Reasoning from Phenomena: Lessons from Newton

Jon Dorling

University of Amsterdam

1. Introduction

On the model of Newton's Principia, the great majority of successful new theories in physics have been introduced by deduction from the phenomena arguments. In such arguments an explanatory theory is deduced from one or more of the empirical facts, or lower level empirical generalizations, which it is designed to explain, by the device of adjoining suitable higher-level theoretical constraints on the form of the required theory. Those theoretical constraints leave certain parameters, the precise form of certain functions, and so on, in the new theory, undetermined, except with the help of the lower level empirical premises.

Although Newton's own complete argument to his inverse square law did contain at least one additional inductive step, it is not difficult to show that deduction from the phenomena arguments can be rigourously deductively valid within modern formal logic, (i.e. all the inductive steps can often be confined to the justification of the premises of such arguments) and that nearly all theoretical advances in physics since Newton have depended partly or wholly on the use of arguments of this general form. Sometimes the high-level theoretical constraints invoked are claimed partly or wholly to follow from a priori justifiable principles, but more usually they are either merely claimed to be plausible inductive generalizations from all experience (as Newton claimed for his three laws of motion which functioned as theoretical constraints in the deduction of his gravitational force law), or, as in most later examples, they are merely claimed to be derived by inductive extrapolation from the successful parts of previous theories.

Now most philosophers would probably be inclined to suppose that it is a truism of probability theory that the conclusion of such a deductively valid argument, namely the new theory thus deduced, must, if validly deduced, end up with at least the initial probability/plausibility of the conjunction of its premises. Hence, since in virtually all examples the higher-level premises are deliberately chosen so as already to appear (possibly with the help of arguments the innovatory theorist has himself just adduced) reasonably, or highly, probable/plausible, and since the lower-level premises have the status of relatively uncontroversial empirical facts and low-level laws, it would then seem that any new theories introduced in this way must themselves automatically be

PSA 1990, Volume 2, pp. 197-208
Copyright © 1991 by the Philosophy of Science Association

https://doi.org/10.1086/psaprocbienmeetp.1990.2.193068 Published online by Cambridge University Press

granted a reasonable, or high, degree of probability/plausibility by scientists. P is probable, P implies Q, therefore Q is probable.

Unfortunately the situation is by no means so simple. For not only have nearly all successful innovations in physics been introduced by arguments of this general form, but so have nearly all unsuccessful innovations in physics, and of course historically there have been far more of the latter. And in most of the latter cases the corresponding deductive justifications from the phenomena, in spite of their logical validity, never did cut much ice with other scientists, even before any direct evidence emerged for the falsity of their conclusions. For those other scientists took these arguments' conclusions as discrediting their theoretical premises: Q is improbable, P implies Q, therefore P is improbable.

To clarify this situation we need to appreciate that such deductive discoveries inevitably lead to revisions of the subjective probabilities initially assigned both to their conclusions and to their premises.

Thus if a seemingly highly unlikely conclusion (for example a surprisingly complicated equation relative to the apparent simplicity of the data requiring explanation) is deduced from premises which had previously seemed very likely to be true, then of course scientists are far more likely as a result of such a deductive discovery to lose confidence in, and to begin to question, one or more of the premises of the deductive argument, than to give their assent to its conclusion. In fact what often happens is that scientists conclude that one or more of the apparently plausible theoretical premises must after all probably be false, without their being able to say which that is. They lose confidence in the truth of the conjunction of premises without necessarily concluding of any particular conjunct that it has become less probable than not.

The opposite kind of example is the following: if theoretical premises which seemed at first, to most other theorists, theoretically unlikely, nevertheless lead to the deduction of an unexpectedly simple new explanatory theory, then as a consequence of that deductive discovery both the new theory and the premises which led to it can end up being assigned much higher degrees of assent than that originally granted to the initial conjunction of premises.

It is possible to give numerous historical examples of both these types of situation. The general rule seems to be that if the theory deduced proves to be more complicated/implausible-looking, than theorists reasonably expected it would have to be, in order to explain the relevant empirical data, then the deductive discovery does more to discredit its premises than to raise confidence in its conclusion; while if the theory deduced proves to be simpler than theorists had anticipated that it would be, then the deductive discovery strengthens scientists' confidence both in its premises and its conclusion. Let us call examples of this second kind impressive deduction from the phenomena arguments. (Of course there are many intermediate examples where the deductive discovery is noted as an interesting technical result, but doesn't induce any marked change in scientists' prior opinions.)

