kind of point-particle that interacts by means of one kind of force, and a theory, T, might postulate 2 (or more) kinds of point-particle, with different masses, perhaps, or charges. In this case, even though T is incompatible with B, it is nevertheless a B-type theory, a more or less disunified exemplification of B. (Only theories which exemplify B perfectly imply B; theories which are more or less B-disunified are incompatible with B.) Analogous remarks concern the ways in which T may be related to physicalism, or B may be related to physicalism. In fact, quite generally, given theses P_r and P_s at levels r and s with $2 \le r < s \le 9$, P_r may be a more or less unified or adequate exemplification of P_s, even though P_r contradicts P_s.

An important non-empirical methodological rule of AOE asserts, in effect, that given two rival level r theses, P_r and Q_r, that one is to be preferred (other things being equal) which exemplifies the accepted level r+1 thesis, P_{r+1}, in the more unified, more adequate way.

The clashes (or disunities) between levels for $r < 5$, and clashes within levels, especially within P_2, serve to drive theoretical physics forward. These pose the problems that physicists try to solve. They are symptomatic of our ignorance. Progress in theoretical physics is to be assessed in terms of the extent to which a contribution promises to bring physics closer to the ideal state of affairs in which P_2 implies both P_1, and P_3 and P_4, P_2 being a candidate for the true, unified theory of everything.

In seeking to resolve clashes between levels, influences can go in both directions. Thus, given a clash between levels 1 and 2, this may lead to the modification, or replacement of the relevant theory at level 2; but, on the other hand, it may lead to the discovery that the relevant experimental result is not correct for any of a number of possible reasons, and needs to be modified. In general, however, such a clash leads to the rejection of the level 2 theory rather than the level 1 experimental result; the latter are held onto more firmly than the former, in part because experimental results have vastly less empirical content than theories, in part because of our confidence in the results of observation and direct experimental manipulation (especially after expert critical examination). Again, given a clash between levels 2 and 3, this may lead to the rejection of the relevant level 2 theory (because it is disunified, *ad hoc*, at odds with the current metaphysics of physics); but, on the other hand, it may lead to the rejection of the level 3 assumption and the adoption, instead, of a new assumption (as has happened a number of times in the history of physics, as we have seen). The rejection of the current level 3 assumption is likely to take place if the level 2 theory, which clashes with it, is highly successful empirically, and furthermore has the effect of increasing unity in the totality of fundamental physical theory overall, so that clashes between levels 2 and 4 are decreased. In general, however, clashes between levels 2 and 3 are resolved by the rejection or modification of theories at level 2 rather than the assumption at level 3, in part because of the vastly greater empirical content of level 2 theories, in part because of the empirical fruitfulness of the level 3 assumption (in the sense indicated above).

It is conceivable that the clash between level 2 theories and the level 4 assumption might lead to the revision of the latter rather than the former. This happened when Galileo rejected the then current level 4 assumption of Aristotelianism, and replaced it with the idea that 'the book of nature is written in the language of mathematics' (an early precursor of our current level 4 assumption of physicalism). The whole idea of AOE is, however that, as we go up the hierarchy of assumptions, we are increasingly unlikely to encounter error, and the need for revision. The higher up we go, the more firmly assumptions are upheld, the more resistance there is to modification.

It deserves to be noted that something like the hierarchy of metaphysical theses, constraining acceptance of physical theory from above, is to be found at the empirical level, constraining acceptance of theory from below. There are, at the lowest level, the results of experiments performed at specific times and places. Then, above these, there are low-level experimental laws, asserting that each experimental result is a repeatable effect. Next up, there are empirical laws such as Hooke's law, Ohm's law or the gas laws. Above these there are such physical laws as those of electrostatics or of thermodynamics. And above these there are theories which have been refuted, but which can be 'derived', when appropriate limits are taken, from accepted fundamental theory – as Newtonian theory can be 'derived' from general relativity. This empirical hierarchy, somewhat informal perhaps, exists in part for precisely the same epistemological and methodological reasons I have given for the hierarchical ordering of metaphysical theses: so that relatively contentless and secure theses (at the bottom of the hierarchy) may be distinguished from more contentful and insecure theses (further up the hierarchy) to facilitate pinpointing what needs to be revised, and how, should the need for revision arise. That such a hierarchy exists at the empirical level provides further support for my claim that we need to adopt such a hierarchy at the metaphysical level.

Having expounded and defended this ten level version of AOE in great detail (see Maxwell,

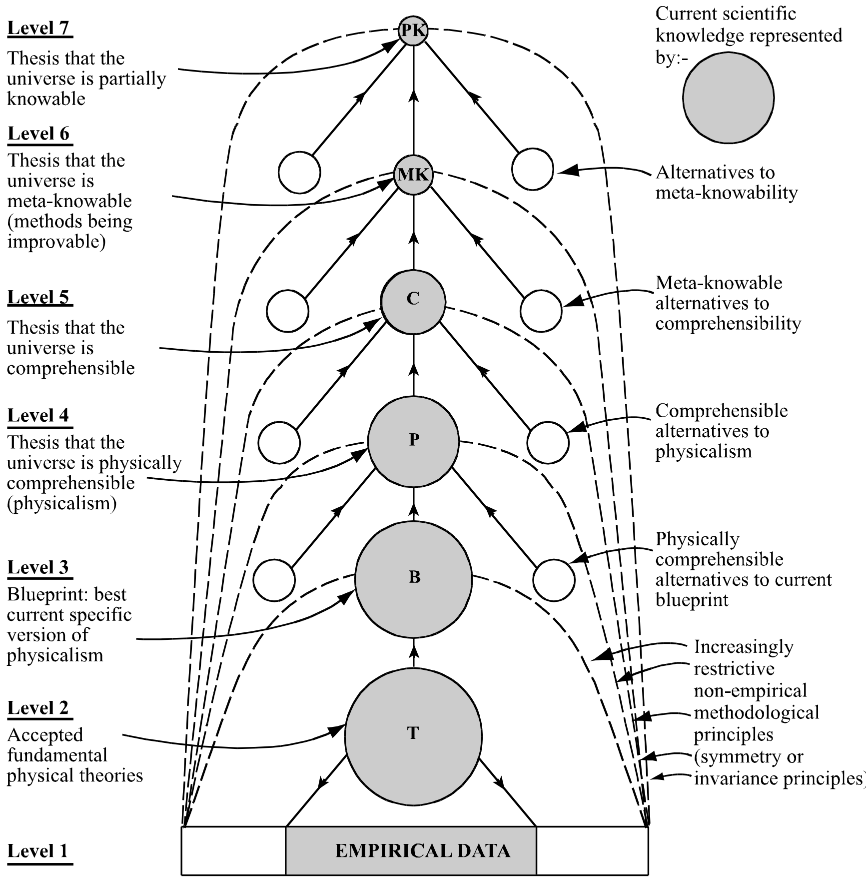

Figure 12: Aim-Oriented Empiricism

1998), I decided subsequently that *ten* levels are perhaps excessive, and I reduced them to *seven* (Maxwell, 2004b): see figure 12. I then complicated the picture again somewhat, by introducing additional levels that explicate the different meanings that may be assigned to physicalism, and to unity of theory. I will say something about this in the next but one section, after I have first discussed the problem of what the unity of a theory might be.

The extended version of AOE just indicated can be generalized, in line with the argument of chapter 5, to become an extended version of aim-oriented rationality. Not just in science, but in life too, aims can be profoundly problematic. Thus, generalizing from science, whenever the aim of any worthwhile endeavour is inherently problematic, it needs to be represented as a hierarchy of aims, these aims becoming less specific, more general, and thus less problematic as one goes up the hierarchy. In this way a framework of more or less unproblematic aims and associated methods is created within which much more specific and problematic aims and methods can be improved as the endeavour proceeds. Aim-oriented rationality, construed in this way, is especially relevant when there are conflicting aims and ideals: it enables those involved to distinguish agreement, high up in the hierarchy of aims, from disagreement, low down in the hierarchy, thus facilitating resolution of conflict. Aim-oriented rationality is no magic cure for conflict, but in facilitating conflict resolution, it could help promote the desire for it by demonstrating that it is feasible.

Figure 13 depicts what this extended version of aim-oriented rationality might look like when applied to the fundamental and profoundly problematic aim and endeavour of creating a wise, civilized world. Figure 13 is the outcome of generalizing figure 12, and then applying the result to the task of creating civilization.

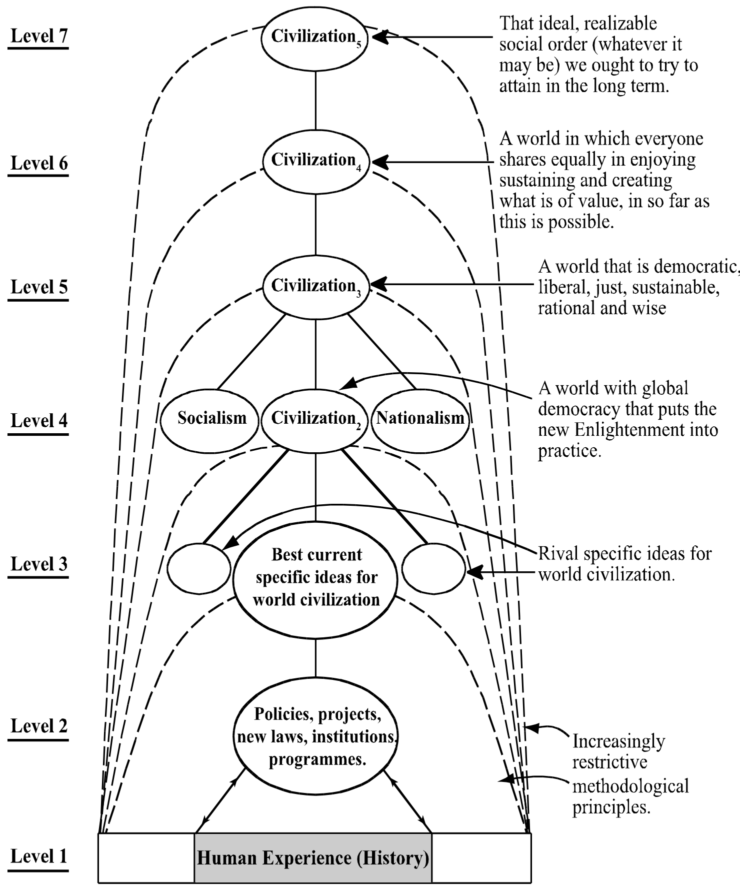

Figure 13: Implementing Aim-Oriented Rationality in Pursuit of Civilization

2 The Problems of Simplicity and Unity of Theory

A further improvement that I have made to AOE has to do with the simplicity, unity or explanatory character of theories. (In what follows I shall concentrate on 'unity'; I will then make a few remarks about 'simplicity'. The 'explanatory character' of a theory depends on just three things: unity, simplicity, and empirical content.)

Everyone recognizes that a theory, in order to be acceptable, must satisfy requirements of unity as well as requirements of empirical success. Horribly *ad hoc*, disunified, complex, aberrant theories, of the kind considered on pages 206-11 above, are just not considered in science, whatever their empirical success might be were they to be considered. But what is unity? Is there just one notion here, or several? How can one capture this notion of the unity of a theory when an apparently beautifully unified theory can always be reformulated so that it becomes horribly disunfied, and *vice versa*, a horribly disunified theory can be reformulated to become unified?[6]

---

[6] Richard Feynman has provided the following amusing illustration of this point (Feynman et al. 1965, 25-10 - 25-11). Consider an appallingly disunified, complex theory, made up of 10^{10} quite different, distinct laws, stuck arbitrarily together. Such a theory can easily be reformulated so that it reduces to the dazzlingly unified, simple form: A = 0. Suppose the 10^{10} distinct laws of the universe are: (1) F = ma; (2) $F = Gm_1m_2/d^2$; and so on, for all 10^{10} laws. Let $A_1 = (F - ma)^2$, $A_2 = (F - Gm_1m_2/d^2)^2$, and so on. Let $A = A_1 + A_2 + \ldots + A_{10}10$. The theory can now be formulated in the unified, simple form A = 0. (This is true if and only if each $A_r = 0$, for r = 1, 2, … 10^{10}).

How are *degrees* of unity to be specified? And how is giving persistent preference to unified theories to be justified? Standard empiricism cannot answer these questions, as we saw on pages 211-14. The problem of 'simplicity' or 'unification', as this group of problems tends to be called, is widely understood to be a fundamental problem of the philosophy of science (Salmon, 1989). Even Einstein recognized the problem, acknowledged he did not know how to solve it, but said that it should be possible to solve (see Maxwell, 1998, pp. 105-6). Can AOE do better?

When I wrote the first edition of this book I was convinced that AOE must be able to solve these problems, in particular the problem of what theoretical unity *is*, but I did not know how it could be done. Four years later, while thinking about the problem in order to distract my attention away from an agonizing pain in my neck caused by a nerve squeezed between two vertebrae, the crucial insight come to me in a flash. Those who have attempted to solve the problem have been looking at entirely the wrong thing. They have been looking at the *theory itself*, its axiomatic structure, its number of postulates, its formulation, its characteristic derivations, the language in which it is formulated. But all this is wrong. What one needs to look at is not the theory itself, but what the theory says about the world, the *content* of the theory in other words. One needs to look, not at the theory, but at the world, or rather at the world as depicted by the theory. At a stroke the worst aspect of the problem of what unity *is* vanishes. No longer does one face what may be called the *terminological* problem of unity – the problem, namely, that the extent to which a theory is unified appears to be highly dependent on the way the theory is *formulated*. Suppose we have a given theory T, which is formulated in N different ways, some formulations exhibiting T as beautifully unified, others as horribly complex and disunified, but all formulations being interpreted in precisely the same way, so as to make precisely the same assertion about the world. If unity has to do exclusively with *content*, then *all these diverse formulations of T, having the same content, have precisely the same degree of unity*. The variability of apparent unity with varying formulations of one and the same theory, T, (given some specific interpretation) – which poses such an insurmountable problem for traditional approaches to the problem (see Salmon, 1989; Maxwell, 1998, pp. 56-68) – poses no problem whatsoever for the thesis that unity has to do with *content*. Variability of formulation of a theory which leaves its content unaffected is wholly irrelevant: the unity of the theory is unaffected.

But now we have a new problem: How is the unity of the *content* of a theory to be assessed? What exactly does it mean to assert that a dynamical physical theory has a unified content?

What it means is that the theory has *the same* content throughout the range of possible phenomena to which the theory applies.[7] Unity, in other words, means that there is just *one*

---

[7] What is invariant throughout the range of phenomena to which the theory applies is what is asserted by the *differential equations* of the theory. A simple example of a differential equation is $dy/dx = 2x$. This represents an infinite family of curves (or functions), $y = x^2 + A$, each curve being given by assigning a different value to the constant, A. This is a general feature of differential equations: they represent infinitely many different curves or functions. In physics, these functions, the so-called 'solutions' of the differential equations of the physical theory, determine how the different physical systems (to which the theory applies) evolve in space and time. It is in this way that *one and the same* differential equation can apply to infinitely many diverse physical phenomena – diverse physical systems which trace out quite different paths through space with the passage of time. (What, it may be asked, is dy/dx? It expresses the rate at which y is changing with respect to changes in x. Equivalently, it expresses the slope of the tangent to the curve of the function in question. If dy/dx is big, a tiny change of x means a big change in y. If dy/dx is small, it needs a big change in x before there is much change in y. Just what one would expect granted that dy/dx is the slope of the tangent to the curve! Differential equations specify the fixed, unchanging way in which quantities change with respect to other quantities, these quantities being, as far as physics is concerned, such things as position, velocity, acceleration, time. Consider a function, such as $y = x^2$. Pick any point (x,y) on the curve of the graph of this function, and consider a second point very close to it $(x + \delta x, y + \delta y)$, where δx and δy are minute numbers. The fraction, $\delta y/\delta x$ is close to representing the slope of the tangent to the curve at the point (x,y). As δx and δy get closer and closer to zero, and $(x + \delta x,$

content throughout the range of possible phenomena to which the theory applies. If the theory postulates *different* contents, *different* laws, for different ranges of possible phenomena, then the theory is *disunified*, and the more such *different* contents there are so the more *disunified* the theory is. Thus 'unity' means 'one', and 'disunity' means 'more than one', the disunity becoming worse and worse as the number of different contents goes up, from two to three to four, and so on. Not only does this enable us to distinguish between 'unified' and 'disunified' theories; it enables us to assign 'degrees of unity' to theories, or to partially order theories with respect to their degree of unity.[8]

All this can be illustrated by considering 'aberrant' and 'non-aberrant' theories of the kind discussed in chapter 9. Thus Newton's theory of gravitation, $F = GM_1M_2/d^2$ is unified in that what the theory asserts is *the same* throughout all possible phenomena to which it applies (all bodies of all possible masses, constitution, shape, relative velocity, distance apart, at all times and places). An aberrant version of this theory that asserts that this law is an inverse cube law after some definite time t_o, so that $F = GM_1M_2/d^2$ for times $t \leq t_o$ and $F = GM_1M_2/d^3$ for times $t > t_o$, is disunified because what the theory asserts is *not* the same throughout the range of possible phenomena to which the theory applies.

Note that special terminology could be introduced to make Newtonian theory look *disunified*, and the aberrant version of Newtonian theory look *unified*. All we need do is interpret 'd^N' to mean 'd^N if $t \leq t_o$ and d^{N+1} if $t > t_o$'. In terms of this (admittedly somewhat bizarre) terminology, the aberrant theory has the form '$F = GM_1M_2/d^2$ ', and Newtonian theory has the 'aberrant' form '$F = GM_1M_2/d^2$ for times $t \leq t_o$ and $F = GM_1M_2/d$ for times $t > t_o$'. But this mere *terminological* reversal of aberrance or disunity does not affect the *content* of the two theories: the content of Newtonian theory remains unified, and the content of the aberrant version (which looks unified) remains disunified. For unity, in other words, we require that the theory is *terminologically* invariant throughout the range of possible phenomena to which it applies when *terminology*, used to formulate the theory, is itself invariant throughout the range of possible phenomena (so that *terminological* invariance implies *content* invariance).

