

MALCOLM R. FORSTER *

UNIFICATION, EXPLANATION, AND THE COMPOSITION OF CAUSES IN NEWTONIAN MECHANICS

Abstract—William Whewell’s philosophy of scientific discovery is applied to the problem of understanding the nature of unification and explanation by the composition of causes in Newtonian mechanics. The essay attempts to demonstrate: (1) The sense in which ‘approximate’ laws (e.g. Kepler’s laws of planetary motion) successfully refer to *real* physical systems rather than to (fictitious) idealizations of them; (2) why good theoretical constructs are not badly underdetermined by observation; and why, in particular, Newtonian forces are not conventional; and (3) how empiricist arguments against the existence of *component* causes, and against the veracity of the *fundamental* laws, are flawed.

I. Introduction

WHEN DOES a successful scientific explanation give us reason to believe in the theoretical ‘world’ it employs? It is possible to give two extremely radical answers to this question. The first is the logical positivist view, which says that scientific explanations *never* give us reason to believe in the theoretical ‘world’. As Friedman (1981) points out most clearly, the positivist need not insist that we must *eliminate* theoretical constructions from science—rather, he may view the theoretical structure used by our best scientific theories as a mere *mathematical representation* of true *empirical* (observable) facts. At the other extreme, we have what Friedman refers to as a very *liberal* and *cavalier* attitude towards the theoretical constructions of the mathematical sciences: “A theoretical explanation gives us reason to believe in its postulated structure (or at least provisionally accept it) whenever it is the *best available* explanation of that phenomenon.” The working assumption of this essay will be that both of these *extreme* viewpoints are wrong. What we need is a more *moderate* realist point of view that holds some middle ground between these two extremes. Some theoretical constructions should be taken seriously as representing aspects of an *unobserved* (and maybe unobservable) reality behind the phenomena, while others may be seen merely as devices for representing the *empirical* data.¹

* Department of Philosophy, University of Wisconsin, Madison, WI 53706, U.S.A.

¹ Friedman’s 1983 book is a detailed account of how this distinction might be drawn in terms of coordinate and coordinate-free representations of space-time theories.

Quite recently, Cartwright (1983) has developed the view that such a distinction should rest on the differences between *causal* and *non-causal* forms of explanation. That is, we should believe in the existence of *causes* postulated by our best causal explanations under certain conditions, but not in the constructs of non-causal explanations (such as those appearing in the *fundamental* laws of physics). Her motto is: “inference to the most likely *cause*” rather than the more liberal and cavalier rubric of “inference to the best explanation”.² Cartwright argues for this difference of attitude on the basis of the very reasonable view that explanation leads to inference in just those cases in which the *explanans* (the thing doing the explaining) is singled out as the only one capable of doing the job. That is, we should believe in the existence of those theoretical entities whose magnitudes are *determined* in some sense by the observed facts they are posited to explain. In such cases, we cannot change the values of the theoretical functions, such as ‘mass’, ‘force’, etc., without getting the explanation wrong. So then the theoretical ‘world’ employed by the explanation of the facts is essential to the success of that explanation, and should be taken seriously.

In Cartwright’s opinion, this “criterion of reality” argues against the existence of component forces: Only the existence of resultant, or composite causes, can be inferred from the best causal explanation. For example, the component gravitational force acting on this cup sitting on the table is counteracted by an equal and opposite force of the table on it, which together produce a zero effect (zero acceleration). By observing the cup’s behaviour (its zero acceleration) we can infer the size of the *resultant* force acting on it (being zero) but this observation fails to determine the size of each component. We might express the fact of the cup’s weight being 2 newtons in terms of a counterfactual conditional: “If gravity were the only force acting then the acceleration would be such and such.” But this does not solve the problem because this conditional does not express an *observable* fact about the cup. The truth value of the counterfactual is not determined by what is observed. So, we could equally well suppose that a gravitational force of 20 newtons is counteracted by an equal and opposite force of the table on the cup and still ‘account’ for the same observational prediction that the cup does not move. Cartwright concludes that positing the existence of a component gravitational force acting on the cup is therefore *redundant* to the success of the explanation, so we should not believe in any such construct.

Ellis (1965) has further argued that even explanations of events in terms of *resultant* forces fail to satisfy the non-redundancy requirement for theoretical inference.

² I have also heard Brian Ellis support Cartwright’s views on this.

In particular, there is an element of *convention* in deciding how to measure the effect of resultant ‘forces’. Different conventions are possible, so this inference also fails to satisfy the non-redundancy requirement. Therefore, we have no justification of the existence of any forces what-so-ever.