Newton's deduction from the phenomena argument to his inverse square law was, of course, of the impressive kind just mentioned. The same is true of many examples in Einstein's papers of arguments of a similar logical form. Though Einstein himself seems never to have noticed how essentially Newtonian in structure his principal derivations of his own succesful new theories were, or to have commented anywhere (as nearly all great previous theoretical physicists had done) on the apparent superiority of such a Newtonian justificational strategy over naive hypothetico-deductivism.

https://doi.org/10.1086/psaprocbienmeetp.1990.2.193068 Published online by Cambridge University Press

(For further discussion of Einstein's use of deductions from the phenomena see Dorling 1991.)

What I want to consider in the present paper is how much logical weight such impressive deduction from the phenomena arguments (which certainly have a persuasive effect on their scientific audiences) really carry. Do they, when deductively valid, in fact legitimate the assignment of high rational probabilities to their conclusions? Newton plainly thought they did. I shall argue that Newton was wrong, but that he was wrong in a surprising way, which he could hardly have anticipated, and which would probably in fact have delighted him.

For I shall argue that what is wrong with 'impressive' deduction-from-the-phenomena arguments is not that the theories which they lead to turn out to be too simple to be exactly true, thus not that they have to be succeeded in the long run by more complicated theories, but rather the reverse: such theories, in spite of superficial appearances of simplicity, have never turned out in the long run to be simple enough: the theories which later replace them, far from being, as they superficially appear to be, more complicated replacements, constitute in fact simpler explanations of the original data—though this often only becomes apparent later, when a sufficiently deep level of mathematical and logical analysis becomes available.

In a later section I shall illustrate this by considering the line of successors to Newton's theory of gravity. But before doing this it will be useful if I digress a little so as to bring readers up-to-date with our current very satisfactory understanding of inductive logic.

2. Digression: modern simplicity-based inductive logic

The centuries-old problem of setting up a completely general, powerful, plausible, and mathematically coherent system of inductive logic, was finally solved by the work of R. Solomonoff in the early 1960's, with some technical improvements by L. Levin in the early 1970's. The Solomonoff-Levin solution treats theories as computer programs for regenerating all the data as output, that is to say as encodings of the data, and assigns prior probabilities to theories according to the number of bits they require as prefix-free programs in the theorist's internal programming language. Such prior probabilities fall off by a factor of two for each extra bit required in the statement of a theory.

The mathematical background to this solution is well-known as Complexity-theory, or as the theory of Kolmogorov Complexity, though it would have been historically more accurate (since Solomonoff's publications anticipated Kolmogorov's), and philosophically more illuminating, to call it Simplicity-Theory or the theory of Solomonoff-Simplicity. One of the fundamental theorems in this theory shows that for any reasonably non-trivial theories their relative priors so-assigned are negligibly dependent on the choice of original programming language.

A recent review of this approach can be found in Li and Vitányi 1991. These authors, apparently following Solomonoff himself, seem to regard these simplicity-based rational priors as a sort of elegant mathematical substitute for unknown philosophically correct priors. I, however, maintain that these simplicity-based priors really are the unique philosophically correct priors.

The attempt to link rational probabilities to simplicity has of course a long history. One can consider Newton's rules of reasoning (with their missing ceteris paribus claus-

https://doi.org/10.1086/psaprocbienmeetp.1990.2.193068 Published online by Cambridge University Press

es added) as simplicity constraints on inductive steps in theory-construction. However those simplicity constraints did not quite work, for Newton's successors such as Boscovich pointed out that Newton's rules of reasoning applied indiscrimately could lead to inductive generalizations inconsistent with one another. In the modern approach it is the overall simplicity of the complete theory which determines its rational probability, not the requirement that each of the theory's epistemically distinguishable ingredients should be the simplest encoding of some particular fraction of the data. But we could not forge the necessary precise link between simplicity and rational probability prior to Solomonoff's and Kolmogorov's precise formal explications of simplicity.