In practice in physics assessments of degrees of unity are somewhat more complex than I have indicated so far because of the following consideration. In assessing the extent to which a theory is disunified we may need to consider *in what way* different, or *how* different, from one another, the different contents of a theory are. A theory that postulates different laws at different times and places is disunified in a much more serious way than a theory which postulates the same laws at all times and places, but also postulates that distinct kinds of physical particle exist, with different dynamical properties, such as charge or mass. This second theory still postulates *different* laws for different ranges of phenomena: laws of one kind for possible physical systems consisting of one kind of particle, and slightly different laws for possible physical systems consisting of another kind of particle. But this second kind of difference in content is much less serious than the first kind (the kind that involves different laws at different times and places).

What this means is that there are different *kinds* of disunity, different *dimensions* of disunity, as one might say, some more serious than others. We can, I suggest, distinguish at least eight different kinds of disunity, as follows.

Any dynamical physical theory, T, can be regarded as specifying an abstract space, S, of possible physical states to which the theory applies, a distinct physical state corresponding to

---

$y + \delta y$) gets closer and closer to (x,y), so $\delta y/\delta x$ gets closer and closer to expressing precisely the slope of the tangent at the point on the curve, (x,y). In the limit, as δx and δy approach zero, so $\delta y/\delta x$ approaches the true value of the tangent, dy/dx. It is not hard to show that in the case of the function $y = x^2$, $dy/dx = 2x$ for any point on the curve (x,y). In this case, as x gets bigger, so dy/dx gets bigger too, just as the graph of the function, a parabola, indicates.) For further details see Maxwell (1998, appendix).

[8] For earlier accounts of my proposed solution to the problem of unity of physical theory see Maxwell (1998, chs. 3 and 4; 2004b, chs. 1-2 and appendix; 2004a; 2004c; and 2007b).

each distinct point in S. For unity, we require that T asserts that the same dynamical laws apply throughout S, governing the evolution of the physical state immediately before and after the instant in question. If T postulates N distinct dynamical laws in N distinct regions of S, then T has disunity of degree N.

Eight different kinds of disunity can now be specified. [These are numbered (8) to (1), rather than (1) to (8), because in the next section these eight increasingly restrictive notions of unity will be incorporated into AOE.] In what follows, in connection with (3) and (2), there are a few physical and mathematical technicalities, which I attempt to explain. Some may find my explanations unhelpful; if so, I hope that (8) to (4) will be crystal clear, and will convey the general idea satisfactorily.)

(8) T divides space-time up into N distinct regions, $R_1...R_N$, and asserts that the laws governing the evolution of phenomena are the same for all space-time regions within each R-region, but are different within different R-regions.[9]

(7) T postulates that, for distinct ranges of physical variables (other than position and time), such as mass or relative velocity, in N distinct regions, $R_1,...R_N$ of the space of all possible phenomena, distinct dynamical laws obtain.

(6) In addition to postulating non-unique physical entities (such as particles), or entities unique but not spatially restricted (such as fields), T postulates, in an arbitrary fashion, N - 1 distinct, unique, spatially localized objects, each with its own distinct, unique dynamic properties.

(5) T postulates physical entities interacting by means of N distinct forces, different forces affecting different entities, and being specified by different force laws. (In this case one would require one force to be universal so that the universe does not fall into distinct parts that do not interact with one another.)

(4) T postulates N different kinds of physical entity,[10] differing with respect to some dynamic property, such as value of mass or charge, but otherwise interacting by means of the same force.

(3) Consider a theory, T, that postulates N distinct kinds of entity (e.g. particles or fields), but these N entities can be regarded as arising because T exhibits some symmetry (in the way that the electric and magnetic fields of classical electromagnetism can be regarded as arising because of the symmetry of Lorentz invariance, or the eight gluons of chromodynamics can be regarded as arising as a result of the local gauge symmetry of SU(3)). If the symmetry group, G, is not a direct product of subgroups, we can declare that T is fully unified; if G is a direct product of subgroups, T lacks full unity; and if the N entities are such that they cannot be regarded as arising as a result of some symmetry of T, with some group structure G, then T is disunified.[11] (See note 11, and below, for clarification.)

---

[9] As I have formulated it here, (8) is open to two somewhat different interpretations. First, for N = 1 we require only that *the same* law operates throughout space in the sense that this would be true even if the law in question asserted that all objects experience a force directed at a unique point in space, and inversely proportional to their distance from that point. Second, for N = 1, we require that *the same* law operates throughout space in the sense that a mere change of position in space of an isolated physical system has no effect on the way the system evolves. An analogous distinction arises in connection with time. In what follows I adopt the second interpretation of (8), which means that a theory which is unified with respect to (8) exhibits symmetry with respect to spatial location, and time of occurrence. As far as the *ad hoc* version of Newtonian theory is concerned, N = 2 for both versions of (8).

[10] Counting entities is rendered a little less ambiguous if a system of M similar particles is counted as a (somewhat peculiar) field. This means that M particles all of the same kind (i.e. with the same dynamic properties) is counted as *one* entity. In the text I continue to adopt the convention that M particles, all the same dynamically, represents one *kind* of entity, rather than one entity.

[11] A few words of explanation. A homogeneous sphere exhibits symmetry in that it can be rotated through any angle about its centre, and it remains the same. Group theory is the mathematical theory of symmetry. Given any symmetric object, there will be a set of operations, a, b, c, … which, when performed on the object leave it unchanged. (In the case of the sphere, the operations are rotations about the centre.) These

(2) If (apparent) disunity of there being N distinct kinds of particle or distinct fields has emerged as a result of a series of cosmic spontaneous symmetry-breaking events, there being manifest unity before these occurred, then the relevant theory, T, is unified. If current (apparent) disunity has not emerged from unity in this way, as a result of spontaneous symmetry-breaking, then the relevant theory, T, is disunified.[12] (See below for clarification.)

(1) According to GR, Newton's force of gravitation is merely an aspect of the curvature of space-time. As a result of a change in our ideas about the nature of space-time, so that its geometric properties become dynamic, a physical force disappears, or becomes unified with space-time. This suggests the following requirement for unity: space-time on the one hand, and physical particles-and-forces on the other, must be unified into a single self-interacting entity, U. If T postulates space-time and physical 'particles-and-forces' as two fundamentally distinct kinds of entity, then T is not unified in this respect.

For unity, in each case, we require N = 1. As we go from (8) to (4), the requirements for unity are intended to be accumulative: each presupposes that N = 1 for previous requirements. As far as (3) and (2) are concerned, if there are N distinct kinds of entity which are not unified by a symmetry, whether broken or not, then the degree of disunity is the same as that for (5) and (4),

---

operations, a, b, c, … form a group, the symmetry group of the object. They must obey the following axioms. (1) There is the identity operation, i, which does nothing. (2) Any two operations, a and b say, can be combined to form a third, c, so that a.b = c. (3) Every operation, a, has an inverse, a^{-1}, so that $a.a^{-1} = i$. (4) Repeated operations are associative, so that a.(b.c) = (a.b).c. There are many different sorts of groups, finite, infinite, discrete, continuous. The symmetry group of the sphere is called SO(3).

A group G is a direct product of subgroups G_1 and G_2, written $G = G_1 \otimes G_2$, if the following three conditions hold: (a) G_1 and G_2 are subsets of G and groups in their own right, (b) $g_1.g_2 = g_2.g_1$, where g_1 is any member of G_1 and g_2 is any member of G_2, and (c) any member g of G is such that $g = g_1.g_2$, for some unique pair belonging to G_1 and G_2 respectively.

In theoretical physics, a symmetry arises when, given any isolated physical system (perhaps of some specific type) some specific kind of change is made to the system, and it evolves in time in just the same way, as if the change had not been made. Thus, given any isolated system, changing merely (a) its location in space, (b) its orientation in space, (c) its time of occurrence, or (d) its uniform velocity, leaves unaffected the way the system evolves. These are space-time symmetries, and apply to all dynamical physical theories (which presuppose that space-time is flat). Lorentz invariance is the name given to the symmetry, postulated by Einstein's special theory of relativity, which any physical system exhibits when its uniform velocity is changed.

There are, in addition, symmetries that apply to specific theories. Thus quantum field theories of electromagneticsm, the electoweak force, and the 'strong' force (which holds quarks together inside protons and neutrons) exhibit a symmetry called 'global gauge invariance'. A feature of the physical state of the system, called the phase, can be changed by any fixed amount everywhere, at some instant, and the system evolves exactly as before. This can be transformed into a local symmetry, called 'local gauge invariance' as follows. At some instant, the phase is changed by different amounts at different places, and to compensate, the field is changed by different amounts at different places, but in ways that are determined by the (arbitrary) changes made to the phase: the result of these two compensating changes is that the physical system evolves as before, as if nothing had been changed. The symmetry groups of these local gauge symmetries of quantum electrodynamics, electroweak theory, and chromodynamics (the quantum field theories of electromagneticsm, the electroweak force, and the strong force respectively), are called U(1), U(1) ⊗ SU(2), and SU(3). Unlike U(1) and SU(3), U(1) ⊗ SU(2) is a direct product of subgroups, as the nomenclature indicates.

A somewhat more detailed, but still informal account of these matters is given in Maxwell (1998, ch. 4, sections 11 to 13, and the appendix, pp. 257-65). For rather more detailed accounts of the locally gauge invariant structure of quantum field theories see: Moriyasu (1983), Aitchison and Hey (1982: part III), and Griffiths (1987, ch. 11). For a delightful informal account of the role of symmetry and group theory in physics, see Zee (1986). For more technical introductory accounts see Isham (1989) or Jones (1990).

[12] For accounts of spontaneous symmetry breaking see Moriyasu (1983) or Mandl and Shaw (1984).

depending on whether there are N distinct forces, or one force but N distinct kinds of entity between which the force acts.

(1) does not introduce a new kind of unity, but introduces, rather, a new, more severe way of counting different kinds of entity. (4) to (2) require, for unity, that there is one kind of self-interacting physical entity evolving in a distinct space-time, the way this entity evolves being specified, of course, by a consistent physical theory. According to (4) to (2), even though there are, in a sense, two kinds of entity, matter (or particles-and-forces) on the one hand, and space-time on the other, nevertheless N = 1. According to (1), this would yield N = 2. For N = 1, (1) requires that matter and space-time are no more than aspects of one basic entity (unified by means of a spontaneously broken symmetry, perhaps).

As we go from (8) to (1), then, requirements for unity become increasingly demanding, with (3) and (2) being at least as demanding as (5) and (4), as explained above.

One qualification ought, perhaps, to be added to the above. Isolated physical systems, that exhibit perfect symmetry related to the symmetries of the underlying theory, may evolve in accordance with a simplified version of the theory. Thus, given Newtonian theory (NT), two spheres of equal mass and dimension, rotating about a point equidistant between them, move in accordance with a simplified version of NT. They rotate uniformly in a circle whose centre is the mid point between the two spheres. This is not to be interpreted as a manifestation of disunity. (One could, of course, consider taking such anomalies seriously, and demand that a perfectly unified theory must be such that it does not permit physical systems which exhibit such symmetries perfectly, to exist.)

Let me now take (8) to (1) in turn, and give, in each case, an example of a theory with some degree of disunity.

(8) T asserts: Up to the last instant of the 21st century, NT holds; from the next instant on, a version of NT holds with the gravitation force repulsive instead of attractive. T, here, is disunified to degree N = 2, in a type (8) way.

(7) T asserts: everything occurs as NT asserts, except for the case of any two solid gold spheres, each having a mass of between one and two thousand tons, moving in otherwise empty space up to a mile apart, in which case the spheres attract each other by means of an inverse cube law of gravitation. T is again disunified to degree N = 2, in a type (7) way.

(6) T asserts: everything occurs as NT asserts, except there is one object in the universe, of mass 8 tons, such that, for any matter up to 8 miles from the centre of mass of this object, gravitation is a repulsive rather than attractive force. The object only interacts by means of gravitation. T, here, is again disunified to degree N = 2, in a type (6) way.

(5) T postulates particles that interact by means of Newtonian gravitation; some of these interact by means of an electrostatic force $F = Kq_1q_2/d^2$, this force being attractive if q_1 and q_2 are oppositely charged, otherwise being repulsive, the force being much stronger than gravitation. T, here, is disunified to degree N = 2, in a type (5) way.

(4) T postulates particles that interact by means of Newtonian gravitation, there being three kinds of particles, of mass m, 2m and 3m. Here, N = 3, in a type (4) way.

(3) T postulates the classical electromagnetic field, composed of the electric and magnetic fields, obeying Maxwell's equations for the field in the vacuum. The symmetry of Lorentz invariance unifies these two fields (see below). Here, N = 1, in a type (2) way.

(2) T is Weinberg's and Salam's electroweak theory, according to which at very high energies, such as those that existed soon after the big bang, the electroweak force has the form of two forces, one with three associated massless particles, two charged, W^- and W+, and one neutral, W^o, and the other with one neutral massless particle, V^o. According to the theory, the two neutral particles, W^o and V^o, are intermingled in two different ways, to form two new, neutral particles, the photon, γ, and another neutral massless particle, Z^o. As energy decreases, the W+, W- and Z^o particles acquire mass, due to the mechanism known as spontaneous symmetry-breaking (involving another, hypothetical particle, not yet detected, called the Higgs particle), while the

photon, γ, retains its zero mass. There appear to be two new, very different forces, the weak and electromagnetic. This theory unifies the weak and electromagnetic forces as a result of exhibiting the symmetry of local gauge invariance; this unification is only partial, however, because the symmetry group is a direct product of two groups, U(1) associated with V^o, and SU(2) associated with W^-, W+ and W^o.[13] This is type (7) unity.

(1) One might imagine a version of string theory without strings, different vibrational modes (perhaps) of empty, compactified six-dimensional space giving rise to the appearance of particles and forces, even though in reality there is only 10 dimensional space-time. Or one might imagine that the quantization of space-time leads to the appearance of particles and forces as only apparently distinct from empty space-time. In either case, N = 1 in a type (1) way: there is just the one self-interacting entity, empty space-time.

In all eight cases, disunity arises because *different* laws govern different regions in the space of all possible phenomena predicted by the theory in question. This is obvious as far as (8) is concerned. In the case of (7), if laws are different depending on whether the value of some variable V is less or greater than some value V_O, then for those parts of the space of all possible phenomena, S, in which $V < V_O$, laws governing phenomena will be different from parts of S in which $V > V_O$. In the case of (6), regions of S in which the unique dynamic object is not present will be different from regions in which it is present. In the case of (5) and (4), regions of S in which only one kind of force or particle prevails will be governed by laws different from other regions in which a different kind of force or particle prevails.

As far as (3) is concerned, the point is perhaps best made by considering the particular case of classical theory of the electromagnetic field formulated so as to conform with Einstein's theory of special relativity. According to special relativity, the electromagnetic field is made up of two fields, the electric and the magnetic fields. On the face of it, there will be regions of S in which there is just an electric field, and other regions in which there is just a magnetic field, which means disunity. According to special relativity, however, a mere change of uniform velocity (with respect to a reference frame) cannot affect the way a system evolves: such a change leaves everything dynamically significant unchanged (as does a mere change of position or orientation in space). However, given any specific electromagnetic field, the way this divides up into an electric and magnetic field *is* changed by a change of uniform motion. In particular a pure electric field will become an admixture of electric and magnetic fields, and a pure magnetic field will become an admixture of magnetic and electric fields. Granted that a mere change of uniform relative motion does not change anything dynamically, or physically, significant, we are obliged to hold that the electric and magnetic fields cannot be separated out in the way required for disunity. There is one unified entity, the electromagnetic field, with electric and magnetic aspects. Both aspects are always present, although, for some quite specific fields, this will not be apparent relative to a reference frame in *one very specific state of motion only*.

The paradigmatic illustration of (2) is, as I have indicated, Weinberg's and Salam's theory of the electroweak force. On the face of it, the four particles of the theory are very different, and cannot be transformed into each other by means symmetry operations. The photon (associated with the electromagnetic force) is massless, whereas the particles associated with the weak force, W^+, W^- and Z^o, have mass (nearly 100 times the mass of the proton), and two of these particles are charged. The underlying electroweak theory nevertheless possesses the local gauge symmetry of U(1) ⊗ SU(2). All these particles at high energies, soon after the big bang, are massless. As the universe cooled, a kind of asymmetry developed in the vacuum, associated with the as-yet undetected Higgs particle, and it is this which creates the asymmetry between the particles of the theory. Something analogous occurs when a uniform block of iron is gradually cooled. The lowest energy state involves the minute magnets associated with the atoms of iron aligning

[13] For further discussion see (Maxwell 1998, 131-40, 257-65 and additional works referred to therein.) See also notes 11 and 12.

themselves so that there is an overall magnetic field in some specific direction. There is a loss of spatial symmetry – the symmetry is 'broken'. The underlying theory of electrodynamics does not, however, pick out any preferred direction. The theory has directional symmetry, even if the block of iron does not.

As far as (1) is concerned, if space-time and particles-and-forces are distinct then, presumably, one region (or possibly point) of S consists of nothing more than empty space-time.

Granted that theoretical physics is pursued in such a way that theories that fail to satisfy (8) to (6) are rejected, whatever their empirical success might otherwise be, it is clear that this means that physics thereby assumes that the universe is such that no physical theory which violates (8) to (6), with N > 1, is true. This accords with AOE but violates standard empiricism. Standard empiricism cannot solve the problem of what theoretical unity is because it cannot endorse the crucial point that unity applies to the *content* of theories (and at the same time hold that unity considerations may over-ride empirical considerations) because this would commit standard empiricism to holding that science permanently accepts a metaphysical thesis (no disunified theory is true), which contradicts a basic tenet of the doctrine. Standard empiricism can only solve the problem by becoming inconsistent!

That the problem of what unity is can be solved granted AOE, but cannot be solved granted standard empiricism, is an enormous success for the former view.