While this essay will accept the non-redundancy requirement as a useful “criterion of reality”, I will argue that Cartwright’s and Ellis’s application of it is overly restrictive. Ellis and Cartwright only consider *causal explanations of single events*, viz. the acceleration of a body at a particular time. But we will see that in the history of astronomy at least, theoretical “causes” are actually posited as the causes of global *phenomena*, which are observed as statistical regularities in the data and described by phenomenological laws. The phenomena observed are the global “effects” of unobserved “causes”.³

William Whewell’s philosophy of science makes precise sense of this now unpopular mode of parlance. It is the *coefficients* in the phenomenological laws of the mathematical sciences that represent the “causes” of the phenomena. Thus the “causes” of refraction phenomena are represented by the refractive indices appearing as coefficients in Snell’s law and the “causes” of gravitational phenomena are represented by the mass coefficients in Newton’s inverse square law of gravitation. And the “causes” of selection phenomena exhibited in the evolution of biological populations are represented by the fitness coefficients in the set of equations relating rates of growth as functions of the numbers of each genotypic subpopulation. The fitting of these equations to the data by statistical methods provides us with theoretical measurements of these coefficients and a way of quantifying the “causes” of the phenomena. The mathematical model employed receives its confirmation when the “causes”—so measured—connect in various ways with the “causes” of other phenomena; forming a *consilience of inductions*. It is when this happens that a merely phenomenological representation of the empirical facts turns into a *bone fide explanation* of the phenomena. Then we start to take the postulated “causes” more seriously. In section 5, the detailed example of modern planetary astronomy will speak against the anti-realist conclusions of Duhem, Ellis, and Cartwright in favour of a moderate realist view of the history of science.

II. Cartwright's Argument Against Component Forces

The general thrust of Cartwright’s argument against the reality of component causes might be reconstructed as follows: (1) Inference to the best explanation

³ The explanation of “global facts” is infrequently discussed explicitly in the literature, but two notable exceptions are that I know of are Friedman (1974), and Hooker’s response (1980).

works only for causal explanations of single (token) events. (2) Inference to the best explanation works only for explanations satisfying the non-redundancy requirement. Therefore; (3) At most one cause can be inferred to exist from the best causal explanation of any single event, for otherwise, the magnitude of the cause cannot be determined from the magnitude of its effect. (4) Explanations by composition of causes posit two, or more, component causes to explain the occurrence of a single event. Therefore; (5) We can never (justifiably) infer the existence of component causes.

The first question that needs answering concerns premise (1). What does Cartwright have in mind when she speaks of inference to the best *causal* explanation, or inference to the most likely cause? She states that:

Causal reasoning provides good ground for our beliefs in theoretical entities. Given our general knowledge about what kinds of conditions and happenings are possible in the circumstances, we reason backwards from the detailed structure of the effects to exactly what characteristics the causes must have in order to bring them about.⁴

But, on page 29, she frankly admits that she “will not offer a model of causal explanation,” but will give an account of *causal laws* from which “certain negative theses follow”. One of these negative theses is of particular interest to us. Falling under a causal law (a kind of *statistical* law between *types* of event) such as $C \rightarrow E$, where C and E label *types* of events, is not sufficient to explain an instance, or token, of E because “a single phenomenon may be in different domains of various causal laws, and in many cases it will be a legitimate question to ask, ‘Which of these causal factors actually brought about the effect on this occasion?’” There may be two (statistical) causal laws applicable to events of type E , say $C_1 \rightarrow E$ and $C_2 \rightarrow E$, and either an event of type C_1 or an event of type C_2 (but not both) caused the E event *on this particular occasion*. For example, experimentation on a large number of lemon trees may show that saturating the roots in water causes the trees to die ($C_1 \rightarrow E$), and that applying defoliant also causes them to die ($C_2 \rightarrow E$). But the cause of death for any *particular* tree is either water saturation or the defoliant, not both, even when the tree is both water-logged and sprayed with defoliant. Causal *laws* (being statistical in nature) hold between event *types*, while Cartwright is thinking of *causes* as event *tokens*. So, to explain an event, according to Cartwright, we need to refer to the *particular* event that caused it *on that occasion*. Causal laws do not give us that information, and so falling under a causal law is not sufficient to explain the *event*.

No matter how long we stare at the dead lemon tree, we may not be able to tell whether it died from the defoliant sprayed on it or from the water saturation of its

⁴ Cartwright (1983), p. 6.

roots. On the surface both alternative explanations are equally good, so both explanations fail the no-redundancy requirement. This consideration motivates the introduction of premise (2) which says, in this context, that the cause should be *determined* by its effect. There is a clear statement of this by William Whewell worth quoting:

Axiom II. *Causes are measured by their effects*

Every effect, that is, every change in external objects, implies a cause, as we have already said: and the existence of the cause is known only by the effect it produces. Hence the intensity or magnitude of the cause cannot be known in any other manner than by these effects: and, therefore, when we have to assign a measure of the cause, we must take it from the effects produced.⁵

For this to be satisfied, it is necessary that there is at most one cause to each effect. For if there were two magnitudes to be inferred from the same observed event, the two magnitudes would not be independent of one another, and we should say that both magnitudes are measures the *same* cause. By observing only that the lemon tree is dead, we can infer only the occurrence of the disjunctive event that *either* it was sprayed with defoliant *or* it had water-logged roots. Similarly, in the cup example, it can only be the *composite* or *resultant* force that is measured by the magnitude of the effect (the acceleration of the cup). Cartwright therefore allows that the *resultant* force acting on the cup is real, but that there is no justification for believing that the *component* gravitational force is 2 newtons, since this force has no separate effect of its own.