So we now have an adequate theory of rational probabilities, ones which are language-independent for all practical purposes, and this theory yields an inductive logic satisfying the aspirations of all probabilistic inductivists from Laplace through Carnap to the Finns, an inductive logic which is in no way restricted to observational-language theories, since theoretical terms and parameters will be preferred whenever they can be introduced so as to shorten our theoretical encodings of the data.

This does not mean that different theorists now have to assign essentially the same priors to new theories. On the contrary, the number of bits required to add a new theory to a given epistemic system will depend not only indirectly on all the other evidence that that system has been developed to account for, but also in addition on just what definitions and constructs have already been introduced into the system in the interests of overall reduction of total epistemic program-length. The relative simplicity assigned by a theorist to a new theory, and hence the relative prior probability he should assign to it, is given by the number of bits he has to add to his total epistemic system to incorporate it, but this number will depend not only on his whole intellectual and experiential background, but also on the logical and mathematical skill with which he has constructed his epistemic system to date, for example on what notations he has already introduced so as to effect earlier encoding economies. For any new abbreviations he introduces in order to shorten a theory will themselves count as part of the cost in bits of that theory. Thus a differential geometer may give General Relativity theory a higher rational prior than an experimental physicist would, simply because the former's total epistemic system has already economically encoded much of the necessary technical apparatus. The rational prior a theorist should give to a theory depends on the actual economy of its encoding in his own epistemic system (or perhaps, if the approach is extended to a meta-level, on the economy of encoding he believes he could give to it in that epistemic system) and not on some hypothetical optimal encoding which he has not yet discovered. Thus different encodings are treated in the first instance as different theories for the individual in question, and an improved encoding as an improved theory. (Our more usual concept of theory would only emerge at a later level by adding the rational priors of different programs provably delivering the same output.)

We also have to be careful to distinguish the apparent number of raw data bits from the number of bits required for encoding that data. For example it is not the case that in deduction from the phenomena arguments, we can simply add the raw data bits for the experimental premises to the bits originally needed for the theoretical premises of the deduction, in order to bound from above the bits needed for the newly deduced theory, and hence to bound from below its rational prior. For that theory may have a very low simplicity ranking among theories consistent with those theoretical premises on their own. This means that the raw data bits for the experimental premises may grossly underestimate the number of bits we need to add to the epistemic system as a whole (to retain its coherence), in order to accommodate that data. Thus an unexpected null result of an experiment may require adding lots of bits to the total epistemic

https://doi.org/10.1086/psaprocbienmeetp.1990.2.193068 Published online by Cambridge University Press

system, even though it is (when we ignore the theoretical background) formally the simplest possible result and thus involves fewest raw data bits. In fact until we've carried out theoretical investigations, such as the construction of deduction from the phenomena arguments, we won't know how many bits each particular experimental result adds to our epistemic system. The Solomonoff-Levin approach thus only allows us to derive relative rational probabilities from computable relative simplicities when the latter computations are carried out on the theorist's total epistemic system. Formal simplicities of data, or of whole theories, considered in isolation from a total epistemic system, may thus often be a very poor guide to Solomonoff-simplicity based rational priors. Thus in practical applications the new inductive logic behaves much more like subjective-Bayesianism, than like what one might first expect of a formal-simplicity-based inductive logic.

However while Solomonoff-Levin inductive reasoning seems satisfactory, indeed correct, as our underlying theory of epistemic rationality, there are tricky issues concerning its effective implementation in any finite physical system such as ourselves. The problems are twofold. First the number of alternative programs, of alternative epistemic systems or theories, which theoretically might need to be taken into consideration at any point, is, though finite, unacceptably large, indeed of astronomical dimensions even for a negligibly small amount of empirical input data. Hence, in practice, all alternative theories, whose theoretical priors are large enough to be relevant to a given theoretical choice, cannot possibly even be explicitly listed. Secondly, we are confronted with a problem of noncomputability. There is no effective decision procedure as to whether even a quite short program will yield any data predictions as output at all, let alone the data we are trying to regenerate.