Is AOE *required* in order to solve the problem of unity? Could it not be argued that a view which acknowledges, merely, that science makes the metaphysical assumption 'the universe is such that no theory that is disunified in senses (8) to (6) is true' is able to solve the problem of unity? There are two objections to such a claim. First, this fails to provide a rationale for biasing choice of theory unified in senses (5) to (1). Second, such a conception of science lacks the rationality of AOE: it dogmatically upholds its one metaphysical assumption (which might after all be false), whereas AOE allows science to *modify* such assumptions in the light of the empirical progress achieved by the rival research programmes to which rival assumptions lead. AOE is permanently committed only to assumptions required to be true for the enterprise of acquiring knowledge to meet with any success at all.

I have formulated the above eight requirements for unity as applying to the individual theory. Formulated in this way, there is an obvious objection. In the case of requirements (6) to (2), the methodological demand that an acceptable theory be unified can always be satisfied trivially: given a theory disunified to degree N = 6, let us say, this can always be split into six theories, each unified with N = 1. The way to cope with this objection is to interpret (8) to (1) as applying to the totality of fundamental physical theory, and to empirical laws if there is no theory which predicts and explains them.

I now consider briefly three questions that may be asked in connection with this proposed solution to the problem of unity of physical theory.

First, what of 'simplicity'? Is this the same as 'unity', or something distinct? The 'simplicity' of a theory can be interpreted as having to do, not with whether the *same* laws apply throughout the space of possible phenomena predicted by the theory in question, but rather with the *nature* of the laws, granted that they are the same. Some laws are simpler than others. In order to overcome the objection that simplicity is formulation dependent it is essential, as in the case of unity, to interpret 'simplicity' as applying to the *content* of theories, and not to their *formulation*, their *axiomatic structure*, etc. For details, see Maxwell (1998, pp. 157-9). It is a further great success of AOE that it succeeds in distinguishing sharply between these two aspects of the problem of what the explanatory character of a physical theory *is*, namely the *unity* aspect, and the *simplicity* aspect, and succeeds in solving both.

On the face of it, mere terminological simplicity can play no important heuristic or methodological role in physics at all because, given any *unified* theory, it can be made as simple or complex as we like by appropriate choice of terminology. But what is paradoxical about the role of simplicity in physics is that terminological simplicity does, in practice, seem to be highly significant heuristically and methodologically. How is this paradox to be resolved? A part of the answer is that what matters, for physics, is that a theory should be simple when formulated in terminology appropriate to a good, acceptable metaphysical blueprint – terminology that, for example, conforms to the symmetries of the blueprint. In addition, it is important that different laws and theories, applicable to different phenomena, should *all* be simple when formulated in the same appropriate terminology. The demand that all physical laws and theories should, as far as possible, be formulated in a common terminology appropriate to the best available blueprint means that terminological simplicity ceases to be something that can always be cooked up artificially, and becomes something that is heuristically and methodologically significant. (For details see Maxwell, 1998, pp. 110-3.)

Second, does the question of whether laws governing a range of phenomena *remain the same* throughout their range of application have an unambiguous answer, in view of Goodman's 'grue' and 'bleen' paradox (Goodman, 1954)? Adapting Goodman's notions slightly, an object is grue if it is green up to the last moment of the 21st century, and blue thereafter; an object is bleen if it is blue up to the last moment of the 21st century, and green thereafter. Are not grue and bleen just as good predicates as blue and green? If the colours of objects change dramatically at the end of the 21st century, so that blue objects become green, and green objects blue, can we not, with equal legitimacy, say that there is no change, objects continue to be grue and bleen? This much discussed paradox is, in my view, very largely a red herring. On the face of it, the distinction made above, between formulation and content, suffices to dismiss the paradox. The sentence 'This object is grue' (S) may, as far as its written form is concerned, be invariant through the end of the 21st century, but what this sentence asserts, its content, is not invariant. To this, the reply may be made that the *content* of S may be regarded as being invariant. But this is not what is ordinarily meant by 'invariant' or 'remain the same': the above account of unity of theory appeals to the ordinary meaning of 'invariant' or 'remains the same', and not the perverse grue and bleen meaning. Two additional points. It should be noted that the Goodman paradox implicitly accepts the ordinary meaning of 'remains the same' in employing the terminology of 'grue' and 'bleen', terminology which remains the same, in the ordinary sense, throughout the end of the 21st century. Second, that the content of grue and bleen is not invariant with respect to the passage of time – unlike the content of blue and green which is invariant – is demonstrated by the point that if objects really are grue and bleen, and a person is convinced of this, then he can tell, by looking at grass and sky, whether or not the 21st century has ended, whereas the same is not true with respect to green and blue. Grue and bleen implicitly refer to a specific time in a way in which green and blue do not.

Third, Goodman's point concerning the ambiguity of 'remains the same' may seem to gain support from the mathematical notion of a function as a rule which takes one from one set of numbers to another. According to this notion, the two functions (1) $y = 3x$ for all x, and (2) $y = 3x$ for $x \leq 2$ and $y = 4x$ for $x > 2$, are equally good functions. Both functions 'remain the same' as x increases and passes through the value $x = 2$. Clearly, we need a narrower notion of function than this if we are to be able to distinguish between functional relationships which do, and which do not, 'remain the same' as values of variables change. We need to appeal to what may be called 'invariant functions', functions which specify some fixed set of mathematical operations to be performed on 'x' (or its equivalent) to obtain 'y' (or its equivalent). In the example just given, (1) is invariant, but (2) is not. (2) is made up of two truncated invariant functions, stuck together at $x = 2$. Functions that appear in theoretical physics are *analytic*; that is, they can be represented as a power series (Penrose, 2004, pp. 112-4). Analytic functions are repeatedly differentiable. Such functions have the remarkable property that from any small bit of the function, the whole function can be reconstructed uniquely, by a process called 'analytic continuation'. All analytic functions are thus invariant. The latter notion is however a wider one, and theoretical physics might, one day, need to employ this wider notion explicitly, if space and time turn out to be discontinuous, and analytic functions have to be abandoned at a fundamental level.[14]

This concludes my discussion of what it means to say of a theory that it is simple and unified. AOE not only solves the problem of what simplicity and unity *are*; it also solves the problem of why it is *rational* for science persistently to accept only those theories that are sufficiently simple and unified (as well as being sufficiently empirically successful, of course).[15] Standard empiricism fails to solve both problems.

It deserves to be noted that (8) to (1), in addition to explicating what it means to say of a dynamical theory that it is unified, also explicates eight different meanings that can be given to *physicalism*. Physicalism(n), for n = 8, 7, … 1, can be interpreted to assert: the universe is such that the true theory of everything is unified in an (n) type way, with N = 1. This will be exploited in the next section. More generally, the above provides us with the means to throw a two-dimensional grid over all possible *partially physically comprehensible* universes. We can interpret physicalism(n,N) to assert: the universe is such that the true theory of everything is disunified in an (n) type way to extent N, with n = 8, 7, … 1, and N = 1, 2, … ∞.

3 A Further Extension of Aim-Oriented Empiricism (AOE)

The above layered interpretation of physicalism makes possible another version of AOE, relevant specifically to physics and modern science since Galileo. The different versions of physicalism(n, 1), as n goes from 8 to 1 correspond, in this version, to increasingly substantial and restrictive metaphysical theses and associated methods: see figure 14. Physicalism(4-2) are on the same level since they are all but equivalent to one another. As we descend the hierarchy, from level (8) to (1), theses become increasingly specific, increasingly potentially fruitful for future progress in theoretical physics, but also increasingly likely to be false and in need of revision. The corresponding methodological requirements for unity, as explicated in the last section, become increasingly demanding, but also increasingly speculative and uncertain. The totality of physical theory, at any given stage in the development of physics (except when a candidate unified theory of everything has been proposed and accepted) will only satisfy these methodological rules partially; a new theory, in order to be an advance from the standpoint of unity, must lead to a new totality of theory satisfying the methodological rules better than the previous totality.

In figure 14, each version of physicalism is taken to assert that the true theory of everything is unified to the full extent (in that sense) with N = 1. This restriction could conceivably be relaxed if the search for unity persistently failed.

Even with the restriction relaxed, however, the version of AOE depicted in figure 14 may turn out to be false. If we exclude from consideration physicalism(n = 8, N = ∞) which permits anything, AOE as depicted in figure 14 assumes that the universe is at least partially physically comprehensible in the sense that phenomena occur in accordance with physical laws which are more or less unified, the traditional distinction between laws and initial conditions being presupposed. But even though the universe is physically comprehensible, the traditional distinction between laws and initial conditions might not be observed. As we shall see in the next section, the true theory of everything might be cosmological in character, and might specify unique initial conditions for the universe. This possibility, and other possibilities of this kind, could no doubt be accommodated

---

14 For a fascinating discussion of the problems that arise in connection with the wider notion of what I have called ‘invariant function’, see Roger Penrose’s discussion of what he calls the ‘Eulerian’ notion of function: Penrose (2004, 6.4).

15 For further details concerning this solution to the problem of unity of physical theory, see Maxwell (1998, especially chs. 3 and 4 and the appendix; 2004b, chs. 1, 2, and appendix, section 2; 2004c; 2004d).

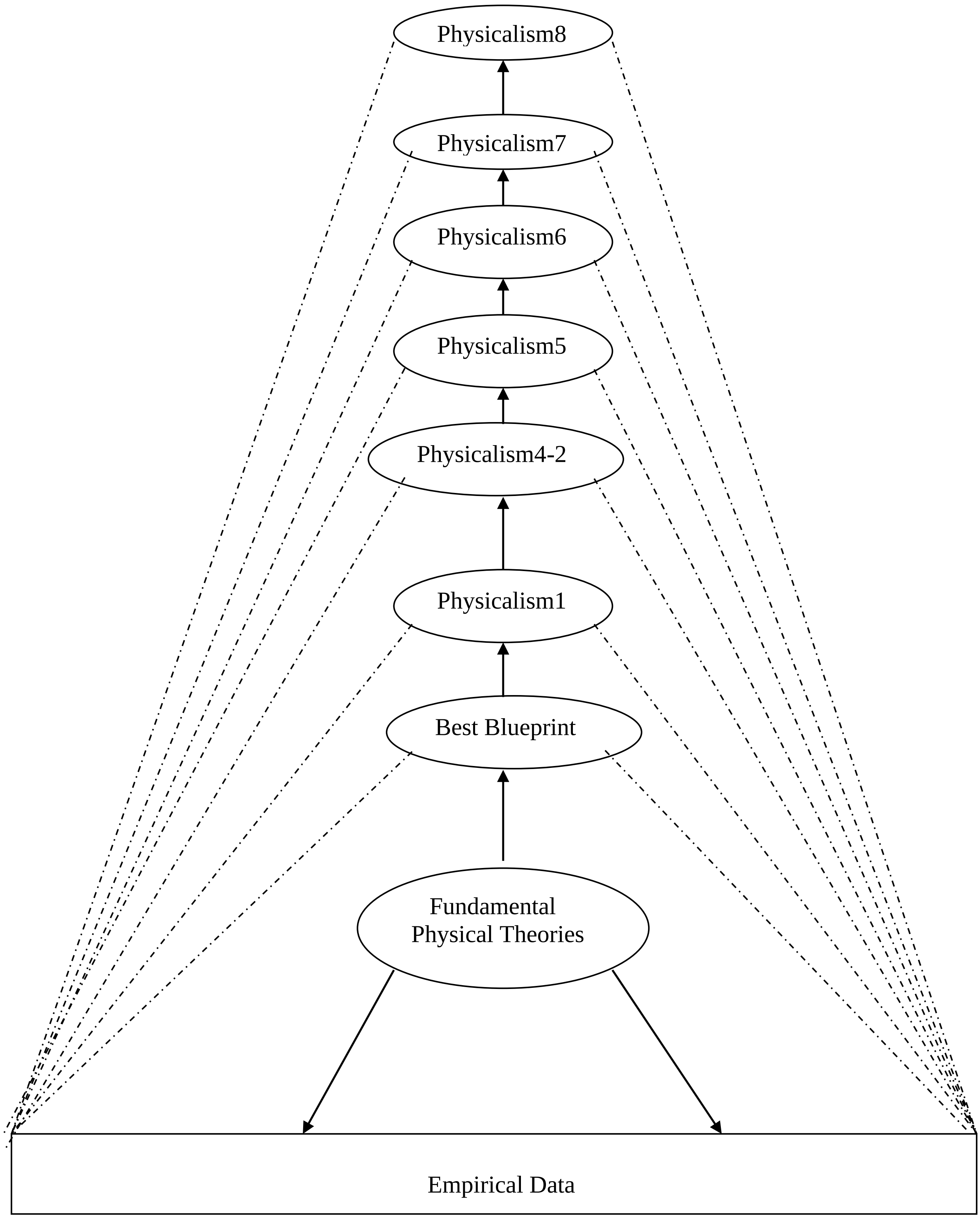

Figure 14: Another version of Aim-Oriented Empiricism (AOE)

within a modified version of the above view. But there are other possibilities, of philosophical interest even if of no interest to physics as at present constituted, which cannot be so accommodated. Perhaps God is ultimately responsible for all natural phenomena, or some kind of cosmic purpose or cosmic programme analogous to a computer programme (as has been suggested). In these cases the universe would be comprehensible but not physically comprehensible – even though it might mimic a physically comprehensible universe.

In order to accommodate these, and other such, possibilities we need to embed the version of AOE depicted above in the version depicted in figures 13 or 12: see figure 15.

Figure 15: Yet another version of Aim-Oriented Empiricism

4 Atomistic and Cosmic Physicalism

A basic motivation for making explicit metaphysical assumptions implicit in the methods of physics is that it provokes us into inventing new metaphysical possibilities, which we might not otherwise have considered. We are much more victims of *implicit* assumptions – of assumptions we deny making – than of assumptions we make explicit. This consideration prompts the question: Are alternatives to physicalism(n,1) with n = 8 …1, conceivable?

The eight versions of physicalism depicted in figure 14 all hold that the physical universe, at any given instant,[16] is made up of two distinct aspects, which we may call **U** and **V**. **U** is what is depicted by the true physical theory of everything, T. It is inherent in all phenomena, everywhere, at all times. It does not itself change, but determines (perhaps probabilistically) the way that which changes does change. **V**, by contrast, is what does change and vary, from moment to moment, and from one place to the next. **U** and **V** together, at one instant, determine (perhaps probabilistically) **V** at the next instant.

This distinction between **U** and **V** can be traced back to atomism, the very first version of physicalism put forward by Democritus some two and a half thousand years ago. Given atomism, **U** consists of the unchanging properties of atoms and space, while **V** consists of the changing (relative) positions and motions of the atoms. As modern physics developed, ideas about the nature of **U** and **V** have changed, but the distinction itself has persisted up to the present. After Newton, rigid atoms interacting only by contact were transformed into point-atoms surrounded by rigid, centrally-directed fields of force. Here, **U** consists of the unchanging properties of the point-atoms and their surrounding fields of force (including the way the force falls off with distance and the affect it has on other point-atoms), while **V** consists of the changing (relative) positions and motions of the point-particles. Then it emerged, as a result of Maxwell's theory of the electromagnetic field and Einstein's theory of special relativity, that force fields are not rigid. Changes in the field take time to travel. This led to a new *unified field* version of physicalism, according to which everything is made up of an extended, self-interacting, unified field (matter being simply especially intense regions of the field). On the one hand there are changing, variable features of the field, **V**; and on the other, there are the unchanging features of the field, **U**, which determine how **V** changes, and which correspond to the laws of the true theory of the field. Subsequent developments have led to further changes in ideas as to what **U** and **V** are, but have not undermined the distinction itself.

It is no accident that the atomism of Democritus sharply distinguishes **U** and **V**. Atomism arose as an attempt to solve the problem of change, in particular the problem Parmenides posed with his argument that change involves a contradiction, and his view that the universe is a homogeneous, unchanging sphere.[17] Parmenides argued that change is impossible because the

---

[16] Or on any given spacelike hypersurface, looking at things from the standpoint of general relativity.

[17] This story has been told brilliantly by Karl Popper: see Popper (1998)

non-existent cannot exist, hence the world must be full, and hence there can be no room for movement or change. Democritus accepted the argument but rejected the conclusion. There is change, hence the non-existent must exist. The non-existent or, as we might say today, the void surrounds Parmenides's homogeneous, unchanging universe. Other Parmendian universes exist in the void. These can be shrunk down to a minute size, put in relative motion – and we have atomism. Each Democritean atom is a miniature Parmenidean universe. Atomism solves the problem posed by Parmenides by retaining as much as possible of the Parmenidean homogeneous, unchanging universe, but at the same time modifying this view just sufficiently to allow for change and diversity. Atomism solves the general problem of change – the problem of understanding how something can both remain the same *and* change – by segregating very precisely those aspects of atoms which do not change, and those which do change, the key to the distinction between **U** and **V**.

But there is another possible response to Parmenides. The universe as depicted by Parmenides – a homogeneous unchanging sphere – is a very special, uniquely unified state of the universe, the big bang state. This unified, initial state of the universe is unstable: spontaneous symmetry breaking occurs, and the universe evolves into a state made up of a great number of *virtual* big bang states. Today, every space-time point is made up of just one thing: a fleetingly existent, virtual big bang state.

Quantum theory can be interpreted as asserting that for very short intervals of time there is uncertainty of energy, and this permits so-called *virtual* particles to come into existence in the vacuum and almost immediately cease to exist. According to *cosmic physicalism* – the alternative to atomism as a response to Parmenides – every minute space-time region is composed, not of virtual particles, but of the virtual universe in its initial, unified, Parmenidean state. Before the big bang, unity is real and all disunity is virtual. After the big bang, disunity is real and unity is virtual. In a sense, there is only the big bang state. Variety and change come from the different ways in which instantaneously existent, virtual big bang states of the universe are inter-related.

There are, then, two distinct versions of physicalism which we may call *atomistic* and *cosmic* physicalism. They can be regarded as arising as a result of giving different responses to the challenge posed by Parmenides's impossible physically comprehensible universe.

Atomistic physicalism takes the Parmenidean universe to depict **U** – that aspect of the universe which does not change and which determines the way that which changes, **V**, does change. Initially, **U** represents the properties of the atom. Subsequent developments in theoretical physics have transformed **U**, so that it may be taken to represent the invariant properties of a unified field, a quantum field, space-time of variable curvature, and so on. Despite these developments, the distinction between **U** and **V** persists, and it is this which is the hallmark of atomistic physicalism.