Hume and Mill are two well known philosophers who have views on causation almost identical to Cartwright's, and yet they disagree with her about the composition of causes. Both take the relata of the causal relation to be particular events, and both *define* the cause of an effect to be unique. As Hume puts it, "The same cause always produces the same effect and the same effect never arises but from the same cause".⁶ This implies, among other things, that if a force acting due north produces motion due north in one instance, the *same* force cannot be a cause of motion north-east in another instance, because these effects are different. What causes the north-east motion is a north-east force. Yet Hume and Mill allow for the existence of component *causes* by introducing the notion of *component effects*. Thus, when the force due north is combined with the force due east, it still produces its motion due north as a *part* of the total effect. Causes have parts because effects have parts:

⁵ Butts ed. (1968), p. 81.

⁶ Hume's *Treatise*, Book I, Part III, Section XV.

When an object increases or diminishes with the increase or diminution of its cause, it is to be regarded as a compounded effect, derived from the union of the several different effects which arise from the several different parts of the cause.⁷

So, a single effect can have many causes because a single effect can have many parts. Mill's agrees with this position in examples of mechanical phenomena:

In this important class of cases of causation, one cause never, properly speaking, defeats or frustrates another; both have their full effect. If a body is propelled in two directions by two forces, one tending to drive it to the north and the other to the east, it is caused to move in a given time exactly as far in both directions as the two forces would separately have carried it, and is left precisely where it would have arrived if it had been acted upon first by one of the two forces and afterward by the other.⁸

After quoting this passage, Cartwright confesses that she is totally unconvinced by this point of view.⁹ The problem for her is to see what *empirical* justification there is for the existence of causal parts. The strategy of allowing for causal parts by decomposing the effect into parts simply transfers the question to a problem about the decomposition of effects. *If* we know that forces compose by the law of vector addition, and *if* we know that there are only two component forces, and we know their directions, then we can infer the magnitudes of those component forces from the magnitude of the resultant. But this fact does not solve Cartwright's problem. There are many pairs of directions along which the resultant vector can be resolved; the problem is to give an empirical criterion for choosing among them. And in the cup case, even if we take the two directions for the decomposition of the resultant (up and down) as given, we still cannot uniquely determine the magnitude of the components in this special case. So, the problem of giving an account of how to infer the magnitudes of component forces has not been solved.

While Cartwright's skeptical conclusion sounds quite convincing in these cases, it is important for her to generalize the argument to include historically more important examples. For this reason she frames the argument in more general and familiar terms. In a nutshell, she alleges that there is a fundamental trade-off between the two demands on inference to the best explanation traditionally made by realists. The first demand is that the best explanations are those that make essential use of the most general and fundamental laws, which 'cover' a wide variety of different phenomena. She agrees that this should be required of an adequate explanation. But this requirement conflicts, in Cartwright's view, with the second realist desideratum—the

⁷ *Ibid.*

⁸ J. S. Mill, *System of Logic*, Book III, Ch.VI..

⁹ Cartwright (1983), pp. 60-61.

‘facticity’ requirement that the best explanations should correctly describe *how bodies behave*. “Really powerful explanatory laws of the sort found in theoretical physics do not state the facts”.¹⁰ And conversely:

Many phenomena which have perfectly good scientific explanations are not covered by any laws. No true laws that is. They are at best covered by *ceteris paribus* generalizations - generalizations that hold only under special conditions. The literal translation is ‘other things being equal’; but it would be more apt to read ‘*ceteris paribus*’ as ‘other things being right.’ ... *Ceteris paribus* generalizations, read literally without the ‘*ceteris paribus*’ modifier, are false. They are not only false, but held by us to be false; and there is no ground in the covering-law picture for false laws to explain anything. On the other hand, with the modifier the *ceteris paribus* generalization may be true, but they cover only those few cases where the conditions are right. For most cases, either we have a law that purports to cover, but cannot explain because it is acknowledged to be false, or we have a law that does not cover. Either way, it is bad for the covering-law picture.¹¹

The structure of Cartwright’s argument is easily understood in terms of our previous examples. If we want to infer that the lemon tree died from water saturation of its roots, we may appeal to a law that says “Water saturation of roots kills lemon trees.” While this law (as stated) is completely general, it does not always hold true in cases in which the tree has been sprayed with a defoliant. In some such cases, the defoliant will be the cause of death. To make the law true we must add a *ceteris paribus* modifier: “If no defoliants have been applied, then water saturation of roots kills lemon trees.” But now the law is no longer general in that it does not cover the cases in which defoliant has been applied.