It might at first seem that the latter problem could be solved by discarding only those theories actually computed to be inconsistent with the data so far, and re-normalizing the absolute Solomonoff-Levin priors assigned to the remainder to unity. However in practice this strategy won't do, because in practice inductive reasoning at a meta-level will make it, for most of the theories which would be left in by such a procedure, wildly unlikely that they would actually be consistent with the data so far. (For example most of the high-theoretical-prior theories left in by such an elimination procedure would be crazy theories such as the theory that all the input data bits so far received agreed precisely with the binary expansion of π from the billion billion billionth place onwards. Even a Popperian would surely not wish to assign a high position in his epistemic rank-ordering of surviving theories, to not-yet-actually falsified theories for which there was not yet the slightest reason to suppose that they were consistent with the existing data.)

Moreover, as we shall see, in practive there will often be actually simpler theories which yield all the data and which have been overlooked by the theorists. This means that in practice we only get relative rational probabilities from the Solomonof-Levin approach and not absolute rational probabilities. Whenever we have overlooked yet simpler theories, the absolute rational probabilities can be much lower for the known theories than most of us suppose. Only the relative rational probabilities for the known theories are computable and accessible to us.

Finally I should emphasize the obvious fact that this new inductive logic makes no pretence at being a logic of discovery. How far one can go in writing computer-implementable discovery algorithms for the discovery of shorter encodings of realistic data remains an open question. (I am personally optimistic: presumably we should be able to write discovery algorithms for discovering at least all the theories which

https://doi.org/10.1086/psaprocbienmeetp.1990.2.193068 Published online by Cambridge University Press

humans could discover, and this may well include all theories which are actually true in the empirical world.)

3. The replacement of Newton's theory of gravity by successively simpler theories

Consider the series of successors of Newton's theory of gravity.

(i) The exponent 2, in the inverse square law, functioned as an unexplained constant in Newton's own theory: its value could only be derived, and then only approximately, by arguing backwards from the astronomical data of Kepler and others. The unsatisfactoriness of this was emphasized by Newton's eighteenth and nineteenth century successors. They therefore, without actually changing its predictive content, replaced Newton's action-at-a-distance theory of gravitation by a field theory in which the exponent 2 emerged as a consequence of the 3-dimensionality of space. This was a simpler theory, even on Popperian criteria: for it require fewer empirically-derived parameters.

While this first change did not lead directly to new predictions, it did lead nevertheless to a change in scientific strategy with respect to the treatment of recalcitrant data. From Newton's point of view there was nothing to rule out the possibility of additional terms in the gravitational force formula depending on different powers of the distance. Such additional terms were indeed proposed from time to time from the eighteenth century onwards. But believers in the new, more geometrically-explanatory, formulation of the theory were prohibited from taking such additional terms seriously, and had to seek elsewhere for the explanation of any predictive anomalies. From Newton's point of view there was also nothing to rule out an exponent in the force formula close to 2, but not exactly 2. Nineteenth century data (in particular data on the perihelion of Mercury) in fact led Newcomb and Hall to 'deduce from the more-accurately known phenomena' that this exponent was not exactly 2, but rather 2.0000001573. But no believer in the moregeometrized version of the theory, could entertain this change as a very serious theoretical possibility, since it sacrificed encoding gains which had already been made and added many additional bits to the theory, so he had to consider other possible explanations of the recalcitrant behaviour of Mercury. (One might at first think that such a small deviation from the inverse square law could be accommodated within a geometrical theory by postulating a slightly curved spatial geometry for the universe: but it was already known that the introduction of a curved geometry would not yield a correction of this kind.)

(ii) Newton had, through the kinematics he adopted as part of the mathematical framework of his theory, committed himself to a peculiarly degenerate geometry of temporal intervals. Among other disadvantages, Newton's choice here had the consequence that although his own theory required forces to explain accelerations, accelerated motion was not distinguishable within his own kinematics from non-accelerated motion: unlike the Euclidean case where a curved line differs in its intrinsic metrical properties from a straight line (only the latter is an extremal relative to the metric), Newton's geometry of temporal intervals failed to distinguish a curved line in space-time from a straight one, i.e. failed to distinguish accelerated from uniform rectilinear motion.