Cosmic physicalism, by contrast, takes the Parmenidean universe to be a special, uniquely unified state of the universe – the big bang state. According to cosmic physicalism, the true theory of everything, T, specifies the properties of the universe in this state. At a fundamental level, the distinction between **U** and **V** does not arise. The distinction only arises when spontaneous symmetry breaking has occurred, and the universe consists of momentarily existing virtual big bang states. **V** consists of the different, changing ways in which these momentarily existing big bang states are inter-related.

There are other striking differences between these two versions of physicalism. Cosmic physicalism is inherently *cosmological* in character, whereas atomistic physicalism is not. According to cosmic physicalism, T of itself specifies the initial state of the universe, whereas according to atomistic physicalism, initial conditions are required in addition to T to specify the initial state of the universe. Cosmic physicalism is inherently *probabilistic*, since spontaneous symmetry breaking is an inherently probabilistic process, whereas atomistic physicalism may be deterministic or probabilistic. Cosmic physicalism must be quantum mechanical to the extent, at least, of incorporating the quantum mechanical distinction between *actual* and *virtual*. Atomistic physicalism makes no such demand.

The two versions of physicalism specify very different conditions for underlying unity to become apparent in as simple a way as possible. According to atomistic physicalism, this happens when the physical system being considered is as simple as possible – the vacuum, or a one particle system or, somewhat more complex, a two particle system. According to cosmic physicalism, it is exactly the opposite: underlying unity is made manifest in a system consisting of *everything* – the entire universe in a very special state, the initial big bang state.

Theoretical physics so far has presupposed atomistic physicalism. But it is possible that cosmic, and not atomistic, physicalism is true. Elsewhere I have indicated a number of recent developments in theoretical physics, from the increasingly variable and dynamic character of space-time as suggested by general relativity and quantum field theory, to the idea of spontaneous symmetry breaking and probabilism as suggested by the electroweak theory, which can be regarded as pointing in the direction of cosmic physicalism. For further details of the view, and arguments in support of the view, see Maxwell (2004b, appendix, section 5).

5 The Solution to the Problem of Verisimilitude

Physics advances from one false fundamental physical theory to another, and from one false level 3 blueprint to another. What, in this case, does it mean to say that physics is making progress? This is the problem of verisimilitude. Popper (1963, pp. 231-7) proposed a solution to the problem but, as we saw in chapter 9, this fails.

Philosophers of science, viewing the matter from a standard empiricist perspective, tend to regard the fact that physics advances from one false theory to another as having very negative implications for scientific progress. That physics will continue in this way has even been dubbed 'the pessimistic induction' (Newton-Smith, 1981, p. 14). But viewed from the perspective of aim-oriented empiricism (AOE), this manner of progression is actually to be expected, if physics really is making progress, and the universe really is physically comprehensible. For, if a theory, T_o, is precisely true throughout some restricted domain of phenomena D then, granted physicalism,[18] T_o must specify precisely what does not change, **U**, throughout all phenomena in D, and the way **U** determines how things change in D. But, according to physicalism, **U** exists unchanged throughout all phenomena. Thus, if T_o specifies the nature of **U** in D, it will be a straightforward matter to extend T_o so that it specifies **U** for all physically possible phenomena, T_o thus becoming the true theory of everything, T. Conversely, if T_o cannot be extended in this way to apply correctly to all phenomena, then T_o cannot be precisely true within D: T_o must be false. In brief, physicalism implies that a physical theory can only be precisely true of *anything* if it is (capable of being) precisely true of *everything*.

Granted, then, that physics proceeds, not by attaining T in one bound, but rather by developing a succession of theories that apply, with ever increasing accuracy, to ever wider ranges of phenomena until eventually a theory of everything is attained, it is inevitable, granted physicalism, that physics will progress by the development of theories that are all *false* throughout their domains of application until the ultimate, unified true theory of everything is attained (which will be precisely true about everything).[19] Since physicalism predicts that physics will progress in this way, the fact that physics has so far thus progressed can only count in favour of physicalism: it cannot count against physicalism and AOE, as some have supposed.[20]

There is just one conceivable exception to this argument. It is possible that the form of T (or the nature of **U**) might be such that T reduces to an especially simple form for an appropriately

---

[18] 'Physicalism' here, as elsewhere where there is no suffix, means 'physicalism(n,1) with $n \leq 4$'. We require a version of physicalism which asserts that there is an invariant **U** throughout all phenomena that are physically possible (according to that version of physicalism).

[19] Or rather, precisely true about that aspect of what exists which determines the way events evolve everywhere, at all times, throughout all phenomena.

[20] See, for example, Laudan (1980), Newton-Smith (1981).

simple or symmetric kind of system. Thus two spherical bodies of equal mass rotating about the point midway between them exemplify a law much simpler in form than Newtonian theory.[21] In having only what remains of T when it has been reduced to just such an especially simple form for some simple or symmetric system, one would have a true theory, but a theory *not* easily extendable to recover T. However, even if such a simplified version of T were to be formulated, it is most unlikely, before the discovery of T, that it would be correctly interpreted to apply only to appropriately symmetric kind of system. One would need to have T in order to know how to specify correctly systems to which the simplified version of T applies precisely. Interpreted to apply to a broader range of systems, the simplified version of T will not be precisely true.[22] It is in any case likely that the perfectly symmetrical system will not be a physical possibility in the actual universe. This is the case as far as the system consisting of two bodies rotating around each other, mentioned above, is concerned. However far away from other bodies this system might be, Newtonian theory, nevertheless, predicts that other bodies will slightly perturb the system, thus ensuring that it is not precisely symmetrical.

Given physicalism (and AOE), it is to be expected that physics advances by developing a succession of theories, T_o, T_1, T_2 ...T_n, which, though all false, and though all mutually incompatible, nevertheless deserve to be regarded as getting progressively closer and closer to the truth, T. But what does it *mean* to speak, here, of T_o ... T_n getting 'progressively closer and closer to the truth'?[23]

AOE solves the problem as follows. T_o ... T_n get 'progressively closer and closer to the truth', T, if and only if: T_n can be 'approximately derived' from T (but not vice versa), T_{n-1} can be 'approximately derived' from T_n (but not vice versa), and so on down to T_o being 'approximately derivable' from T_1 (but not vice versa).

In order to explicate the key notion, here, of 'approximate derivation' let us consider a special case. Let us take T_o to be Galileo's version of the heliocentric theory (G), T_1 to be Kepler's laws

---

[21] Given NT, precisely the right initial conditions, and nothing external to the system disturbing its evolution, the two spheres move in circular orbits with uniform speeds about the point midway between them. Interestingly enough, given GR, this is no longer the case: the rotating spheres radiate gravitational waves, and thus, very gradually lose energy. The spheres slowly spiral inwards – something that has been observed in the case of a double star system.

[22] And on the other hand if it is precisely true when applied to all physically possible systems that differ slightly from the symmetric systems then, granted physicalism, it will be readily extendable to become T.

[23] It is important, in my view, to regard the problem of verisimilitude as being a problem that arises, in the first instance, and perhaps exclusively, in connection with progress in fundamental theoretical physics. This is where the problem arose in the first place, with the discovery, the realization, that theoretical physics advances from one false theory to another, and yet does genuinely seem to be making progress. The problem, interpreted in this way becomes, if anything, even more acute when it is appreciated that if physics really is making progress towards depicting the comprehensible structure of the physical universe, as AOE implies, then physics *ought* to make progress by advancing from one false theory to another. There must, it seems, be a solution to the problem: What can it mean to talk of progress in these circumstances? Some have interpreted the problem in a much wider way, as the problem of specifying what it can mean, quite generally, to say of a succession of false propositions, p_1, p_2, … that they get, progressively closer and closer to the truth. But it is not at all certain that there is a solution to this more general problem. We do not need to solve this more general problem to say what we mean by progress in parts of natural science outside physics. This is because in these other areas of natural science it does not happen, in the same way, that science advances, predictably and rigidly, from one false theory to another. Harvey's theory that the heart pumps blood around the body, put forward long ago, still seems true today, and it is not easy to see how it could ever turn out to be false. The idea that *all* natural science advances from one false theory to another is itself, quite simply, false. In my view, then, the fact that the solution to the problem of verisimilitude, proposed here, is restricted to theoretical physics does not mean that this proposal is limited or inadequate. The problem – and the solution – *need* to be restricted in this way. Theoretical physics is where the problem belongs.

of planetary motion (KL), T_2 to be Newtonian theory (NT), and T to be Einstein's theory of general relativity (GR). What does it mean to say that NT can be 'approximately derived' from GR, KL can be 'approximately derived' from NT, and G can be 'approximately derived' from KL? Let us take the case (considered briefly in previous chapters) of approximately deriving KL from NT.

This can be done in three steps. *First,* NT is restricted to N body systems interacting by gravitation alone within some definite volume, no two bodies being closer than some given distance r. *Second,* keeping the mass of one object constant, we consider the paths followed by the other bodies as their masses tend to zero. According to NT, in the limit, these paths are precisely those specified by KL for planets. In this way we recover the *form* of KL from NT. *Third*, we reinterpret this 'derived' version of KL so that it is now taken to apply to systems like that of our solar system. (It is of course this *third* step of reinterpretation that introduces error: mutual gravitational attraction between planets, and between planets and the sun, ensure that the paths of planets, with masses greater than zero, must diverge, however slightly, from precise Keplerian orbits.)

The approximate derivation of G from KL is even simpler: only two steps are required. *First*, KL is restricted to systems for which the elliptical paths of planets take the form of circles; and *second*, this restricted version of KL is then reinterpreted to apply to all systems to which KL applies.

The approximate derivation of NT from GR is, by contrast, somewhat more complicated. *First*, GR is restricted to systems of bodies with mass travelling along geodesics. *Second*, we consider the paths of the bodies as distances between the bodies are increased, relative velocities tend to zero, and the curvature of space-time tends to the limiting case of flat space and time. *Third*, the resulting laws are reinterpreted to apply to bodies of any mass travelling at any relative distance and velocity. In this way, we arrive at an instrumentalistic mimic of NT which asserts (in effect): bodies move *as if* there is a force of gravitation such that $F = ma$ and $F = Gm_1m_2/d^2$. According to GR, there is no force of gravitation; there is, rather, space-time that is curved by the presence of mass, or energy-density. Massive bodies travel along geodesics in this curved space-time, a geodesic being the equivalent of a straight line in curved space. The force of gravitation has disappeared. Since GR makes no reference to force, it is not possible to derive from GR a version of NT that asserts that the force of gravitation exists. It is possible, however, to derive a version of NT that makes precisely the same predictions as NT, which is all that we require.[24]

Quite generally, we can say that T_{r-1} is 'approximately derivable' from T_r if and only if a theory empirically equivalent to T_{r-1} can be extracted from T_r by taking finitely many steps of the above type, involving (a) restricting the range of application of a theory, (b) allowing some combination of variables of a theory to tend to zero, and (c) reinterpreting a theory so that it applies to a wider range of phenomena.

It is important – for this proposed solution to the problem of verisimilitude – that the true theory of everything, T, is not presupposed to be unified or comprehensible. We want the idea that successive theories get closer and closer to the truth to be applicable in as wide a range of possible universes as possible. We don't want this notion to be applicable in only physically comprehensible universes. The demand that the successive theories can all be derived from the true theory of everything, T, does place constraints on T, but it does not mean that T must be unified or comprehensible. The first step to be taken in approximately deriving T_n from T is to restrict the range of application of T to a specific kind of system. It is quite possible for T to be sufficiently unified as far as this specific kind of system is concerned to approximately imply the more or less unified theory, T_n, and yet for T to be seriously disunified for all other phenomena.[25]

---

[24] For details see Schutz (1989: 205-208) or Rohrlich (1989).

[25] This corrects Maxwell (1998, p. 214) where I said that this solution to the problem of verisimilitude 'requires AOE to be presupposed'. On the contrary, it is important that AOE and physicalism are *not* presupposed.

This solution to the problem of verisimilitude can be exploited to solve the problem of what it means to say, of a succession of level 3 blueprints, B_0, B_2, ... B_n, that they get closer and closer to the true blueprint, B. Here, B is a blueprint of the true theory of everything, T. T implies B, but not *vice versa*. B, roughly, specifies the *kind* of entity precisely specified by T. B specifies symmetries which T must observe if it is to accord with B. Given B, T is the simplest theory there is compatible with B. Let T_0, T_2, ... T_n, be the simplest possible physical theories corresponding to B_0, B_2, ... B_n respectively. Then we may say that the blueprints, B_0, B_2, ... B_n get progressively closer and closer to B if and only if T_0, T_2, ... T_n get progressively closer and closer to T (in the way just explicated).

This proposed solution to the problem of what it means to say of a succession of blueprints that they are getting closer and closer to the true blueprint is likely to be misleading unless T is unified, and physicalism is true. Otherwise it would be possible for B_0, B_2, ... B_n to be progressively exemplifying physicalism more and more adequately, and at the same time getting closer and closer to B, even though B itself fails drastically to exemplify physicalism. To this extent (and to rule out this counter-intuitive possibility), physicalism and AOE do need, I think, here, to be presupposed.

It is worth noting just how ubiquitous 'approximate derivations' of the above type are in physics. When empirical predictions are derived from a physical theory approximations are very frequently made during the course of the derivation. Higher order terms in some expansion are set to zero; complicated expressions reduce to simple ones as a result of the neglect of effects deemed to be sufficiently minute. All such 'approximate derivations', to be found everywhere in physics, are logically invalid in just the same way in which the derivations of KL from NT, and NT from GR, are invalid. It is legitimate to regard such 'derivations' as valid insofar as it is an easy, if pedantic, matter to turn them into valid derivations by replacing the precise conclusion with an approximate one. None of this ought to seem problematic to anyone with any first hand familiarity with physics.

In one important respect, the above solution to the problem of verisimilitude is unsatisfactory. If a series of theories, T_o ... T_n progressively approaches the truth, T, then, as we move from T_o to T_n, more and more of the form of T will be captured by the successive theories. This justifies regarding T_o ... T_n as constituting improving theoretical knowledge of the nature of the basic dynamic structure of the universe. Nevertheless, T_o ... T_n are all false. We do not have progress in knowledge in the sense of a progressive capturing of more and more empirical truth.

I have remarked above, however that, even though successive accepted physical theories are all false, we nevertheless regard them as making progress because they 'apply, with ever increasing accuracy, to ever wider ranges of phenomena'. This certainly seems to be true of the sequence G, KL, NT, GR, and of other such sequences of physical theories (from classical to quantum physics). Can a bit more precision be given to this idea that T_2 is 'closer to the truth' than T_1 because the predictions of T_2 are more accurate than those of T_1, and apply to a wider range of phenomena? It can.

The important point to appreciate, of course, is that accepted physical theories, despite being false, nevertheless make a vast amount of true *approximate* predictions. It is these true approximate predictions of T_2 and T_1 that we need to compare. Furthermore, the theories we are interested in make predictions about the way physical systems or states of affairs evolve in time. It is the true approximate predictions, made by T_1 and T_2, about how systems evolve in time that we need to compare. This we can do as follows.

We consider predictions that the theories – T_1, T_2 and the true theory of everything, T – make of any isolated system of the form:

[Theory + state of the system at time t_1] $\rightarrow$ state of the system at time t_2.

What is derived, here – the specification of the state of the system at time t_2 – is the prediction of the theory. T_1, T_2, and the corresponding specifications of the state of the system at time t_1, and the predictions – the derived specifications of the states of the system at time t_2 – are all false. But these *false* specifications of the states of the system at times t_1 and t_2 imply true approximate specifications. In the case of Newtonian theory applied to the solar system, for example, such a true approximate specification would assert that each planet is located within such and such a region of space, having such and such a range of possible velocities (and would not give the precise position and velocity).

We can now declare that T_2 is closer to the truth than T_1 if:-

(a) The true approximate prediction of T_2 is more accurate, more precise, than the true approximate prediction of T_1;

(b) The true approximate specification of the initial state, at time t_1, associated with T_2, is at least as accurate, as precise, than the one associated with T_1.

(c) T_2 yields true approximate predictions of phenomena about which T_1 is silent (but T_1 makes no such predictions about which T_2 is silent).

If (a) to (c) hold, we can declare that T_2 makes more precise predictions than T_1 about more phenomena and is, in that sense, closer to the truth than T_1.

Why do we need clause (b)? Because we want to capture the idea that, if scientific progress is taking place, then increasingly accurate *predictions* are being made on the basis of specifications of initial states which are at least do not decrease in accuracy. In fact these specifications of initial states will, no doubt, increase in accuracy as the predictions increase in accuracy. In the limit, when the true theory of everything is reached, T provides the means for true, precise specifications of initial and final states of the system (even though such specifications could not be made in practice).

In spelling out this second account of what it means to say of two false physical theories that one is closer to the truth than the other, I have slurred over some details concerned, in the main, with what it means, precisely, to say that one specification of the state of a system is more accurate than another. As these details are rather fussy and unilluminating, I have relegated them to an appendix to be found at the end of this chapter.

Even if (a) to (c) hold for T_1 and T_2, and T_2 is closer to the truth than T_1 in the sense just explicated, it still might be the case that T_2 makes wildly false predictions about phenomena about which T_1 is silent. In other words, T_2 might be much more accurate than T_1 about phenomena to which both theories apply and might make true approximate predictions about additional phenomena about which T_1 says nothing, but might, in addition, make wildly false predictions about further phenomena about which T_1 is silent. Even though having much more truth content than T_1, T_2 would also have much more falsity content. If ever such circumstances arose in scientific practice, would we hold T_2 to be, nevertheless, an advance over T_1?

We might. T_2 might be accepted as a better theory than T_1, as long as it is restricted, in an *ad hoc* fashion, to phenomena for which it yields true approximate predictions. (Something like this is done when quantum theory is restricted in an *ad hoc* fashion so as not to apply to classical measuring instruments, for which it gives drastically false predictions.)