The same argument can be used to question the status of Newton’s law of gravitation as it is used to explain the behaviour of our cup. In its general form, the law of gravitation states that “two bodies exert a force between each other which varies inversely as the distance between them, and varies directly as the product of their masses.” According to that formula, suppose that we calculate the component gravitational force on the cup towards the earth to be 2 newtons. But is this law true in this particular case. Cartwright claims that it is not because this general law does not satisfy the ‘facticity’ requirement. It does not correctly predict that the acceleration of the cup towards the earth is zero. The reason is that there is another force acting on the cup—the normal force applied by the table. So, to make the law true we should add a *ceteris paribus* modifier of the following form:

¹⁰ *Ibid*, p.3.

¹¹ *Ibid*, pp.45-46.

*If there are no forces other than gravitational forces at work, then two bodies exert a force between each other which varies inversely as the distance between them, and varies directly as the product of their masses.*¹²

In this instance, the modifier only demands that *non-gravitational* forces are absent, but in fact Cartwright should make a far stronger qualification than that. For her argument - in its most general form - denies a realist attitude towards the decomposition of any resultant gravitational force into *gravitational* components as well. For instance, if we consider the 3-body problem in celestial mechanics in which the moon is attracted simultaneously by both the sun and the earth, the law of gravitation tells us the magnitude of both components. These two component forces then combine to produce the *resultant* motion (acceleration), and neither instance of the law in isolation tells us the how the moon behaves. The *ceteris paribus* modification of the law should therefore read:

If there are no other forces at work (of any kind), then two bodies exert a force between each other which varies inversely as the distance between them, and varies directly as the product of their masses.

Only this modified law satisfies the facticity requirement. In this way, we see that the problem that Cartwright raises is a very serious problem even *within* the domain of autonomous disciplines such as planetary astronomy. Or, to put the point in another way, once we solve the problem within planetary astronomy, the resolution of the problem Cartwright actually considers will be straightforward.

Of course, we can ‘visualise’ the reality of component forces in terms of *tendencies* (Mill),¹³ *causal influences* (Creary, 1981), or *powers* (Hume), or other metaphysical entities, but such a intuitive metaphysical description achieves no more than our ‘counterfactual’ formulation of the weight of the cup in terms of how it *would* behave *were* the table not supporting it. The fact remains that the table is supporting the cup. The problem with causal powers, or tendencies, or influences, is well known: they lack the ‘proper’ grounding in empirical fact: “Hume taught us that ‘the distinction, which we often make betwixt *power* and the exercise of it, is ... without foundation’. It is just Hume’s illicit distinction we need here: the law of gravitation claims that two bodies have the *power* to produce a force of size $Gm'm/r^2$ they do not always succeed in the *exercise* of it”.¹⁴ And in reference to Creary’s (1981) casual ‘influences’, Cartwright emphasizes that she is “not opposed to them because of any

¹² *Ibid.*, pp.57-58.

¹³ *System of Logic*, Book II, Ch.X.

¹⁴ Cartwright (1983), p.61.

general objection to theoretical entities,” but rather because she thinks that “every new theoretical entity which is admitted should be grounded in experimentation, which shows up its causal structure in detail”.¹⁵ The problem with introducing causal powers, tendencies, or influences, is that we then end up denying the facticity of laws.

Cartwright, like Hume, simply embraces this anti-realist conclusion. She allows that fundamental laws *are* about causal powers, causal influences, causal parts, or what-have-you, but only while insisting that these are entities referred to in the fictitious *models* and *idealizations* in which the *ceteris paribus* conditions of our laws are satisfied. This is how we explain and organize the facts of our experience on Cartwright’s view.

The realist, who wants to resist this conclusion, seems to face a dilemma; if he accepts the causal powers story (or something like it), as well as the *truth* of fundamental laws as being about elements of reality, he must repudiate the moderate empiricist demands of confirmation and testability. Indeed, realists nowadays do commonly denounce all loyalty to any version of empiricism. But in this essay, I hope to show how he can meet the empiricist challenge head on. As indicated earlier, the argument will be that Cartwright’s initial premise—that inference to the best explanation works only for single (token) events—is too restrictive. If we construe “causal” explanations more broadly as explaining *global facts* and statistical *regularities*, then such explanation does satisfy the non-redundancy condition in such a way as to justify the existence of component “causes”, and eventually the existence of component forces. Before developing this idea, however, the next section will continue to formulate the empiricist challenge even more ‘forcefully’ in terms of Ellis’s celebrated argument for the conventionality of forces.

III. Ellis's Argument for the Conventionality of Forces

One possible reply to Cartwright’s argument against component forces involves an appeal to an intuitive account of inductive inference: We *can* infer the existence of ‘component’ forces when no other forces are acting and the ‘component’ force is equal to the resultant force—so why not *generalize* the inference to all situations what-so-ever if we can do so without contradiction? That we can *consistently* make such generalizations is proven by the consistency of the metaphysical picture of component forces as influences, powers, or tendencies of the last section. There are two things wrong with this reply. The first is that it tacitly assumes that we have a general account of the *rules* by which such generalizations are made. It assumes that

¹⁵ *Ibid.*, pp.66-67.

we have a defensible solution to the infamous problems of induction. If we did, then a solution to Friedman's problem (that of deciding which features of our explanatory theories represent a reality beyond the phenomena) would be at hand and that would decide the issue. The problem is that the realists do not seem to have an adequate account of induction.