Only after the work of Minkowski did it become clear that Newton had not in fact chosen the simplest mathematical possibility for the geometry of space-time. Newton's implicit transformation group, which simply added the Galilean transformations to the transformations of the Euclidean group, was mathematically less simple, more artificial, and less unified, than one which picked instead the combination of the Lorentz and Euclidean groups. And later work, from a more synthetic-geometrical point of

https://doi.org/10.1086/psaprocbienmeetp.1990.2.193068 Published online by Cambridge University Press

view, showed that more of the structure and simple theorems of Euclidean geometry were retained in Minkowski's proposed space-time geometry, than could be retained in Newton's space-time geometry. Thus Newton's geometry proved not to be the simplest generalization of Euclidean spatial geometry to include temporal intervals as well as spatial ones, i.e. to extend spatial geometry into a space-time geometry, but a more complicated alternative requiring more independent geometrical axioms.

This inductive error—a failure to assign the higher probability to the mathematically simpler theory consistent with the data—was what had forced Newton into the additional epistemically improbable conclusion that his dynamics required forces to produce sometimes epistemically unascertainable effects. But this latter defect simply disappeared once the preferred space-time geometry was adopted: absolute accelerations became identifiable with independently-measurable metrical curvature: thus kinematics and dynamics were no longer at odds with one another, and the conjunction of their respective axioms no longer reduced their joint probability and simplicity.

(iii) However this could only take us to a special relativistic theory of gravitation. (Einstein seems at one time to have thought that such a theory could not be consistent with all the data then known observationally, e.g. with that from the experiments of Eötvos. But Nordström and others soon proved Einstein technically mistaken in this belief.) There remained nevertheless unnecessary data bits in any such special relativistic theory of gravitation, ones which had already been mysteriously introduced by Newton; namely the unexplained proportionality of gravitational and inertial masses in Newton's gravitational theory. In fact Newton had quite bizarrely introduced a force field when all that the data evidently required was not a force field but an acceleration field. Only through a seemingly accidental cancellation of unnecessarily-introduced terms arbitrarily set equal to one another, did Newton's force field reduce to the equivalent acceleration field.

One way to solve this difficulty was simply to introduce a special-relativistic acceleration field for gravity. But Einstein found an even simpler alternative. For any such theory, when fully formalized, would contain two axioms where one would do: namely one axiom requiring that the space-time geometry was everywhere flat (i.e. requiring that the full Riemann curvature tensor vanished everywhere) and another axiom specifying gravitational departures from geodesic trajectories as a function of the gravitational source distribution in space-time. But given that gravitation was already at an observational level an acceleration field and not a force-field, these two axioms could, with a saving in bits, simply be combined into one axiom specifying the curvature in a curved space-time geometry as a function of the gravitational source distribution, i.e. one would no longer prescribe that the Riemann curvature tensor was everywhere zero, but make its value a function of the gravitational source distribution. Einstein's general theory of relativity introduced precisely this formal simplification, by combining two axioms into one with a small resultant saving in bits.

(iv) Einstein's theory was not the only theory which would do this or, it first seemed, the simplest. For Nordström's scalar theory of gravitation, in which the fully contracted curvature tensor, the curvature scalar, is simply taken proportional to the rest-mass density, seemed to have a much simpler field equation. However Einstein observed that if one began with a Lagrangian, rather than with field equations, his own theory was formally the simpler, and that it also unified the gravitational behaviour of matter and light in a way which no scalar theory could do, since light has zero rest mass. Both these considerations argued that Einstein's theory was simpler within the context of the rest of physics, since the similarities between matter and light had been a

https://doi.org/10.1086/psaprocbienmeetp.1990.2.193068 Published online by Cambridge University Press

source of encoding gains in physical theory from Newton to the twentieth century, and it had been clear since the eighteenth century (if not earlier: Fermat) that action principles were one of the formally simplest way of formulating dynamical laws.