This second way of explicating what it means to say that T_2 is closer to the truth than T_1 would be characterized by Popper (1963, chapter 3) as ‘instrumentalistic’, in that it amounts to declaring that T_2 is a better *instrument* than T_1 for predicting phenomena (T_2 predicting more phenomena more accurately). This explication does not capture the idea that T_2 is closer to the truth than T_1 because T_2 is a more accurate characterization of the ultimate explanatory structure of the universe. But for that idea, we can turn to the first proposal, spelled out above. This second proposal is intended only to supplement the first. Taken together, the two proposals provide, I claim, an acceptable solution to the problem of verisimilitude as this arises in the context of theoretical physics.

6 The Problem of Induction

In chapter 9 I argued that AOE succeeds in solving the problem of induction, something which no version of standard empiricism can do. I have left unchanged what I said in that chapter of the first edition of 1984, since it is in my view essentially correct. But there have been developments since 1984, as I have already indicated, and some of these reveal the following inadequacies in the argument of chapter 9. To begin with, I argued (see page 225) that the best way we can improve knowledge in a partially comprehensible universe is to assume perfect comprehensibility, and fail to discover it. But, as I have already mentioned, situations might arise in which this is not correct. Again, the argument of chapter 9 fails to exploit properly the divergent, schematic accounts of the early evolution of natural science depicted on pages 232-5. Yet again, I argued that standard empiricism fails to solve the problem of verisimilitude, and the problems of simplicity, but I failed to explain how AOE solves these problems. Solutions to these problems (as we shall see) are required for the solution to the problem of induction. Again, the solution to the problem of induction sketched in chapter 9 makes essential use of the idea that AOE provides a framework for the *improvement* of false metaphysical assumptions – or blueprints – at the lowest level in the hierarchy of metaphysical assumptions, but no account is given there of what it *means* to say of two false metaphysical theses that one is an 'improvement' over the other. Even more serious, and closely connected to the previous point, I fail to explain how the aim-oriented empiricist solution to the problem of induction overcomes what may be called the circularity objection. This objection is that it is invalidly circular to appeal to some metaphysical thesis in order to justify the success of science, and then appeal to the success of science in order to justify acceptance of the metaphysical thesis. AOE seems, if anything, to intensify this circularity objection, in that it is a proud boast of the view that it captures and facilitates positive feedback between improving theoretical knowledge, and improving accepted metaphysical theses and associated methods. Publications of mine subsequent to 1984 have to a considerable extent put right these inadequacies in the argument of chapter 9: see Maxwell (1998, especially chs. 4 and 5; 2002b: 2004a; 2004b, chs. 1 and 2, and appendix, sections 2 and 6; 2004c; 2005b; 2005e; 2006a; 2007a; 2007b). In what follows I draw the various threads of these arguments together to form a line of argument as strong and succinct as possible, and one that makes amends for the deficiencies of the account of 1984.

In chapter 9 I pointed out that there are two parts to the problem of induction, namely:

1. *The Theoretical Problem*: What grounds are there for holding that theories accepted in accordance with the methods of science embody knowledge, granted that our aim is to improve our theoretical knowledge and understanding of (aspects of) the universe?

2. *The Practical Problem*: What grounds are there for holding that theories accepted in accordance with these methods embody knowledge sufficiently reliable and trustworthy to form a basis for action?

To these two, a third part should be added:-

3. *The Methodological Problem*: What precise methods ought science to employ in accepting and rejecting theories in the light of evidence?

Problems 1 and 2 differ because they presuppose different aims or purposes for which theories are accepted. If our aim is to improve theoretical knowledge and understanding of the universe, as in problem 1, it may be more important that a theory we accept is fruitful, in suggesting further fruitful lines of research for example, than that its empirical predictions are reliable. Just the reverse is the case if the aim is that presupposed by problem 2. It would seem, on the face of it, that we have no reason to suppose that a theory accepted for the purposes of theoretical knowledge and understanding would invariably be the same as that accepted for practical purposes, for the sake of *technological applications* and *action*. In scientific practice, rather remarkably, these two very different purposes do often lead to the acceptance of the same theory – although requirements that arise in connection with 2 may be more stringent than those that arise in connection with 1. (This latter point is not, perhaps, surprising. If a cosmological theory should turn out to be false, only some professional cosmologists may be disappointed; but if a theory employed in practical contexts, such as designing aeroplanes or developing drugs, should turn out to be false, people may well die. Naturally, in such *practical* contexts, we need to be more certain of truth, insofar as we can be, than in exclusively *theoretical* contexts.)

Whereas problems 1 and 2 both require that some kind of rationale or justification be provided for accepting theories in accordance with scientific method given one has such and such aim in mind, problem 3, by contrast, makes no such request for a rationale or justification. In order to solve problem 3, all one needs to do is specify the methods of science correctly.

If one looks at the history of attempts to solve the problem, one finds that most of the attention has been on problem 2. Problem 3 tends to be overlooked, the presumption being, it would seem, that that part of the problem can easily be solved. This attitude is a very serious mistake.

There is a vast literature on the problem of induction: see, for example (Kyburg, 1970; Swain, 1970; Watkins, 1984; Howson, 2002 – and references given therein). Most commentators hold that, despite this vast literature, the problem remains unsolved, and hardly any advance has been made towards its solution. Very few philosophers claim to have solved the problem, and when such claims are made, almost everyone else disagrees with them. Karl Popper is one of the few philosophers to have claimed to have solved the problem but, as he acknowledges himself, hardly anyone else agrees.[26] The problem has been around for over 250 years and has, it seems, stubbornly resisted endless attempts at solving it, so much so that, in recent years philosophers of science have grown weary of the problem, and no longer expect it to be solved, or indeed think it solvable. Given all this, why should my claim that AOE solves the problem be taken seriously for a moment?

It should be taken seriously because I can point to a *reason* why earlier attempts at solving the problem have failed. They have failed because they have presupposed (some version of) standard empiricism. Without even distinguishing problems 1 to 3, they have sought to justify acceptance of theories selected by methods prescribed by standard empiricism. *But this is to attempt to justify the unjustifiable.* Standard empiricism is, as we have seen, hopelessly unrigorous because it suppresses substantial metaphysical assumptions made by science which influence what theories are accepted and rejected in addition to empirical considerations. In short, all earlier attempts at solving problem 2 have failed because invalid answers to problem 3 have been carelessly presupposed. The crucial first step in solving the problem of induction is to give the correct solution to problem 3. The correct solution is AOE. Previous attempts have failed because they have got this crucial first step wrong. The solution to be proposed here deserves attention because this crucial first step is got right.

Traditionally, the problem of induction is viewed as the problem of how claims to theoretical knowledge – especially theoretical scientific knowledge – can be justified given that no theory can be verified however much evidence may be accumulated in its favour. What the above considerations indicate is that the problem should be viewed in a quite different way. We should rather view the persistence of the problem of induction as an indication that there is something seriously wrong with the whole conception of science that is being presupposed by the way the problem is formulated. And we should formulate the problem, rather, like this: How do we need to change our views about the nature of science so that the problem of how theories are established on the basis of evidence no longer arises? The problem of induction is important because it provides a test for the adequacy of views about science. In order to be acceptable, a view as to what the aims and methods of science ought to be must lead to the solution to the problem of induction. We might also say: the task is not to *justify* science; rather, it is to see how science must be *changed* so that the problem of induction no longer arises.

Some of this is implicit in Popper's attempted solution. His proposed solution involved changing dramatically our whole conception of science, in that it is recognized that scientific theories can be *falsified* but not *verified* (a point now quite widely accepted, but once heresy). Popper's proposal, quite exceptionally, does make an important contribution towards solving the problem, precisely because it involves changing our view about science in a way that it needs to change, if the problem is to be solved. But Popper does not go far enough in this respect. Ultimately, his proposed solution fails. For, despite its revolutionary aspect, in one respect Popper's falsificationist conception of science is thoroughly conventional, in that it is a version of

[26] 'I think I have solved . . . the problem of induction. . . However, few philosophers would support the thesis that I have solved the problem' (Popper, 1972, p. 1).

the untenable standard empiricism. Popper's proposed solution fails because it does not even solve problem 3 above – the problem, merely, of specifying the methods of science.[27]

In order to solve the problem, we need to take matters one step further than Popper's falsificationism: we need to adopt AOE.

But does this suffice to solve problem 3? Even if AOE, with an appropriate choice of metaphysical blueprint at level 3 in the hierarchy of theses, solves the problem of specifying the methods of theoretical physics, does this suffice to solve the methods of natural science as a whole? I have three points to make in response to this question.

First, the problem of induction only arises in a pristine form in connection with theoretical physics. This is because all other branches of natural science presuppose relevant results of some other, explanatorily more fundamental natural science. Put crudely, biology presupposes chemistry which, in turn, presupposes physics. As a result, two kinds of consideration uncontroversially govern choice of theory in biology, let us say: empirical considerations from 'below', and relevant results of explanatorily more fundamental sciences, such as chemistry and physics, from 'above'. Thus, within biology – as should be clear even to a standard empiricist – evidence alone does not decide what biological theories are accepted and rejected: relevant parts of chemistry and physics play a role as well.[28] In order to confront the problem of induction in its naked, pristine form, we need to concentrate our attention on theoretical physics, since this is the only branch of natural science which does not have a more fundamental branch to presuppose. What this means, in turn, is that, as far as tackling the problem of induction is concerned, it suffices that AOE depicts the methods of theoretical physics; it does not need to specify the methods of natural science as a whole. (And given the explanatorily fundamental role of theoretical physics in the natural sciences, if the problem of induction can be solved for the former, this will suffice to solve it for the latter as well.)

Second, because theoretical physics is explanatorily fundamental in natural science, there is an important sense in which AOE, in being applicable to theoretical physics, as depicted above, in figure 12 let us say, is applicable to the whole of natural science.

But, third, there is a much more detailed and accurate way in which the general idea of AOE is applicable to the diverse methods of *all* the diverse branches of natural science. Different branches of natural science have different aims, and make different presuppositions, even if they are all inter-related in the way just indicated. Thus, a major aim of biology is to discover what survival value features of living things have – an aim that presupposes Darwin's theory of evolution. Such an aim and presupposition does not arise within the context of physics,

---

[27] That Popper espouses standard empiricism is clear from his advocacy of his demarcation criterion: a theory, in order to be scientific, must be falsifiable (which renders unfalsifiable metaphysical statements unscientific). And, as I remarked in chapter 9, Popper defends the doctrine explicitly in defending '*the principle of empiricism,* which asserts that in science, only observation and experiment may decide upon the *acceptance or rejection* of scientific statements, including laws and theories' (1963, p. 54). It might be thought that Popper's espousal of 'metaphysical research programmes' in his later publications represents a change of attitude towards the scientific status of metaphysics, but it does not. Although 'indispensable for science' such research programmes are, nevertheless, according to Popper 'more of the nature of myths, or of dreams, than of science' Popper (1982, p. 165). For a discussion of this point see Maxwell (2007a).

[28] Very occasionally, when a biological theory clashes with accepted chemistry or physics, it may happen that the biology is found to be correct and it is the chemistry or physics that needs to be revised. This happened when Kelvin, employing then current knowledge in physics, calculated that the earth could not have existed long enough for evolution to have occurred in the way described by Darwin, because if it had it would have cooled long ago to a temperature far below its present value. It turned out, subsequently, that Kelvin's calculations were incorrect because they ignored the heat generated by naturally occurring radioactivity associated with some of the constituents of the earth. Biology was right, physics was wrong. But this way of resolving such a clash is very infrequent; the norm is for it to be resolved the other way round. All this is, of course, all but demanded by AOE, as I made clear when expounding the view.

cosmology or inorganic chemistry. Again, geology has the historical aim of discovering how various features of the earth's surface were created in the past: theoretical physics as it has been conducted up to the present does not have any such historical aim.[29] Specific aims and presuppositions of these types, made by specific branches of natural science, lead to the adoption of specific methods, designed to help achieve the specific aims, and corresponding to the specific presuppositions. These diverse methods of the diverse branches of the natural sciences, corresponding to diverse aims and presuppositions, can be accurately captured by the general idea of AOE. All we need to do is add one or more levels below level 3 of figure 12 to take into account specific presuppositions and aims of the specific natural science we are interested in, and associate relevant additional methods with these additional assumptions. In this way we can accurately capture the specific aims and methods of as wide a range of specific scientific specialities as we please, even to the extent of capturing accurately aims and methods of highly restricted scientific specialities. We can also, in this way, capture the *evolving* aims and methods of a scientific speciality by specifying the evolving specific presuppositions of that speciality. AOE is sufficiently flexible to capture both what is common to all of natural science and at the same time what is specific to diverse branches of natural science, however specialized they may be, and however much they may evolve with time.

Granted that AOE solves problem 3, it remains to be shown that it solves problems 1 and 2 as well.

In tackling problems 1 and 2, it is important not to formulate them in a way which renders them insoluble. Thus it is no good formulating the problem of induction as the problem of how physical theories can be verified by evidence, since such theories cannot be so verified. Nor should the problem be formulated as 'How can we have some grounds for holding that an empirically successful theory is true?', since the historical record tells us that even the most empirically successful physical theories turn out eventually to be false, and AOE tells us that all dynamical physical theories, not generalizable to all phenomena, are false. The above *theoretical* and *practical* problems of induction need to be reformulated slightly, along the following lines, to make this point explicit:-

1. *The Theoretical Problem*: What grounds are there for holding that a physical theory, accepted in accordance with the methods of science, embodies knowledge in the sense that, even though it may be false, it is a step towards the truth, granted that our aim is to improve our theoretical knowledge and understanding of (aspects of) the universe?

2. *The Practical Problem*: What grounds are there for holding that a physical theory, accepted in accordance with the methods of science, embodies knowledge in the sense that it will continue to yield true empirical predictions in standard regions of application, to standard degrees of accuracy, in a way that is sufficiently reliable and trustworthy to form a basis for action?

In tackling the problem of induction, it is important to appreciate just how strong the reasons are for holding that scientific knowledge makes presuppositions that are metaphysical and cosmological. There is the argument already encountered: in persistently failing even to consider endlessly many empirically more successful, disunified rivals to accepted physical theories, physics makes a persistent metaphysical and cosmological assumption to the effect that the universe is such that no disunified theory is true. But even more striking, our most humble, prosaic, common sense claims to knowledge of things in our immediate environment make metaphysical and cosmological presuppositions. The proposition 'This chair on which I now sit will continue to exist and support me for the next 30 seconds' implies 'No cosmic convulsion is

---

[29] The idea of spontaneous symmetry breaking, if taken literally as an historical event which, in a sense, transformed manifest basic laws of physics, does give to theoretical physics a kind of historical aspect. This is apparent, too, in the emphasis given in theoretical physics to the study of conditions at or immediately after the big bang. If ever cosmic physicalism becomes the accepted blueprint, physics would acquire an even stronger historical character.

now occurring far away in the universe which will spread with near infinite speed to engulf and destroy the earth and everything on it, including me and my chair, in under 30 seconds time'. If the first proposition is true, then the second must be true as well, which means the first implies the second. Thus, if I know the first proposition, I at least implicitly know the second as well. Even our most trivial, common sense, observational claims to factual knowledge, which include knowledge of matters a mere second or two into the future, presuppose knowledge of metaphysical theses about the entire universe. If we deny that we have such cosmological knowledge we are obliged to deny, also, that we have trivial factual common sense knowledge of our immediate surroundings. Bereft of cosmological knowledge, we have scarcely any factual knowledge at all. These considerations can perhaps be taken in two ways: as establishing either extreme scepticism (we know nothing), or that we need to adopt a more conjectural conception of knowledge, one which is such that 'knowledge' of the entire cosmos does indeed become possible. But in any case, if even our most humble, limited, common sense items of particular, factual knowledge make presuppositions about the entire cosmos, it ought to occasion no surprise that our theoretical scientific knowledge, so vastly more burdened with empirical content, so much more precise and wide ranging in predictive power, makes such cosmological presuppositions as well.

It is, in a way, very odd that AOE has not been seen as the obvious view to adopt as the first step towards solving the problem of induction. Everyone agrees that evidence underdetermines theory. And yet, in practice, most of the time, very few theories contend for acceptance. Almost all of the infinity of rival theories that are compatible with the available evidence that always exist, never in scientific practice make their presence felt. It is entirely reasonable to conclude that this is because hidden, unacknowledged assumptions made by scientists, in addition to the evidence, exclude these infinitely many rivals. The obvious first step to take, in tackling the problem of induction, one would think, is to make these hidden, unacknowledged assumptions explicit. It is just this that one does if one is confronted by an invalid inference from correct premises to a correct conclusion: make explicit additional implicit premises which, once acknowledged, turn the invalid inference into a valid one. Why not take the analogous step in connection with scientific 'inference' from evidence to theory (even if in this case, strictly speaking, no valid inference results)? It is just this which AOE does. It makes explicit implicit metaphysical assumptions concerning the knowability and comprehensibility of the universe which have the effect, when added to evidence, of tightly restricting theories that receive, and deserve, scientific attention (disunified rivals that are compatible with the evidence being excluded). There is the problem, of course, that the metaphysical thesis most effective in so restricting theories worthy of consideration – the metaphysical blueprint at level 3 – is most likely to be false: but the hierarchical framework of AOE is designed to help us put that right: it is designed to help us critically assess, and improve, this probably false thesis. Why has not AOE been adopted as the first step towards solving the problem of induction long ago? A version of the doctrine has been in the literature, after all, since 1974 (see Maxwell, 1974).

One reason may have to do with the demise of so-called 'rationalism'. Once upon a time some philosophers, the 'rationalists' – notably Descartes, Spinoza and Leibniz – held that some substantial theses about the nature of the universe could be established by reason alone. Some philosophers today may think that appealing to metaphysics in order to solve the problem of induction can only be successful if rationalism is correct, and the relevant metaphysical theses can be established by reason alone. Evidence cannot establish the truth of metaphysical theses so, if anything, it must be reason that one has to call upon to do the job. But rationalism is, nowadays, severely discredited. How could reason alone establish the truth of substantial theses about the universe? With the demise of rationalism – so the thought runs – comes the demise of the idea of appealing to metaphysical theses in order to solve the problem of induction.