Secondly, there is another well known argument in the literature (Ellis, 1965) which purports to show that such a simple account of inductive generalization would not suffice to solve the problem anyway.¹⁶ Ellis (1965) has argued that even resultant forces do not exist because the *effects* that we use to measure them are not determined solely by what we observe but partly by *convention*. We measure resultant forces by the accelerations, but accelerations are determined relative to some frame of reference, and our choice of reference frame is conventional. In his 1976 paper, Ellis endorses a reconstruction of his (1965) argument for the conventionality of forces due to Hunt & Suchting (1969) as follows:

- (1) The distinguishing feature of forces generally is that in some sense their existence entails and is entailed by the existence of effects they are supposed to produce. Or, necessarily, there is an *X*-force *if* there is an *X*-effect of that force.
- (2) There is always an element of convention in deciding what we should regard as an effect.
- (3) Hence, to the extent to which there is this element of convention, the existence of forces is also conventional.

The strength of Ellis's position rests on his second premise, which he supports with clear and interesting arguments. First, he discusses Newton's law of inertia, which states that if no force is present a body will continue to move in a straight line with uniform speed. The contrapositive of this tells us that when any body accelerates towards us, then there exists an impressed (gravitational) force. But, asks Ellis: "What reason have we, independent of the law of inertia, for saying that gravitational forces exist?"¹⁷ To say that we know that gravitational forces exist because celestial bodies or terrestrial projectiles do not move with uniform motion in a straight line is "obviously to beg the question." For the only obvious way to define an inertial frame is in reference to a particle subject to no forces. But this "is simply to assume the

¹⁶ Brian Ellis no longer adheres to the conclusions of his 1965 paper, and now seems to hold a view closer to that argued in this paper (Ellis et al, 1986). But there is still considerable disagreement between the conclusions of this essay and Ellis's general philosophical position, as described in his 1985 essay for example.

¹⁷ Ellis (1965), pp. 41-42.

truth of the law whose truth we wish to establish.” The allegation is that the law of inertia has no empirical content—*i. e.*, that it is *tautological* in this sense.

To focus his argument, Ellis presents an *alternative* law of inertia as follows: “Every body has a component of relative acceleration toward every other body in the universe directly proportional to the sum of their masses and inversely proportional to the square of the distance between them - *unless it is acted upon by a force*”.¹⁸ The concept of ‘force’ implicitly defined by this law is different from Newton’s ‘definition’, yet the theory obtained from it is *empirically* indistinguishable from Newton’s. Thus, Ellis concludes that the postulation of Newtonian forces in the standard mechanical model of the world is merely a *redundant* explanatory device—nothing more than a convenient mathematical fiction. There is no *empirical* justification for believing in the existence of Newtonian gravitational forces. This conclusion is even stronger than what Cartwright wants, for it argues against the existence of *all* gravitational forces, whether they be component forces or resultant forces.

The intuition behind Ellis’s argument is easy to explain, and easy to generalize. Ellis’s idea is that “a system is acted upon by a force (or forces) if and only if we consider that the system persists in an unnatural state or that it is changing in an unnatural way.”¹⁹ The conventionality of forces results from the conventionality of deciding what is to be regarded as a *natural state of motion*. In the case of gravitational phenomena, a natural state of motion is defined in terms of a frame non-rotating with respect to the fixed stars and centered near the sun, and this choice is conventional, argues Ellis. The same argument can be applied to other ways of measuring gravitational forces. For instance, if we modify the cup example slightly, and imagine that the cup stretches a *spring* by a certain amount, it might be thought that we then have an objective determination of its weight from observing the amount that the spring is deformed. But the observed deformation of the spring is actually the deformation of the spring *relative to its natural length*. But the *natural* length of the spring is defined as its length *when no forces are acting*, and so the concept of “natural length” *presupposes* the criterion for the existence of forces we are trying to explicate. We can only avoid this vicious circularity by conventional fiat. The effect of the cup on the spring is measured as the *difference* between its extended length and its ‘natural’ length, but what ‘natural’ means in this context is a matter of convention. Similarly, we measure the acceleration of a projectile relative to a conventionally chosen frame of reference. Forces are conventional because effects are conventional. So, forces do not exist.

¹⁸ *Ibid.*, p.49.

¹⁹ *Ibid.*, p.45.

The observed facts do not seem to uniquely determine what unobserved ‘forces’ should explain them. The standard Newtonian explanation in terms of Newtonian ‘forces’ is a *redundant* explanation of how bodies behave, since we could just as well use Ellis’s ‘forces’ as defined from his “law of inertia” to explain the same facts. So, the Newtonian explanations of mechanical phenomena appear to violate the non-redundancy requirement of a moderate empiricism, and such explanations are not good enough for inferences to the best explanation. Or so it seems.