(v) Newton had already established the wave properties of light, and had conjectured connections between light and electrical attractions and repulsions, though it took more than another century before a consistent theory embodying these phenomena was available, namely the electromagnetic field theory of Maxwell. From the point of view of the Newtonian space-time geometry that theory's equations seemed quite complicated, but in the simpler space-time geometry of Minkowski, it became clear that all that was involved was the replacement of Faraday's geometrical lines-of-force explanation of electro-statics, with point-charges in space as sources, by planes-of-force emanating from the world-lines of charges in space-time as sources. Entirely analogous geometrical constraints then ensure that where the world line of the source becomes curved there is necessarily a propagated change in its associated planes of force, with all the properties of an electromagnetic wave. So taking into account the simplest space-time geometry, and the need to relate inverse power laws to geometry, if unnecessary data bits were not to appear in the theory, Maxwellian field theory became the simplest explanation of electrostatics.

However, a field theory of radiation coupled to a particle theory of matter was not really internally coherent. It led to infinities at the location of point-particles (or at the boundaries of extended particles), and to further infinities (the Rayleigh-Jeans catastrophe) when energy exchanges (later also momentum and angular-momentum exchanges) between the particles and the field were considered. Attempts to resolve the latter difficulty by returning to a particle theory of radiation coupled to a particle theory of matter never really got off the ground theoretically, and the only viable theoretical alternative was then to set up a wave-theory (i.e. a field theory) of matter coupled to the existing wave-theory (i.e. field theory) of radiation. The simplest wave-theory of matter consistent with the simplest space-time geometry was then discovered, by the combined efforts of Schrödinger and Dirac, to be modern relativistic wave-mechanics.

Due to an unfortunate quirk of history, the mathematical techniques for properly understanding this new theory not yet being available, physicists for the next seventy years (1926-1996) did not realise that this theory already without more ado also predicted and explained the particle properties of matter and radiation, and they thought that a further mysterious complication known as second-quantization was necessary, and not being able to understand this, pretended that the fundamental entities were not waves, but 'wave-particles'. In fact mass, charge, and energy quantization already comes out of the exact equations of the ordinary wave theory as a mathematical consequence of the self-coupling of the matter field via the electromagnetic field (and therefore does not need to be added independently) but this self-coupling term was ignored because it made the equations too difficult to calculate with prior to the nineteen eighties. And the so-called 'wave-packet-collapse' is really a pseudo-phenomenon due to the fact that physicists had neglected half the solutions (the advanced solutions) of their coupled time-symmetric equations, and ordinary wave-interference between these solutions and the others already predicts and explains at a classical level all the supposed wave-packet collapse phenomena. Thus the move that had been made in 1929 and 1930 to a more complicated and epistemically wildly improbable theory was, in hindsight, unnecessary. The simpler earlier theory was really actually correct. (For more extended discussion of my unorthodox contentions here, Dorling 1987.)

(vi) However there remained a gross improbability in fundamental physical theory due to the now wildly disparate treatment of gravitational and other forces. The for-

https://doi.org/10.1086/psaprocbienmeetp.1990.2.193068 Published online by Cambridge University Press

mer force was built into the geometry while the latter were still treated as essentially classical forces (i.e. as non-geometrized potentials in quantum-mechanical Lagrangians). This disunity was removed in the later '60's and '70's, when it was realised that once one treats the internal non-spatio-temporal degrees of freedom of the elementary particle fields as determining an internal geometry analogous to space-time geometry, one can perform the same trick as Einstein performed and replace all the other forces by geometrical curvatures, in each case combining two equations into one, with a resulting small saving in bits. This change in the direction of greater simplicity constituted the recent gauge-theoretical revolution in physics. Within this broader geometrical framework, Newton's second law of motion in effect reduces to Newton's first law of motion: all matter now moves uniformly along the straightest possible lines in the surrounding (generalized) curved geometry. Accelerated, non-geodesic, motions, no longer really exist in this fuller geometry.