But this objection collapses the moment one adopts a quasi-Popperian conception of knowledge, and acknowledges that all our knowledge is conjectural in character, it being just as impossible to justify the truth of scientific theories as scientific metaphysics, whether by an appeal to reason or evidence.

A more serious objection has to do with the apparent invalid circularity involved in appealing to metaphysical theses in order to solve the problem of induction – something that has already been alluded to. Such an approach would seem to involve justifying the success of science by an appeal to metaphysical principles, which are in turn justified by the success of science. But, as Bas van Fraassen has put it in a striking phrase (which I have quoted on other occasions), 'From Gravesande's axiom of the uniformity of nature in 1717 to Russell's postulates of human knowledge in 1948, this has been a mug's game' (van Fraassen, 1985, pp. 259-60). How does AOE escape this charge?

The first point to note is that, quite independent of any claim to solve the problem of induction, a conception of science – call it presuppositionism – which acknowledges that science makes a persistent metaphysical assumption concerning unity is *more rigorous* than any standard empiricist conception which denies this. Intellectual rigour demands that assumptions (or conjectures) that are substantial, influential, problematic and implicit need to be made explicit. In persistently accepting simple, unified theories in preference to empirically more successful disunified theories, science thereby does make an (implicit or explicit) metaphysical assumption. Rigour demands that this assumption be acknowledged explicitly. Presuppositionism does this, but standard empiricism does not. This means that presuppositionism is more intellectually rigorous than any version of standard empiricism.

Attempts at solving the problem of induction, if they are to have any hope of success, must begin with the most *rigorous* conception of science available. It is clearly hopeless trying to justify the unrigorous, and therefore unjustifiable. This means that the actual situation is the exact opposite of what van Fraassen declares. The only hope we have of solving the problem of induction is to begin with presuppositionism, unless something better turns up; all views which reject presuppositionism, being inherently unrigorous, are doomed to failure.

A view that is even more rigorous than presuppositionism is available, namely AOE. This is more rigorous because it does not just rigidly and dogmatically accept some metaphysical thesis of unity, but instead accepts a hierarchy of theses, thus facilitating the critical assessment, and revision, of the more substantial theses in this hierarchy, those most likely to be false, in the light of the empirical success and failure of associated research programmes, and other considerations. AOE is more rigorous than presuppositionism because it focuses attention on those assumptions most likely to be false, and most likely to need revision and improvement, at the same time providing a relatively unproblematic framework within which such revision and improvement may proceed.

Granted all this, the conclusion is clear: attempts to solve the problem of induction must begin with AOE; all other approaches are doomed to failure.

But this does not solve the circularity problem. Indeed, it may even be judged to make this problem worse. For AOE has something like circularity built into it quite explicitly; it is even upheld as its greatest virtue and triumph. The whole point of the view, after all, as I have just emphasized, is to facilitate the critical assessment of theses low down in the hierarchy in the light of the empirical success and failure of science. Successful theorizing may lead to a revision of level 3 blueprint ideas; such ideas constrain what is accepted at the level of testable theory (level 2). How, then, does AOE overcome the circularity objection?

In order to solve the problem, I shall argue, we need to see science as accepting a metaphysical thesis which, if true, renders the circularity of AOE legitimate, the reasons for accepting this thesis making no appeal to the success of science whatsoever.

It may be asked how AOE can in practice work at all if physical theories are both constrained by the current level 3 metaphysical thesis, and at the same time are able to modify this level 3 thesis. How can choice of theory both be influenced by, and influence, choice of level 3 thesis? The answer is that, as one goes up the hierarchy of levels of AOE, so the corresponding theses become more and more resistant to modification. Level 2 theories are only acceptable if sufficiently empirically successful, and sufficiently in accord with the best available thesis at level 3. But if attempts to develop theories in accordance with this level 3 thesis persistently fail, and a theory emerges that accords with the thesis at level 4 but clashes with the current level 3 thesis, then this thesis will be modified to accord with the new theory. Far greater persistent empirical failure would be required before this would legitimately lead to the rejection of the level 4 thesis of physicalism, and the adoption of some rival comprehensibility thesis, especially if this differed substantially from physicalism. Such a development would be dramatic and revolutionary indeed, for it would involve changing the whole nature of natural science. An intellectual earthquake would be needed before the level 5 thesis of comprehensibility deserved to be modified. It is the increasing resistance to modification as one goes up the hierarchy that makes it possible for theses accepted at one level both to be influenced by, and to influence, theses accepted at the next level up. The increasing resistance to modification that arises as one goes up the hierarchy is justified by the point that theses become increasingly contentless as one goes up the hierarchy, thus being increasingly likely to be true. It is also justified by the point that, as one goes up the hierarchy, theses become increasingly close to being such that their truth is required for science, or the pursuit of knowledge, to be possible at all.

As I have already pointed out above, a similar two-way influence takes place between theory and evidence. If a theory clashes with evidence then, in general, the theory will be rejected. This will occur especially if the clashing evidence consists of a number of different kinds of experimental result, each kind of experiment being repeated, and being subject to expert critical scrutiny. But the opposite also takes place in science. A clash between theory and an experimental result may lead to the experimental result being rejected. Many experiments are very difficult to perform. It may take weeks before the apparatus involved works properly. Early experimental results that clash with established theory are regarded as indications that the apparatus is not working properly, and are rejected. In short, the two-way influence between theory and metaphysics, demanded by AOE, also takes place between theory and evidence, as every scientist would acknowledge.

Here, now, are nine further preliminary remarks concerning my proposed solution to the problem of induction, and the circularity problem in particular.

First, within the framework of AOE, no attempt is made to justify the truth of a physical theory by an appeal to a blueprint, the truth of the blueprint in turn being justified by an appeal to the empirical success of the theory. Physical theories, whatever their empirical success, and metaphysical assumptions, whatever their position in the hierarchy, and however fruitful in helping to generate empirical progress, remain conjectures. All our knowledge is presumed to be conjectural in character, even though we may conjecture that some parts are rather more conjectural than others.

Second, even though AOE provides no arguments for the *truth* of theses in the hierarchy, it does provide arguments for *accepting* these theses, granted that the aim of science is to acquire knowledge of the truth, insofar as this is possible. It is important to recognize just how different these two things are. To illustrate the point, Popper (1959) sets out to justify accepting that theory which has the greatest empirical content (other things being equal), even though that is the theory which is most likely to be false. He does so on the grounds that it is the theory with the greatest empirical content which we can most readily discover to be false (if it is false), discovering falsehood in this way being the means by which science makes progress. This Popperian justification for *accepting* a theory is diametrically opposed to any attempted justification of the *truth* of the theory, since the justification involves accepting that theory most likely to be *false*!

Third, and backing up the two points just made, it is important to remember that accepted physical theories (at level 2) and the best available blueprint (at level 3) will be *incompatible* with one another as long as no candidate theory of everything has been accepted (as at present). The circularity inherent in AOE can hardly be interpreted as any kind of attempt to justify the truth of the accepted blueprint by an appeal to the empirical success of accepted theories, in turn justified by an appeal to the blueprint, if these two are *incompatible* with one another.

Fourth, the rationale behind making explicit metaphysical theses implicit in the persistent scientific acceptance of unified theories is not to justify the truth of these theses. Quite the contrary, it is to make these theses available for sustained critical scrutiny (in the hope that they can be improved).

Fifth, it is vital to remember that it is not just theoretical knowledge in physics that presupposes and requires some metaphysical and cosmological knowledge. As we have seen, even our most trivial items of common sense knowledge about our immediate environment (that include knowledge of things mere seconds into the future) contain, implicitly, some knowledge about the entire cosmos. It hardly overstates the situation to say that we have no factual knowledge of anything if we do not have some knowledge, even if meagre, of everything. Failing to acknowledge the metaphysical, cosmological presuppositions of science cannot be anything other than intellectually dishonest. As I have stressed, merely acknowledging such presuppositions as an explicit part of conjectural scientific knowledge in itself enhances the intellectual rigour of science.

Sixth, it might seem, despite points one to four above, that any attempt to solve the problem of induction by appealing to some metaphysical or cosmological thesis *must* provide some grounds for holding that this thesis is true. But this seems hopeless: neither an appeal to evidence, nor an appeal to reason, could conceivably, it would seem, do the job. But what this demand neglects is that it presupposes an untenable, standard empiricist conception of science. It presupposes that science got going when it dissociated itself from metaphysics and concentrated on assessing claims to knowledge empirically. This view is hopeless, both as an historical account and as a prescription as to what ought to go on: see, for example, figure 9 of chapter 9 and associated text. As point five above makes clear, science cannot get going by dissociating itself from metaphysical presuppositions: instead, science gets going and proceeds by developing and preferring those metaphysical theses which seem best to promote progress in knowledge. The proper task, in other words, is not to provide arguments for the truth of some metaphysical thesis, but to provide arguments for the claim that one such thesis helps promote the growth of scientific knowledge better than rival theses. (But it is precisely arguments of this kind which introduce apparent invalid circularity.)

Seventh, the reasons that will be given for accepting metaphysical theses at the various levels of AOE are versions of the following: this thesis is the best to adopt, at its level of generality, given that the aim is to improve knowledge of the truth, insofar as this is possible. There are three reasons of this type, namely: this thesis (1) needs to be true for the pursuit of knowledge to meet with any success at all; (2) holds out the greatest hope for progress in knowledge, if true; and (3) is in fact associated with progress in scientific knowledge – or what seems to be progress in scientific knowledge – in that it is the blueprint of the most empirically successful research programme.

Eighth, it is clear from point six, that there can be no knock-down, definitive solution to the problem of induction. This is because (a) one cannot list all possible theses that might be considered at each level, and (b) as science progresses, the thesis at level 3 is almost bound to change, and even theses higher up in the hierarchy may change.

Ninth, it is the *methodological* character of AOE that creates the circularity problem. If the methods of AOE, evolving in the light of the empirical success and failure of rival research programmes, had no more than *heuristic* force, suggesting merely that one kind of hypothesis might be sought rather than another, there would be no serious problem. What creates the problem is that these evolving methods of AOE have what may be termed *methodological* force: they influence (but in a fallible and revisable way) what theories are to be accepted and rejected, along with empirical considerations.

Quite enough preliminaries! I now sketch how, in my view, AOE solves the problem of induction, taking the version of AOE depicted in figure 12 above, beginning at the top of the hierarchy, and working down to accepted physical theories at level 2. What follows is only a sketch; a fuller account would be couched in terms of the version of AOE depicted in figure 11. For an account along these lines see Maxwell (1998, ch. 5).

Level 7: *Partial Knowability*. The universe is such that we possess and can acquire some knowledge of our immediate environment as a basis for action. If this is false, we cannot acquire knowledge whatever we assume. Accepting this thesis as an item of scientific knowledge can only help, and cannot sabotage, the pursuit of knowledge whatever the universe is like. We are justified in accepting this thesis as a permanent item of scientific knowledge even though we have no grounds for holding it to be true.[30] It should be noted that this is a thesis about the entire cosmos, and not just about our local environment.

Level 6: *Meta-Knowability*. The somewhat more precise thesis that the universe is 'meta-knowable', which means that the universe is such that there is some rationally discoverable assumption about it which leads to improved methods for the improvement of knowledge.

As I have already acknowledged, the notion of 'rationally discoverable' is problematic. As the phrase is used here, no thesis about the universe is rationally discoverable if it is grossly *ad hoc*, like a theory that is disunified in a type 8 to 6 way, and the *ad hoc* phenomena, postulated by the thesis, lie beyond our experience. Any such thesis is one of infinitely many rivals, all equally arbitrary, there being no rationale to prefer the given thesis.

Meta-knowability bring us to my proposed solution to the circularity problem. Permitting metaphysical assumptions to influence what theories are accepted, and at the same time permitting the empirical success of theories to influence what metaphysical assumptions are accepted, may (if carried out properly), *in certain sorts of universe*, lead to genuine progress in knowledge. Meta-knowability is to be interpreted as asserting that *this is just such a universe*. And furthermore, crucially, reasons for accepting meta-knowability make no appeal to the success of science. In this way, meta-knowability legitimises the potentially invalid circularity of generalized AOE (GAOE), and of AOE.

Relative to an existing body of knowledge and methods for the acquisition of new knowledge, possible universes can be divided up, roughly, into three categories: (i) those which are such that the meta-methodology of GAOE or AOE can meet with no success, not even apparent success, in the sense that new metaphysical ideas and associated methods for the improvement of knowledge cannot be put into practice so that success (or at least apparent success) is achieved; (ii) those which are such that AOE appears to be successful for a time, but this success is illusory, this being impossible to discover during the period of illusory success; and (iii) those which are such that GAOE, and even AOE, can meet with genuine success. Meta-knowability asserts that our universe is a type (i) or (iii) universe; it rules out universes of type (ii).

Meta-knowability asserts, in short, that the universe is such that AOE can meet with success and will not lead us astray in a way in which we cannot hope to discover by normal methods of scientific inquiry (as would be the case in a type (ii) universe). If we have good grounds for accepting meta-knowability as a part of scientific knowledge – grounds which do not appeal to the success of science – then we have good grounds for adopting and implementing AOE (from levels 5 to 2). Meta-knowability, if true, does not guarantee that AOE will be successful. Instead it guarantees that AOE will not meet with illusory success, the illusory character of this apparent success being such that it could not have been discovered by any means whatsoever before some date is reached.

[30] Sooner or later, this thesis will be falsified. Our current scientific knowledge tells us that, one day, the sun will become a red giant and engulf the earth; acquisition of knowledge, and life itself, will no longer be possible on earth.

If AOE lacks meta-knowability, its circular procedure, interpreted as one designed to procure knowledge to the extent that this is possible, becomes dramatically invalid, as the following consideration reveals. Corresponding to the succession of accepted fundamental physical theories developed from Newton down to today, there is a succession of severely disunified rivals which postulate that gravitation becomes a repulsive force from the beginning of 2150, let us say. Corresponding to these disunified theories there is a hierarchy of disunified versions of physicalism, all of which assert that there is an abrupt change in the laws of nature at 2150. The disunified theories, just as empirically successful as the theories we accept, render the disunified versions of physicalism just as scientifically fruitful as unified versions of physicalism are rendered by the unified theories we actually accept. The circularity inherent in AOE is invalid because it can be employed so as to lead to the adoption of *disunified* theories and metaphysical theses just as legitimately as it can be employed to lead to the adoption of *unified* theories and metaphysical theses. This is the case, at least, if AOE is bereft of meta-knowability. But if we have good reasons to accept meta-knowability as a part of scientific knowledge, then we have good reasons to reject *disunified* versions of physicalism: these lack the crucial requirement of rational discoverability. If we have good reasons to accept meta-knowability as an item of scientific knowledge, and these reasons make no appeal to the success of science, then the circularity inherent in AOE ceases to be invalid: meta-knowability asserts that the universe is such that empirical success achieved by implementing AOE will not be illusory *in a way which could not discovered by any means before a certain date*.

But what reasons have we for accepting meta-knowability that make no appeal to the success of science? One argument is simply this. As the pursuit of knowledge, and science, have developed over the millennia, GAOE and AOE have in fact been put into practice. Metaphysical presuppositions have been revised in the light of which seem to meet with the greatest empirical success – from myths, religious views, the ideas of the Presocratics, the ideas of Plato, Aristotle, Galileo, Boyle, Newton, and Boscovich, to field ideas, ideas associated with quantum theory and string theory, to physicalism. Even empirical methods have been revised in the light of metaphysical revisions. For example, given Aristotelian metaphysics, with its denial that precise mathematical laws govern natural phenomena, there is little point in performing precise experiments to decide between rival theories. This changes dramatically once Galileo's metaphysics is accepted, according to which 'the book of nature is written in the language of mathematics' (an early statement of physicalism). Suddenly, it becomes highly pertinent to perform precise experiments, of the kind performed by Galileo involving, for example, rolling balls down inclined planes, to try to determine what precise mathematical law governs the fall of bodies near the surface of the earth. Granted, then, that GAOE and AOE have been put into practice over the millennia and right up to the present, science is more rigorous if the metaphysical assumption, implicit in this practice, is made explicit. This is the case even if this explicit thesis remains a conjecture *with no other reasons being given for its acceptance* over and above that it is implicit in scientific practice. No other justification for accepting meta-knowability explicitly in the hierarchy of GAOE and AOE is required (in order to render science more rigorous).[31]

Can anything more be said? I think it can. We can argue that, as a result of accepting meta-knowability, we may have much to gain and little to lose. In accepting meta-knowability we decide, in effect, that it is worthwhile to try to improve knowledge about how to improve knowledge. We take seriously the possibility that the universe is such that we can discover something rather general about its nature which will enable us to *improve* our methods for

---

[31] No circularity is involved here, because no attempt is made to justify acceptance of meta-knowability by an appeal to the success of science. The argument is that we should make meta-knowability explicit even if AOE science is entirely unsuccessful. We should make explicit metaphysical assumptions implicit in our methodological practice whether this practice meets with success or not.

improving knowledge. Not only do we hope to learn about the world; we hope to learn about how to learn about the world, and we are prepared to implement a meta-methodology (GAOE) which capitalizes on this possibility should it turn out to be actual. To fail to try to *improve* methods for improving knowledge on the grounds that apparent success might prove to be illusory is surely to proceed in a cripplingly over-cautious fashion. Any attempt at improving knowledge may unexpectedly fail, including the attempt to improve methods for improving knowledge. But eschewing the attempt to learn because it may fail cannot be sound: such an excuse for not making the attempt always exists. In accepting meta-knowability we do not assume, note, that the universe *is* such that GAOE will meet with success. We assume, merely, that it is such that *if* GAOE or AOE appears to meet with empirical success, this success will not be illusory in a way which could not have been discovered prior to the illusory character of the success becoming apparent. But this is an entirely sensible assumption to make. Nothing is to be gained from foregoing the attempt to acquire knowledge because of the fear that future, inherently unpredictable changes in the laws of nature may occur which render knowledge acquired obsolete.