It is implicitly assumed in Ellis’s argument, as in Cartwright’s, that a force is introduced into scientific discourse in order to explain the occurrence of *single* spacio-temporally localized events, such as the instantaneous acceleration of a particular body towards the earth, or the extension of a certain spring at a given time. I do not deny that forces, and other theoretical entities, explain such *events*. And I agree with Ellis and Cartwright that *this* explanatory role is not strong enough to justify the existence of forces. But Ellis and Cartwright seem to overlook is the possibility that theoretical “causes” may be justifiably introduced to explain the global physical “effects” described by phenomenological laws. For example, the introduction of the property of gravitational mass into mechanics is justified by its role in the explanation of various *statistical regularities* in the astronomical data. Although forces themselves are not “causes” in this sense, their justification flows from the success of this more fundamental type of “causal” explanation. To make sense of all this, we will return to the work of the nineteenth century philosopher and historian of science, William Whewell.

IV. Whewell's Philosophy of Scientific Discovery

The fundamental issue at hand is the status of the mathematical constructions that scientists introduce to explain the phenomena recorded by their measuring instruments. These phenomena may be created in the laboratory or they may occur naturally as in the case of astronomy. The question is: How are our theoretical entities (such as ‘mass’, ‘force’, etc.) constructed, or ‘defined’, in terms of the values of the variables directly recorded in the laboratory or the observatory?

The aim of this section is to show that an answer to this question already exists in the writings of William Whewell, who is best known for his notions of *consilience of inductions* and theoretical unification. Whewell’s account of scientific discovery will provide us with a philosophical theory of scientific explanation and inference that will hold middle-ground between the austere anti-realism of the logical positivists and the cavalier and overly liberal attitudes of many modern-day realists.

4.1: *The Colligation of Facts*

According to Whewell, “the Colligation of ascertained Facts into general Propositions” may be considered as containing three steps, which he labels as (1) the *Selection of the Idea*, (2) the *Construction of the Conception*, and (3) the *Determination of the Magnitudes*. And in mathematical investigations, these three steps correspond to (1) the determination of the *Independent Variable*, (2) the *Formula*, and (3) the *Coefficients*, ... or (1) the *Argument*, (2) the *Law*, and (3) the *Numerical Data*, in a Table of an astronomical or other *Inequality*.²⁰

Upon the *selection of the independent variable*, we may plot many pairs of values of the dependent and independent variables on a ‘scatter diagram’. The second stage of the process is to *select the formula*, defining a family of functions such as the family of all linear functions, which then connects the dependent variable with the independent variable. The formula might be written as a generic equation, such as $Y = a.X + b$, where Y is the dependent variable and X is the independent variable, and a and b are undetermined coefficients. The third step in the colligation of facts is to *determine the magnitudes* of these coefficients by finding the closest fitting curve of the specified family of curves by employing, say, the method of least squares. The values of a and b for this closest fitting curve determine the magnitudes of the coefficients. These three steps are familiar to any modern researcher in the empirical sciences.

Because Whewell himself did some important empirical work on the tides,²¹ he frequently uses this as an illustration of how the colligation of facts works. The aim of tidalogy is to discover the laws governing the height of successive high waters, H , at a particular place. The first step in a colligation of these facts is to name an *independent variable*. In this example, one such variable is the distance from syzygy of the Moon, D (the angle between the Moon and the Sun as seen from the Earth), which increases from 0 to 360 degrees about once every 4 weeks. When the height of high waters, H , is plotted on a graph against this variable, D , a roughly sinusoidal variation in H is detected of a period of 180 degrees, with maximums at approximately $D=0$ ($=360$) and $D=180$. In this case (of fitting a sinusoidal curve to the data) the coefficients will be the amplitude, period, and the phase, of the fitted curve. (See Fig. 1.)

Clearly, this sinusoidal curve will not fit the data exactly; there will always remain an *unexplained* variation of the dependent variable (H) above and below the best

²⁰ Butts ed. (1968), pp. 210-211.

²¹ *Ibid*, p.3.

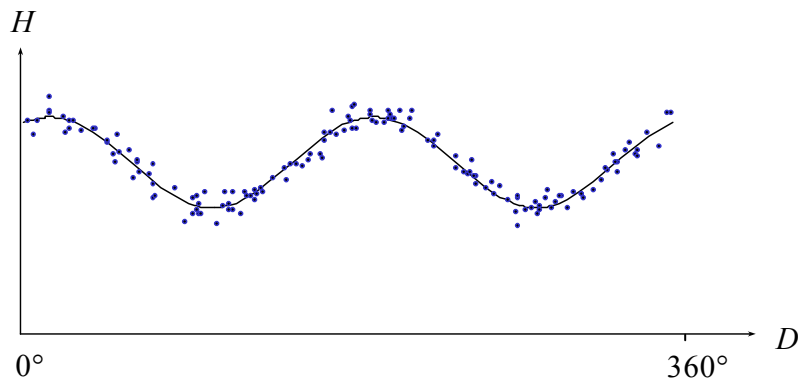


Fig. 1. The (fortnightly) variation of high-water tidal marks with respect to the phases of the moon: The semi-menstrual inequality. The error is Gaussian.