(vii) As a by-product of this inductive simplification it became clear that Einstein's theory of gravitation had itself not gone quite far enough. For mathematically, although Einstein had made the curvature a function of the gravitational sources, there was an analogous tensor, the torsion tensor, which was still required to vanish everywhere and played the role of a flat background geometry. And this feature of space-time geometry was related in precisely the same way to the six-parameter sub-group of rotations and boosts in the ten-parameter Poincaré group underlying special relativity as the ordinary curvature was related to the four-parameter Abelian sub-group corresponding to translations in space and time. Einstein's theory was the gauge theory of the latter subgroup, not of the whole group. Einstein's argument from cause-effect reciprocity for space-time curvature depending on the distribution of matter was equally applicable to space-time torsion, and taking this consideration seriously generates a further twenty-four equations determining the torsion in addition to Einstein's sixteen field-equations determining the curvature. The result is the so-called U4 theory, or Einstein-Cartan-Sciama-Kibble theory, of gravitation. It is less mathematically arbitrary than Einstein's original theory. It also allows elementary particles such as fermions to function as sources of gravitational fields, which was not really possible in Einstein's original theory since the the natural energy-momentum tensor for fermions is anti-symmetric rather than symmetric as would be required by Einstein's original unmodified field equations. However this new less arbitrary theory still yields the same predictions as Einstein's original theory as far as ordinary macroscopic gravitational effects are concerned.

(viii) However even with these improvements physical theory is still not formally as simple as one might reasonably expect. For it still contains arbitrary coupling constants. In particular the gravitational coupling constant has to be regarded as a real-valued parameter equivalent to an infinite-bit ingredient in the theory, and there is nothing in the orthodox theory to explain why this parameter is not zero, i.e. why gravitation exists at all. However there is a recent programmatic theory which would overcome just this difficulty, and explain why gravitation exists at all and has the strength it does, namely Super-string theory. In fact Super-string theory not only eliminates various otherwise unavoidable infinities in the quantum theory of the other forces, but has no consistent solutions which do not include Einsteinian gravitation, and requires a non-zero gravitational coupling constant. The empirical value of this coupling constant should actually be calculable within this theory. Unfortunately the theory is not yet well enough understood for us to be able actually to carry out this calculation. Super-string theory remains to this extent still a merely programmatic theory. But should it prove right we will have to conclude that Newton was wrong in concluding that gravity was not an essential property of matter, and thus wrong in

https://doi.org/10.1086/psaprocbienmeetp.1990.2.193068 Published online by Cambridge University Press

concluding on that basis that the presence of gravity in our world required a free creative act by an omnipotent Deity.

4. Discussion

What I want to emphasize about this brief review of the subsequent history of Newton's theory is that every change can in fact be viewed as one of formal simplification. Things do not seem like this to the layman because the layman does not realise how complicated the mathematical formalization has to be of all the background assumptions about the world which he takes for granted, and how much mathematical arbitrariness is implicit in naive formalizations of these background assumptions. But with the benefit of deeper mathematical understanding, we can see that arbitrarinesses here can be eliminated simultaneously with the elimination of arbitrarinesses in Newton's original explicit theory: such non-evident and evident arbitrarinesses can be made to cancel each other out, yielding what is overall a succession of mathematical simplifications of the theory of the world. This is what has happened in physics so far, and it is reasonable to suppose that it will continue in the future. (The most natural inductive inference here would be to the conclusion that Galileo and Einstein were right in thinking that our actual universe will ultimately turn out to be the simplest possible physical universe.)

This history creates, however, the following problem for deduction from the phenomena arguments, even for the most "impressive" deduction from the phenomena arguments. We would have liked these to yield actual high rational probabilities for the theories to which they lead. At first it seems they must do this because those theories are deduced from background assumptions which seem to have high probabilities. Any known rival theory can be shown to be inconsistent with one or other of these background assumptions and thus can be shown to be less probable in the then state of knowledge.

But the trouble is that such background assumptions must go well beyond the actual evidential data. And while they seem at the time to be the simplest mathematical generalizations consistent with that data (and were this so, this would entitle them to high Solomonoff-Levin rational probabilities), nevertheless subsequent mathematical investigation has always shown that there were formally simpler alternatives which had been overlooked, and which, had theorists known of them at the time, would have thus to have been assigned higher rational probabilities than the alternatives actually chosen.

Underlying this situation is a fundamental mathematical feature of the new inductive logic. Relative rational probabilities for any rival theories known to predict the same data can always be computed. It is enough to write those theories as programs and to count the number of program-bits required for each theory. Absolute rational probabilities are a different matter.

We would know these too if we knew that there were no simpler theories predicting the same data. For Solomonoff-Levin rational probabilities fall off fast enough with numbers of additional bits, for more complicated theories, even the disjunction of all more complicated theories, generally (though there are occasional exceptions) to get a low relative rational probability. So the problem is not the Popperian problem of the universe actually probably turning our more and more complicated.