Neither partial knowability nor meta-knowability excludes the possibility that such inherently unpredictable events occur. Even though we accept these theses, we might, nevertheless, still discover and accept that unpredictable changes in the laws of nature do occur (if they did occur). We might live, or come to live, in a world in which inherently inexplicable, unpredictable events occur quite often. Objects vanish, or abruptly appear; substances abruptly change their properties; bridges collapse, mountains vanish, houses turn into elephants, trees become daffodils. People die as a result, but life might nevertheless go on, and it might be possible, not just to improve knowledge, but to improve knowledge about how to improve knowledge. Meta-knowability asserts that, if we have had no such experience of them, such events do not occur. We are justified in ignoring the possibility that such events may occur in future in both science and life because, if they occur in the future nothing, in the nature of things, can be done to anticipate their occurrence, or evade the harm they may cause. It is this which provides the grounds for accepting meta-knowability as an item of scientific knowledge.

Hume, famously, argued that what exists at one moment cannot necessarily determine what exists at the next moment. If he is right, we may well feel that anything may happen at any moment – just because there can be nothing in existence now to *determine* (perhaps probabilistically) what will exist next. However, elsewhere I have shown that Hume is wrong, and it *is* possible that what exists at one instant necessarily determines what exists at the next moment (Maxwell, 1968; 1998, pp. 141-155). Since this is possible, it is, in my view, madness not to assume that what exists now *does* necessarily determine what exists next. Recognizing that Hume's arguments, here, are invalid is bound to affect ideas about how likely it is that utterly inexplicable, inherently unpredictable events will occur, as long as we do not seem to have had any experience of them.

Accepting meta-knowability, then, puts on record our decision to try to learn how to learn – to try to improve assumptions and associated methods in the light of improving knowledge and understanding, in the light of which seem best to promote empirical progress. This goes on, after all, in a thoroughly acknowledged and uncontroversial manner at the empirical level. New knowledge can give rise to new technology, new instruments and experimental techniques – from the telescope and microscope to the cyclotron – which are in turn employed to help create new knowledge. At the empirical level, uncontroversially and fruitfully, there is a kind of circular, positive feedback between improving knowledge and improving observational and experimental methods for the further improvement of knowledge. Something analogous has long gone on too, implicitly, in scientific practice, at the theoretical level. Science would be more rigorous, and even more successful, if this latter was explicitly recognized and acknowledged.

I have argued that we are justified in ignoring the possibility that apparent success achieved as a result of implementing GAOE might turn out to be illusory in a way we could not possibly have discovered. Are we justified, however, in ignoring illusory apparent success of a less fiendish kind – apparent success which we could have discovered to have been illusory, if we had tried harder? Do not GAOE and AOE always carry the danger that they will actively create the illusion of success – metaphysical assumptions and methods being chosen to promote the illusion of success in the pursuit of knowledge?

GAOE and AOE are better equipped to defeat this danger than any other rival methodology for science.

Consider the best that any version of standard empiricism can do to defeat illusory success. First, accepted observational and experimental results can be subjected to sustained critical scrutiny. Experiments can be repeated in different laboratories by different scientists; and essentially the same experiment can be performed in different ways in an attempt to eliminate errors associated with one type of experiment. Second, accepted laws and theories can be severely tested, a variety of consequences being put to the test. Third, rival laws and theories can be developed in order to disclose crucial experiments which may falsify the accepted laws and theories, and which would not otherwise have been thought of: these crucial experiments can then be performed. These three standard empiricist procedures for detecting illusory empirical success are all important.

But AOE science can go further. In addition, it can subject the current best blueprint, and associated methodological principles, to sustained critical scrutiny. It can actively seek to develop improved versions of this blueprint. It can even criticize and develop alternatives to metaphysical theses higher up in the hierarchy, at level 4, and even higher (see figure 11). AOE comes with a framework that facilitates sustained critical scrutiny of current aims and methods, assumptions and methods; it provides meta-methodological machinery for the development of alternative possible aims and methods - alternative vantage points from which any illusory success of current aims and methods may be much more readily detected. Basic blueprint assumptions of a science do much to determine what kind of evidence is acceptable within that science. A change of blueprint may lead to a change in what constitutes acceptable evidence – a point illustrated above in connection with the transition from Aristotle to Galileo. There is always the danger that a science seems to make great empirical success and fails to discover that this success is illusory because the evidence required to reveal this is declared illegitimate by the accepted blueprint. Thus the demand within physics that experimental result be repeatable prevents physics from discovering miracles – unique, unrepeated events – on empirical grounds. In order to discover the illusory character of such apparent empirical success it may be necessary to view matters from the standpoint of a modified blueprint, with modified standards for what constitutes an acceptable empirical result. AOE encourages the development of such modified blueprints, whereas standard empiricism does not even recognize the need for them. (Any view which specifies a fixed metaphysical assumption for science, on one level, is no better than standard empiricism in the respect just discussed.)

That AOE is better equipped to discover illusory empirical success than rival views provides a decisive rebuttal of the charge that there is an inherently invalid circularity in the manner in which AOE adjusts assumptions and methods in the light of empirical success and failure. On the contrary, AOE science is in a better position to detect such illusory success than science conducted in accordance with any rival view. AOE can modify its aims and methods, its assumptions and methods, in the direction of those which seem to produce the greatest empirical success – thus implementing something like positive feedback (and circularity). At the same time, AOE provides means for discovering when such apparent success is illusory in a way that is better, more effective, than any rival view.

This concludes my discussion of the solution to the circularity problem, and the reasons for accepting meta-knowability in preference to any rival thesis at this level.

Level 5: *Comprehensibility*. The thesis that the universe is comprehensible in some way or other, there being *something*, or an aspect of something (kind of physical entity, God, society of gods, cosmic purpose, cosmic programme or whatever) that runs through all phenomena, and in terms of which all phenomena can, in principle, be explained and understood. Almost all (perhaps all) cultures possess a myth, cosmology or religious view taken to explain natural phenomena, presupposed by attempts to improve knowledge. Almost all of these are personalistic, animistic or purposive in character: natural phenomena are explained in terms of the actions of gods, spirits, God, or purposes. Acceptance of some version of comprehensibility is often combined, however, with a clause that places strict limits on knowability (this clause being required, perhaps, to protect the thesis against criticism, and to explain away the lack of success of the view in promoting acquisition of knowledge). Thus God is said to be mysterious and unknowable. That the universe is held to be (more or less) comprehensible in almost all cultures is not, however, a good reason to hold it to be worthy of acceptance. Grounds for this stem from the thesis one rung down in the ladder of theses at:-

Level 4: *Physicalism(1,1)*. The thesis that the universe is physically comprehensible, everything being made up of just one kind of physical entity (or perhaps just one entity), all change and diversity being in principle explicable in terms of this one kind of entity. This thesis asserts that the universe is such that some yet-to-be-discovered physical 'theory of everything' (in the current jargon of theoretical physicists) is unified in a type 1 way, and true.

Granted meta-knowability, we are justified in accepting that thesis, other things being equal, which holds out the greatest promise, if true, for progress in empircal knowledge. Physicalism(1,1) satisfies this requirement better than any rival thesis at this level, in that it places more demanding restrictions on any testable theory that is to be ultimately acceptable. (Such a theory must, in principle, predict and explain all physical phenomena, and must be unified in a type 1 way – the most demanding requirement for unity.) Physicalism(1) also indicates a path along which physics may proceed in order to improve empirical knowledge: testable theories need to be put forward and tested that, as far as possible (a) predict ever wider ranges of phenomena, and (b) are ever more unified. In order to develop good new theories, the attempt needs to be made to resolve clashes between existing empirically successful, unified theories. In short, physicalism(1,1), if true, indicates that AOE needs to be put into scientific practice.

But it is not just that physicalism(1,1) holds out the promise of progress; it has been associated, implicitly, with all the great advances in theoretical knowledge and understanding in physics at least since Galileo's time.

All advances in theory in physics since the scientific revolution have been advances in unification, in the sense of (8) to (1) above. Thus Newtonian theory (NT) unifies Galileo's laws of terrestrial motion and Kepler's laws of planetary motion (and much else besides): this is unification in senses (8) to (6). Maxwellian classical electrodynamics, (CEM), unifies electricity, magnetism and light (plus radio, infra red, ultra violet, X and gamma rays): this is unification in sense (5). Special relativity (SR) brings greater unity to CEM, in revealing that the way one divides up the electromagnetic field into the electric and magnetic fields depends on one's reference frame: this is unification in sense (3). SR is also a step towards unifying NT and CEM in that it transforms space and time so as to make CEM satisfy a basic principle fundamental to NT, namely the (restricted) principle of relativity. SR also brings about a unification of matter and energy, via the most famous equation of modern physics, $E = mc^2$, and partially unifies space and time into Minkowskian space-time. General relativity (GR) unifies space-time and gravitation, in that, according to GR, gravitation is no more than an effect of the curvature of space-time – a step towards unification in sense (1). Quantum theory (QM) and atomic theory unify a mass of phenomena having to do with the structure and properties of matter, and the way matter interacts with light: this is unification in senses (5) and (4). Quantum electrodynamics unifies QM, CEM and SR. Quantum electroweak theory unifies (partially) electromagnetism and the weak force: this is (partial) unification in sense (2). Quantum chromodynamics brings unity to hadron physics (via quarks) and brings unity to the eight kinds of gluons of the strong force: this is unification in sense (3). The standard model (SM) unifies to a considerable extent all known phenomena associated with fundamental particles and the forces between them (apart from gravitation): partial unification in senses (5) to (2). The theory unifies to some extent its two component quantum field theories in that both are locally gauge invariant (the symmetry group being U(1)XSU(2)XSU(3)). All the current programmes to unify SM and GR known to me, including string theory or M-theory, seek to unify in senses (5) to (1).[32]

In short, all advances in fundamental theory since Galileo have invariably brought greater unity to theoretical physics in one or other, or all, of senses (8) to (1): all successive theories have increasingly successfully exemplified and given precision to physicalism(1,1) to an extent which cannot be said of any rival metaphysical thesis, at that level of generality. The whole way theoretical physics has developed points towards physicalism(1,1), in other words, as the goal towards which physics has developed. Furthermore, what it means to say this is given precision by the account of theoretical unity given in section 3 above.

In response to this claim it may be objected that theoretical physics could equally well be regarded as pointing towards a less restrictive version of physicalism – one which does not require matter and space-time to be unified, or one which demands only that the true theory everything is no more disunified than in a type 4 way to an extent N = 3, let us say (so that the true theory postulates three kinds of forces). What grounds are there for preferring physicalism(1,1) to physicalism(4,3), let us say? There are at least four, none of course decisive.

Fundamental to the whole argument for AOE is that physics needs to put the Principle of Intellectual Integrity into practice (and I have claimed that AOE can be construed as the outcome of successive applications of this principle). In considering what thesis ought to be accepted at level 4, then, we need to consider what is implicit in those current methods of physics that influence what theories are to be accepted on non-empirical grounds – having to do with simplicity, unity, explanatoriness. There can be no doubt that, as far as non-empirical considerations are concerned, the more nearly a new fundamental physical theory satisfies all eight of the above requirements for unity, with N = 1, the more acceptable it will be deemed to be. Furthermore, failure of a theory to satisfy elements of these criteria is taken to be grounds for holding the theory to be false even in the absence of empirical difficulties. For example, high energy physics in the 1960s kept discovering more and more different hadrons, and was judged to be in a state of crisis as the number rose to over one hundred. Again, even though the standard model (the current quantum field theory of fundamental particles and forces) does not face serious empirical problems, it is nevertheless regarded by most physicists as unlikely to be correct just because of its serious lack of unity. In adopting such non-empirical criteria for acceptability, physicists thereby implicitly assume that the best conjecture as to where the truth lies is in the direction of physicalism(1,1). The Principle of Intellectual Integrity requires that this implicit assumption – or conjecture – be made explicit so that it can be critically assessed and, we may hope, improved. Physics with physicalism(1,1) explicitly acknowledged as a part of conjectural knowledge is more rigorous than physics without this being acknowledged because physics pursued in the former way is able to subject non-empirical methods to critical appraisal as physicalism(1,1) is critically appraised, whereas physics pursued in the latter way cannot do this. Because physicalism(1,1) makes more definite, substantial claims than any rival version of physicalism, it is more open to critical appraisal than rival versions.

A second point to note is that it may well be that, even if some other version of physicalism(n,N) is true, with n > 1 and N > 1, nevertheless our best hope of discovering the truth may still lie in attempting to discover a theory that exemplifies physicalism(1,1), and failing in the attempt. As N becomes bigger, so the number of possible theories of everything compatible with that version of physicalism rapidly increases. (If N = 2, and the universe is made up of two distinct unified, dynamical patterns, there are, nevertheless, in general, infinitely many ways in which these two distinct patterns can be fitted together to make infinitely many different possible universes exemplifying just these two dynamic patterns. The step from one specified unified

[32] For further discussion see (Maxwell 1998, 80-89, 131-40, 257-65 and additional works referred to therein).

pattern to two is the step from one possible universe to infinitely many!) It makes sense to seek the simplest, most discoverable possibility, and design our methodology accordingly. As I mentioned at the beginning of this chapter, one can imagine a universe in which we might have reasons for adopting a methodological rule different from: (A) in order to be ultimately acceptable, a theory must be comprehensive and unified in a type (1) way. An example is: (B) in order to discover the true theory of everything, there need to be infinitely many theoretical revolutions, the number of forces increasing by one at each revolution. We cannot, therefore, just argue that, even if some version of physicalism other than physicalism(1,1) is true, nevertheless our best hope of discovering the truth is to adopt (A), try to discover a theory that exemplifies physicalism(1,1), and fail in the attempt. But we can argue that, *in our current state of ignorance*, our best bet is to adopt (A), and revise our acceptance of physicalism(1,1) if some other version of physicalism should emerge as appearing to fit the progress of physics better. (A number of revolutions have taken place, and each time, the number of forces has gone up by one.)

There is another reason for preferring physicalism(1,1) to any other version, namely: only this version can do justice to the way general relativity unifies gravitation and space-time. This is a step towards type (1) unification in that, according to the theory, gravitation as a force disappears, and we are left with a dynamic theory of space-time. (Matter, or energy-density more generally, tells space-time how to curve: bodies then move along geodesics – the nearest things to straight lines in curved space-time.) It is above all general relativity which holds out the possibility that, not just gravitation, but all the forces and particles may be unified with space-time.

In short, physicalism(1,1) seems to be the best bet when one takes into account (a) its inherent promise of progress, (b) the manner in which it is exemplified in every accepted new fundamental theory in physics, (c) its greater fruitfulness for progress even if some other version of physicalism is true, and (d) the way in which it is suggested by general relativity.

Finally, it needs to be remembered that what we are discussing is reasons for accepting physicalism(1,1) at level 4 *within the context of AOE*. If physicalism(1,1) was a candidate for the *only* metaphysical thesis to be accepted by science, it might well be thought to be much too specific and risky to be regarded as a part of scientific knowledge. But the whole point of AOE is that, as we descend the hierarchy, theses become increasingly specific, risky, tentative, and likely to require rejection, or at least revision. Physicalism(1,1) is bound to have a much more dubious epistemological status than partial knowability, let us say.

This concludes my discussion of reasons for accepting physicalism(1,1) at level 4.

Level 3: *Best Blueprint*. The best available more or less specific metaphysical view as to how the universe is physically comprehensible, a view which asserts that everything is composed of some more or less specific kind of physical entity, all change and diversity being, in principle, explicable in terms of this kind of entity. As I have already mentioned, examples, taken from the history of physics include: the corpuscular hypothesis of the 17th century, according to which the universe consists of minute, infinitely rigid corpuscles that interact only by contact; the view, associated with Newton and Boscovich, according to which the universe consists of point-atoms that possess mass and interact at a distance by means of rigid, spherically symmetrical, centrally directed forces; the unified field view, associated with Faraday and Einstein, according to which everything is made up of one self-interacting field, particles of matter being especially intense regions of the field. Some might argue that the best available blueprint available today is the basic metaphysical idea of superstring theory, or M-theory as it is now called: the universe consists of minute quantum strings that move in 10 or 11 dimensions of space-time, all but four of which are curled up into a minute size, thus escaping detection. In Maxwell (1998, chapter 3) I argue, however, that the best available blueprint is a somewhat more general thesis that I call *Lagrangianism*. What one requires, of course, is a metaphysical idea which unifies key ideas taken from quantum theory and general relativity. My suggestion, along these lines, is probabilistic dynamic geometry of space-time (Maxwell, 1985a, pp. 40-41; 2006b, pp.240-1).

Level 2: Accepted fundamental Physical Theory. All accepted fundamental dynamical theories, or accepted laws governing the way physical phenomena occur if no dynamical theory has been developed that applies to the phenomena in question. In terms of current scientific knowledge, this level consists of the so-called standard model (SM) – the quantum field theory of fundamental particles and the forces between them – plus general relativity (GR). We are justified in accepting these theories because, better than any available rivals they satisfy the two requirements of (a) empirical success, and (b) unity, as explicated above, thus exemplifying (in the best available way) the best level 4 thesis – physicalism(1,1).[33]

This concludes my discussion of reasons for preferring theses at levels 7 to 2 of the version of AOE depicted in figure 12.

It may, perhaps, be conceded that AOE solves the *methodological* problem of induction. And it might just about be conceded that the above discussion solves the *theoretical* problem of induction. But none of this, it may be objected, goes any way at all towards solving the *practical* problem of induction. The reason is very simple. None of the arguments given above for accepting theses at the various levels of AOE provide any grounds whatsoever for believing these theses are likely to be true. They are only reasons for accepting these theses *granted that our aim is to improve knowledge, insofar as this is possible*. What counts, in other words, is the fruitfulness of these theses from the standpoint of improving theoretical knowledge. But when it comes to action, what we require is reasons for holding relevant factual propositions are true. Conjectures, speculations, are not good enough when it comes to risking our own lives, or the lives of others.

I have three points to make in support of the claim that the above does succeed in solving the practical problem (insofar as it is capable of being solved).