fitting curve; this residue being completely *capricious*, or random, with respect to the independent variable recorded. For a sufficiently large number of observations, the *mean* value of H for a fixed value of D will lie on (or very close to) the curve, even though the particular *observed* values of H will not. In fact, this is one method of determining what the curve should be, called the Method of Means. Whewell mentions that Lubbock's first investigations of the laws of the tides of London used this method on over 13,000 observations extending through nineteen years; it being considered that this large number was necessary to remove the effects of accidental causes.²² In this way, "the Method of Means gets rid of irregularities by taking the arithmetical mean of a great number of observed quantities".²³ Later, Whewell writes:

The Argument [independent variable] being thus assumed, the Method of Means is very efficacious in ridding our inquiry of errors and irregularities which would impede and perplex it. Irregularities which are altogether accidental, or at least accidental with reference to some law which we have under consideration, compensate each other in a very remarkable way, when we take the Means of *many* observations. If we have before us a collection of observed tides, some of them may be elevated some depressed by the wind, some noted too high and some too low by the observer, some augmented and diminished by un contemplated changes in the moon's distance or motion: but in the course of a year or two at the longest, all these causes of irregularity balance each other; and the law of succession, which runs through the observations, comes out precisely as if those disturbing influences did not exist.²⁴

In this case, and in the examples to follow, the method of means works because the other variables upon which H systematically depends (such as the Parallax of the

²² *Ibid*, p. 233.

²³ *Ibid*, p. 233.

²⁴ *Ibid*, pp. 231-232.

Moon, or its Declination) vary with time periods quite different from the fortnightly variation of the Distance from Syzygy (or any simple multiples of it). But given that this condition is satisfied, it is important to note that the success of the method does *not* depend upon the ‘unexplained’ variation being small in comparison with the variation ‘explained’ by the law. Whewell cites the example the diurnal oscillations of the barometer, which are very much smaller than the “errours by which they are encumbered and concealed,” which are “hitherto reduced to no law.” But “the result was a clear and incontestable proof of the existence of such oscillations”.²⁵ It is just as well that the existence of an unexplained ‘residue’ does not preclude the possibility of a law, for otherwise there could be few laws to be found in the quantum mechanical domain, in which there is often an *unexplainable* variation in the functional dependence of some observables on others (unless the hidden variable theorists turn out to be right). Given the precise value of one variable, the laws of quantum mechanics predict only the *mean* values of other (non-commuting) observables. Similarly, the laws of tidology only predict the *mean* or expected value of the height of high waters in any particular instance.

Often the observations made *for a given value of the independent variable (D)* may be too few in number to obtain a reliable estimate of the mean of *H*. In such cases, the mean of *H* obtained for one value of *D* might be used to ‘correct’ the value at neighboring values of *D*, and this is what the *method of least squares* succeeds in doing. “*The Method of Least Squares* is a Method of Means, in which the mean is taken according to the condition, that the sum of the squares of the errors of observation shall be the least possible which the law of the facts allows. It appears, by the Doctrine of chances, that this is the *most probable* mean”.²⁶ Whewell’s subsequent remarks are very interesting. He says that “by this method, thus getting rid at once, in a great measure, of the errours of observation, we obtain data which are *more true than the individual facts themselves*.” And, “If we thus take the *whole mass of the facts*, and remove the errours of actual observation, by making the curve which expresses the supposed observation regular and smooth, we have the separate facts corrected by their general tendency. We are put in possession, as we have said, of something more true than any fact by itself is.”²⁷ Of course, Whewell’s comments are not restricted to the method of least squares, but apply to any determination of formula coefficients by statistical methods of estimation. Our next task is to understand what Whewell intended by these remarks.

²⁵ *Ibid*, p. 233.

²⁶ *Ibid*, p. 223.

²⁷ *Ibid*, p. 227.

4.2: The Nature of Scientific Induction

A major disagreement of Whewell with John Stuart Mill over the process of scientific discovery concerned Whewell's insistence that *scientific* induction is always accompanied by some sort of conceptual innovation—the adding of a *new element* to the facts by the mind of the investigator. As Whewell explains it:

Induction is familiarly spoken of as the process by which we collect a *General Proposition* from a number of *Particular Cases*: and it appears to be frequently imagined that the general proposition results from mere juxtaposition of the cases, or at most, from merely conjoining and extending them. But if we consider the process more closely ... we shall perceive that this is an inadequate account of the matter. The particular facts are not merely brought together, but there is a New Element added to the combination by the very act of thought by which they are combined. There is a Conception of mind introduced in the general proposition, which did not exist in any of the observed facts... The pearls are there, but they will not hang together until some one provides the string.²⁸

For example, suppose we perform the following high school physics experiment. Tie a light tape to a heavy object, and throw it out a sixth storey window. As it falls, the string passes through a 'ticker-timer', which punches marks on it at regular time intervals. We then plot the results on a graph, with 'time' being the independent variable on the x -axis and the distance of the each mark along the tape from the end tied to the object plotted on the y -axis. The data points on the graph are the "pearls". If the student is asked to fit a curve to the plotted data points (the "string"), he does not know what to do (he might draw straight lines between adjacent points). The teacher knows, however, that some *parabola* will provide a good - though not perfect - fit, and that a certain parameter in the specification of that parabola will measure the acceleration due to gravity (g). [This value is best obtained by plotting the distance values against the time *squared*, and then determining the slope (g) of the best fitting straight line passing through these points.] It is the teacher who provides the conceptual string.