The problem is the reverse. Absolute rational probabilities can only be assigned, or bounded from below, if we know there are no simpler theories predicting the same data. And we cannot ever know this because for any realistic data there can be no al-

https://doi.org/10.1086/psaprocbienmeetp.1990.2.193068 Published online by Cambridge University Press

gorithm for determining the shortest program which will regenerate that data. Absolute rational probabilities are thus non-computable. This would not matter if we had meta-inductive evidence that we were actually good at finding the shortest possible encodings of realistic data. If we could show that modulo mathematical miracles concerning as yet uncomputed sequences of digits in the development of π, we had every reason to believe that we were often in practice actually succeeding in determining the shortest encodings of realistic data.

Unfortunately the existing meta-inductive evidence points in precisely the opposite direction. Each generation of theoretical physicists discovered that its predecessors had missed what were actually fewer-bit encodings of all the existing data. For this reason even the most impressive deduction-from-the-phenomena justifications, contrary to Newton's own inductive hopes, fail to yield high rational probabilities. At best we can take the conclusions to which such arguments lead as the most preferred of the available theories, and as theories which can be expected to yield correct predictions in the domain in which their background assumptions remain reasonable extrapolations from the data.

We indeed have some a priori mathematical guide as to when we are likely to be nearing the boundaries of such domains. Namely when the value of some physical quantity begins to approximate to the value of what appears to be a fundamental constant in the theory in question or in some related theory. For it is precisely at such points in a theory, or in the relation between two theories, that we expect deeper mathematical insights to yield future changes resulting in potential overall theoretical simplification.

Thus in practice it is reasonable to suppose that when an unexpectedly simple theory is deduced from the phenomena it will remain a very good approximation to the truth in a very extended domain. But the new inductive logic warns us that we cannot conclude that it is probably true, and meta-induction from the history of physics teaches us that it is almost certainly false, not because the truth will turn out to be more complicated, but because the truth will prove to be even simpler when all the relevant data that were available are taken into account.

The Newtonian strategy gives us a way of establishing theories which are rationally more probable than any rivals we are likely to be able to envisage given the current state of theoretical understanding. But at the same time the subsequent history of physics warns us that we will almost certainly be failing to envisage rationally even more probable theories.

The moral seems to be that physicists should spend more time reflecting on the foundations of current theoretical frameworks, than on tinkering with theories to explain particular recalcitrant experimental results. Every feature which is normally taken for granted in our currently successful theories needs to be repeatedly called in question, because we can be almost certain that there will be simpler future theories obtainable by abandoning some such features.

Extrapolating a little further from the history of physics, it is hard to avoid the inductive conclusion that our actual universe is likely ultimately to prove to be, in some meaningful sense, the simplest possible universe. This suggests that a more direct and a priori approach to characterizing the latter structure (e.g. by first laying down informal adequacy conditions on the class of mathematical structures which could count as characterizing possible physical universes, and then looking for the simplest mathematical structure which could meet those conditions) might eventually deliver the jackpot.

https://doi.org/10.1086/psaprocbienmeetp.1990.2.193068 Published online by Cambridge University Press

Thus the results of three centuries of reasoning from the phenomena are more surprising than philosophers seem to realize: for they seem to imply that we may be unduly neglecting a potentially viable alternative more aprioristic strategy.

References

Dorling, J. (1987), "Schrödinger's original interpretation of the Schrödinger equation: a rescue attempt," Schrödinger, *Centenary Celebrations of a Polymath*, C.W. Kilmister (ed.). Cambridge: Cambridge UP.

______. (1991), "Einstein's methodology of discovery was Newtonian deduction-from-the-phenomena," *Scientific Discovery* (provisional title), J. Leplin (ed.), University of California Press (forthcoming 1991).

Li, M. and Vitányi, P.M.B. (1991), "Inductive reasoning and Kolmogorov Complexity," *Proceedings of the 4th Annual IEEE Structure in Complexity Theory Conference 1989.*

https://doi.org/10.1086/psaprocbienmeetp.1990.2.193068 Published online by Cambridge University Press