The first is this. Accept AOE, accept that the level 4 thesis of physicalism is a part of our knowledge, and a sharp distinction can be drawn between *certainty* and *speculation* – a distinction that eludes Popper's account of the matter. Briefly, and roughly, factual propositions which are sufficiently well corroborated and sufficiently in accord with physicalism fall into the category of trustworthy knowledge; all other factual propositions that have not been falsified fall into the category of mere speculation. This, I claim, reflects the way we actually demarcate *trustworthy knowledge* from *mere speculation*. To take an example considered by John Worrall (1989), we do not jump off the top of the Eiffel tower, entrusting our life to the truth of the conjecture that we will float gently down to the ground because this conjecture fails to satisfy the two requirements for trustworthy knowledge. It is no doubt possible to concoct a theory that is more acceptable, according to the methodology of (Popper, 1959), than Newton's or Einstein's theory of gravitation – an *ad hoc* theory concocted to have greater empirical content and success than either – but such a theory would clash severely with physicalism. This demarcates trustworthy knowledge from speculation, but does not provide a *justification* for the distinction. For that, some kind of justification of physicalism is required. Is any forthcoming?

This leads me to my second point. Even our most humdrum, particular, factual items of knowledge about our immediate circumstances, presupposed by our ordinary actions in life, have a cosmological dimension, as we have seen. Cosmological assumptions, or conjectures, are an inevitable part of almost all that we take to be factual knowledge, whether commonsensical or scientific. The crucial question, in the context of practical life, is: which cosmological conjecture,

[33] Some theories clash with physicalism(1,1) more severely than others. What, it may be asked, can this mean? The account of theoretical unity, given above, provides the answer. Given two rival sets of fundamental physical theories, T_1 and T_2, each set aspiring to be comprehensive, T_1 clashes more severely with physicalism(1,1) than T_2 if T_1 is disunified in a more serious way than T_2. (Theories become increasingly seriously disunified as n goes from 1 to 8.) If T_1 and T_2 are disunified in the same kind of way, then T_1 clashes more severely with physicalism(1,1) than T_2 if T_1 has a greater degree of disunity than T_2.

of those available, is to be preferred? The only guideline we have available as to which is most likely to be true is: Which seems best to promote acquisition of empirical knowledge? The answer, as we have seen, is physicalism(1,1).

We have before us, let us suppose, a number of candidate cosmological theses: physicalism(1,1) and theses A, B, C,... (which might include the Aristotelian thesis that everything is to be explained in terms of some overall cosmological purpose, the thesis that natural phenomena exemplify a cosmological computer programme, and the thesis that phenomena occur as a result of the will of God). How should we choose? (We assume the theses are all consistent, and viable cosmological theses in that each can apparently accommodate everything that exists.) One consideration, clearly, is to see which is implicit in our everyday actions, and is presupposed by that part of what we take to be knowledge upon which we base our actions. Let us suppose all the candidates pass this test. The only remaining relevant consideration is: Which thesis holds out the greatest hope of empirical progress, if true, and is actually associated with what seems to be progress in empirical knowledge? An untestable, metaphysical thesis that holds out the promise of progress in empirical knowledge, if true, has a kind of quasi-testable status. If it is adopted as the blueprint of an actively pursued research programme, and this programme, even after decades or centuries of endeavour, makes no substantial progress, this tells against the blueprint. But if, on the other hand, the research programme seems to make rapid, even ever accelerating progress, this tells for that blueprint. What better indication could we have of the truth of the blueprint than that assuming it to be true is uniquely fruitful for the acquisition of knowledge? Given this way of assessing cosmological theses, the grounds for preferring physicalism(1,1) to all other candidates are overwhelming.

But the above argument has, of course, a built in circularity (which no doubt explains why philosophers ignore it). It is perfectly possible, in other words, for natural science to appear to achieve spectacular progress in empirical knowledge – this success being uniquely associated with science presupposing physicalism(1,1) – and yet for physicalism(1,1) to be grossly false. The success might be illusory, either in a way which could in principle be discovered, or in a way which could not, even in principle, until some specific time in the future (when 'the laws of nature abruptly change').

This circularity problem was solved above. If apparent scientific progress is illusory in a discoverable way, well, AOE is uniquely equipped to discover it. The circularity feature of AOE (as far as discoverable illusory success is concerned) provides no grounds whatsoever for not implementing AOE, and accepting the results of AOE science as a basis for action, when these results are sufficiently well corroborated empirically. If, on the other hand, scientific progress is illusory in a way which is not discoverable (until all is revealed), then nothing can be done to guard us against such possible future disasters. Not just AOE science, but *any* methodology, *any* procedure or way of life, must be vulnerable to such undiscoverable illusory success. That AOE is vulnerable to it, and cannot guard against it, is thus no reason whatsoever for not accepting, as a basis for action, the well-established results of AOE science. Since *nothing* can anticipate, and protect us from, such unanticipatable disaster, it's foolish to blame AOE for being unable to anticipate, and protect us from, such disaster. There is here no reason not to accept well-established results of AOE science as a basis for action.

My third and final point is this. Before the scientific revolution, there was much more general awareness, than there is today, that what may be called cosmological circumstances could impact, in perhaps drastic and dreadful ways, on the ordinary circumstances of life. Evil spirits might cast spells and bring catastrophe, even death; comets might bring disaster; the gods might send drought, locusts, storm, the plague, and might even destroy the world. Then came science, and with it the assurance that the natural world is governed by impersonal, utterly reliable physical law. This, it seemed, had been securely established by Newtonian science. Had not Newton himself demonstrated how physical laws can be verified by induction from phenomena|? There remained the niggling philosophical puzzle as to how it is possible to verify laws by means of induction, but this irritating puzzle of induction is best left to philosophers to waste their time on.

This rather common attitude – common at least until recently (scepticism about science having recently become much more widespread) – rests on an illusion. Newton did not establish his law of gravitation by induction from the phenomena, as he claimed to have done. He could not have done this, because it cannot be done.

As it happens, Newton himself anticipated a basic feature of AOE. He recognized explicitly that scientific method makes presuppositions about nature. Three of his four rules of reason, concerned with simplicity, quite explicitly make assumptions about the nature of the universe. Thus rule 1 asserts: '*We are to admit no more causes of natural things than such as are both true and sufficient to explain their appearances*.' And Newton adds: 'To this purpose the philosophers say that nature does nothing in vain, and more is in vain when less will serve; for Nature is pleased with simplicity, and affects not the pomp of superfluous causes' (Newton, 1962, p. 398). Newton understood that persistently preferring simple theories means that Nature herself is being persistently assumed to be simple.

But this aspect of Newton's thought came to be overlooked. The immense, unprecedented success of natural science after Newton was taken to demonstrate that humanity had somehow discovered the secret of wresting truth and certainty from nature, and only the incompetence of philosophers prevented everyone from knowing exactly what this secret amounted to. Even today there are philosophers who think that the problem of induction will only be solved when this secret of how scientists manage to capture truth and certainty is laid bare for everyone to see and understand.

But this is an illusion. Even our most humdrum, particular, practical knowledge of aspects of our immediate environment, as we have seen, let alone the mighty claims to knowledge of science, contains a cosmological element which must remain conjectural. Modern science has, it seems, made a profound discovery about the ultimate nature of the cosmos, namely that it is physically comprehensible. Once AOE is accepted, it becomes clear that this thesis, despite its *metaphysical* and *cosmological* character, is one of the most firmly established theoretical propositions of science (in that physical theories, in order to be accepted, must accord with this proposition as far as possible, and theories which clash with it too stridently are not even considered, even though they would be much more empirically successful than accepted theories if considered). Given this cosmological thesis that the universe is physically comprehensible, the way we in practice distinguish trustworthy knowledge from mere speculation becomes clear. Nevertheless, despite its central place and role in science, the thesis remains inherently conjectural in character. Practical certainty has this usually unacknowledged conjectural and cosmological dimension inherent in it.

As it is, our attitude towards the thesis that the universe is physically comprehensible is highly hypocritical. The fundamental role that it plays in science, in technology, in our whole culture and way of life, is denied. Non-scientists deny it because they do not want to confront the grim implications the thesis has for the meaning and value of human life – the difficulty of seeing how there can be consciousness, freedom, meaning and value if the universe really is physically comprehensible.[34] Scientists deny it, because they do not want to acknowledge that there is an element of *faith* in science. They confidently distinguish science from religion on the grounds that, whereas religion appeals to dogma and faith, in science there is no faith and everything is assessed impartially with respect to evidence. But this, as we have seen, is nonsense. There is an element of faith in science too. The real difference between science and religion – most dogmatic religions that is – is that whereas science subjects its articles of faith to sustained critical scrutiny,

---

[34] Elsewhere I have sought to show how consciousness, free will, the experiential world, meaning and value can exist even though the universe is physically comprehensible: see chapter 10 of the present work and Maxwell (1966; 1968; and especially 2001).

modifying them in the direction of that which seems most fruitful from the standpoint of the growth of knowledge, dogmatic religion does nothing of the kind. We are justified in accepting physicalism as a part of our knowledge, even in the context of practical action, because some such cosmological conjecture must be accepted, and physicalism has proved more fruitful for progress in knowledge than any rival. It is always possible that this success is illusory, and physicalism is no more than a kind of scientific hallucination. But if the success of science is illusory in a way we could not in principle discover, then this is a possibility we face whatever we assume; it is not something we can do anything about, and deserves to be ignored. If, on the other hand, the success of science is illusory in a way which can in principle be discovered, then AOE science provides us with the best means of unmasking the illusion. Either way, physicalism deserves to be accepted even in practical contexts.

A more honest recognition of the presence of cosmological conjectures inherent in science, and inherent even in our most humble items of practical knowledge would involve recognizing that all our knowledge is indeed conjectural in character without, thereby, destroying the distinction we make between practical certainty and speculation.

Popper has done much to create an awareness of the conjectural character of scientific knowledge – helped, of course, by the dethronement of Newtonian science with the advent of general relativity and quantum theory. But in one crucial respect, Popper helped sustain the Newtonian tradition, the *status quo*. He fiercely defended, to the last, the highly traditional, and mistaken, idea that the scientific character of science depends on it being dissociated from metaphysics.[35] Actually, it is all the other way round. If science is to be rigorous, it is essential that it acknowledge – and so throw open to criticism and improvement – metaphysical and cosmological theses implicit in the persistent scientific selection of unified, explanatory theories. And that is just the start of one line of argument leading to the philosophy of wisdom: not just metaphysics, but values, and political committments too, implicit in the scientific endeavour, need to be made explicit, if science is to be rigorous, so that these problematic assumptions and committments can be criticized and, we may hope, improved.

For further arguments intended to show that AOE solves all three parts of the problem of induction, insofar as they can be solved, see Maxwell (1998, ch. 5; 2004b, appendix, section 6; and 2005e).

Appendix

In this appendix I say, in a little more detail, what it means to say that T_2 is closer to the truth than T_1 in the sense that it makes more precise predictions of more phenomena.

In order to do this, we need to consider, as before, the following paradigmatic kind of prediction that a dynamical physical theory makes. We have the physical theory, T_1 let us say, and any isolated physical system or physical state of affairs, S, which is such that T_1 predicts how S evolves in time. Let the specification of the initial state of the system at time t_1, formulated in terminology appropriate to T_1, be S^1_1, and let the specification of some later state of the system, at time t_2, be S^1_2. T_1 might be Newtonian theory, S^1_1 might be a specification of the state of the solar system at some moment t_1, S^1_1 specifying the instantaneous positions, velocities and masses of the sun and planets at time t_1, and S^1_2 specifying the state of the solar system, as predicted by Newtonian theory, at a later time t_2. We have that T_1 and S^1_1, taken together, imply S^1_2; i.e. $(T_1 \& S^1_1) \rightarrow S^1_2$. Newtonian theory plus a specification of the instantaneous state of the solar system implies specifications of future states – future positions and velocities of the sun and planets.[36]

---

[35] For a discussion of this defect in Popper's work, see (Maxwell, 2005b and 2007a).

[36] This simplifies what is, in scientific practice, much more complicated. The way initial and boundary conditions need to be formulated changes from one theory to another, as one moves from Newtonian theory to a field theory, such as classical electrodynamics, to general relativity and quantum theory. These complications are not relevant to the problem under discussion, and can be ignored. Orthodox quantum

In the case of a false theory, such as T_1, T_2 or Newtonian theory, the theory, T_1, and the corresponding specifications of state, S^1_1 and S^1_2, are all false. To take the example of Newtonian theory, it is not just that the theory is false, but the Newtonian specifications of the instantaneous states of the system will be false as well. All these Newtonian propositions presuppose that space is Euclidean, for example, but general relativity tells us that space is not Euclidean, and the true theory of everything, T, is likely to tell us this too. T may depart even more radically from the presuppositions of these Newtonian propositions in that it asserts that space, and perhaps time, are discontinuous, whereas Newtonian theory presupposes that both are continuous. Again, the Newtonian propositions assume that physical systems, such as the sun and planets of the solar system, are made up of classical particles with mass, and with definite positions and velocities at successive moments. But this is denied by quantum theory, and is likely to be denied, if anything, even more emphatically, by the true theory of everything, T.

Despite the presumed falsity of T_1, S^1_1 and S^1_2, T_1 can still issue in true predictions. S^1_2 can be specified in a looser, approximate fashion, S^1_2* say, so that S^1_2* asserts something like 'the state of S is S^1_2 within such and such a range of values of variables (such as relative positions and velocities)'. In the case of Newtonian theory, S^{NT}_2* tells us, not precisely where each planet is and what its precise velocity is, but rather specifies a volume of space within which such and such a planet is located, having a velocity of such and such a range of values. We stipulate that S^1_2* is concocted so that (a) $S^1_2 \rightarrow S^1_2*$, and (b) S^1_2* is compatible with T. (It may be asked: How is it possible for the false proposition S^1_2 to imply the looser, approximate *true* proposition, S^1_2*? A trivial example of this is the following statement, uttered on Wednesday: ''Today is Monday' implies 'Today is a weekday''.[37])

We may stipulate that S^1_2* is as precise as it can be without being false. We don't have to assume that there is just one true approximate specification of S at time t_2. Given one such specification, S^1_2*, it seems reasonable to suppose that others could be generated by making the range of values of one variable, specifying the state of S approximately, a little bit more precise, as long as the range of values of another variable is made, compensatingly, less precise. There may be infinitely many such different, true, minimally approximate specifications of the state of S at time t_2, corresponding to the precise, false specification S^1_2. (We require that there is no true S^1_2** such that $S^1_2* \rightarrow S^1_2**$ but not $S^1_2** \rightarrow S^1_2*$.)

Just as there is a true minimally approximate specification of S at time t_2, namely S^1_2*, so too we can stipulate that there is a true, minimally approximate specification of the state of S at time t_1, namely S^1_1*. And, as before, there may be infinitely many such true approximate specifications.

We have, then, that $(T_1 \,\&\, S^1_1) \rightarrow S^1_2 \rightarrow S^1_2*$.

Everything stipulated about T_1, S^1_1, S^1_1*, S^1_2 and S^1_2* is also stipulated to hold for T_2, S^2_1, etc. In the case of the true theory of everything, T, there are the two true precise specifications of the state of S at time t_1 and t_2, namely S^T_1 and S^T_2.

Now a few remarks about the relations between T_1, T_2 and T. Corresponding to any given pair, S^1_1* and S^1_2* there will be, we may presume, infinitely many different, true, precise specifications of possible states of systems, S^T_1 and S^T_2. Newtonian theory, applied to the solar system, need not take into account the positions of all the constituent atoms, or fundamental particles, of which the planets are composed. Infinitely many re-arrangements of atoms will not affect the way the solar system evolves, as predicted by Newtonian theory. Because the specifications of states of systems, formulated in terms of T, are so much more precise than specifications formulated in terms of T_1 or

---

theory, lacking its own ontology, must appeal to a process of preparation, and cannot specify an initial state in purely quantum mechanical terms, but in this respect the theory is unsatisfactory: see Maxwell (1998, ch. 7).

[37] A better example, avoiding context, would be 'This sphere is an ellipsoid', said of an ellipsoid that is not a sphere. (A sphere is a special case of an ellipsoid, so that all spheres are ellipsoids, but not all ellipsoids are spheres.)

T_2, infinitely many different specifications of the former, will correspond to the same specification formulated in terms of T_1 or T_2. We may presume that, likewise, infinitely many different true approximate specifications of states formulated in terms of T_2 correspond to just one pair, $S^1{}_1*$ and $S^1{}_2*$. [38]

With these preliminaries over, we can now state the conditions that must be satisfied for T_2 to be closer to the truth than T_1. This will be the case if:-

(a) Given any predictive task such that both T_1 and T_2 make true approximate predictions, the prediction of T_2 is more accurate than that of T_1. That is, given that the two corresponding predictions are $(T_1 + S^1{}_1*) \rightarrow S^1{}_2*$ and $(T_2 + S^2{}_1*) \rightarrow S^2{}_2*$, then $S^2{}_1*$ is at least as precise as $S^1{}_1*$ and $S^2{}_2*$ is more precise than $S^1{}_2*$. (That is, $S^2{}_2* \rightarrow S^1{}_2*$ but not $S^1{}_2* \rightarrow S^2{}_2*$, and if $S^1{}_1* \rightarrow S^2{}_1*$ then $S^2{}_1* \rightarrow S^1{}_1*$.)

(b) There are many (presumably infinitely many) true approximate predictions of T_2 (implied by T_2 and appropriate specifications of initial conditions, statements of type $S^2{}_1$), to which there correspond no such predictions of T_1 (but not *vice versa*).

In other words, T_2 is closer to the truth than T_1 if (a) everything true that T_1 predicts, T_2 can predict with greater accuracy, and (b) T_2 makes true predictions about which T_1 is silent or can predict nothing true whatsoever.

Not only are the true approximate predictions of T_2 more precise than those of T_1; the true approximate specifications of initial conditions of T_2 are at least as accurate as those of T_1.

[38] It is just possible that, for some range of phenomena to which T_1 applies, the theory is so badly false that it is difficult to see what true approximate predictive statements it implies. If this is the case, then this range of phenomena must be ignored, and only that range considered for which T_1 does yield true approximate predictions.