As any mathematician knows, no finite number of distance values within any neighborhood suffices to determine the instantaneous velocity or acceleration at that point, even approximately. The data set is always finite, so the choice of a parabolic formula is not determined by the facts—it is merely the guess that has been verified by physicists, who have found it to be correlated to other independently measured quantities obtained from other experiments (which the student knows nothing about). The data in our imaginary experiment are consistent with any number of other curves—some, for example having sharp oscillations of small amplitude superposed

²⁸ *Ibid*, pp. 140-141.

on the parabola, giving wildly varying values of the instantaneous accelerations. There is nothing observed by our students that rules these out; in fact our best mechanical theories today suggest that the motion is not actually continuous at the micro-level. But, while the ‘locally’ observed facts in any particular experiment do not determine what form of law should be applied, it is impossible to extract any significant information from the data without *some* choice of formula. This is Whewell’s insight on the nature of scientific induction.

Another reason why the formula fitted is not *determined* by the data in a single induction is due to the unavoidable presence of ‘errors’ (either observational errors or ‘errors’ due to the action of other laws). The Method of Means seeks to nullify the effects of ‘error’ by accumulating many observations for each value of the independent variable, as already illustrated by the example in tidology. In such cases, it is plainly impossible to find a function that passes through every point, since a mathematical function (by definition) can only have one value assigned to one value of the independent variable. In this case, it might be argued, the accumulation of mean values uniquely determines the law governing the *means*. But given that the independent variable is continuous (ranging over an infinite number of possible values), no finite number of observations by themselves can determine the mean at every point. So, we must first select a *formula*, thereby adding a “new element” to the data.

Thus, induction proceeds from the particulars already observed to unobserved instances of the same type *via the intermediate step* of introducing a *conception* to the facts not already contained in them. The choice of conception, or formula, is not itself determined by the facts it is initially introduced to explain, but once the choice is made, the coefficients can be measured. This choice of formula can later be empirically verified by the correlation of its coefficients with other coefficients independently obtained from other inductions: what Whewell calls the *consilience of inductions*. And if the conception turns out to be *general* in this sense of ‘applying’ to many different inductions, then we have greater confidence in the form of the original law and in the proper physical significance of its coefficients. Instantaneous acceleration is a physically significant conception because it enters into the formulation of higher-level laws. The fact that the interpolation of the law to intermediate times is successful as well (to a certain approximation) is also an important means of verification, but is not the whole story.

4.3: *The Method of Residues*

As mentioned already, the ‘error’ between the *observed* value of a variable and the theoretical value assigned by a law may be due to either the imperfections of observation, *or to the operation of other laws*. It is the latter case that is important to

the discovery of *component* laws. For example, the dependency of the tides on the angle of syzygy given by the *semimenstrual inequality* has already been explained. But the dependency of the height of high waters, H , on other variables such as the Moon's parallax (P), or the Moon's declination (C), is also each discoverable by the Method of Means by averaging out the effects of the other variables over a large number of observations. Or, we can consider the *difference* between the observed values and the theoretical mean values given by the first law, and discover the law by which this *residue* varies by a *second* application of the colligation of facts. Whewell describes this method as follows:

Aphorism XLVII: The Method of Residues consists in subtracting, from the quantities given by Observation, the quantity given by any Law already discovered; and then examining the remainder, or Residue, in order to discover the leading Law which it follows. When this second Law has been discovered, the quantity given by it may be subtracted from the first Residue; thus giving a Second Residue, which may be examined in the same manner; and so on.²⁹

The residue variation above and below the semi-menstrual inequality for the tides is partly explained by component laws connecting H with the parallax of the moon (P) and the declination of the moon (C). The total observed "effect" consists of a *set* of instances of the four-tuple (C, P, D, H) ; viz. the set $\{(C, P, D, H)\}$. The method of means extracts three component "effects" from this from the data sets $\{(D, H)\}$, $\{(P, H)\}$, and $\{(C, H)\}$; each yielding the component laws $H = f(D) + R$, $H = g(P) + R'$, and $H = h(C) + R''$, respectively, where the residues R , R' and R'' , are random variables distributed 'normally' around the theoretical means given by the functions $f(D)$, $g(P)$, and $h(C)$, respectively. If, instead, we apply the method of residues starting with the law $H = f(D) + R$, we then discover that $R = g(P) + R_1$, and finally that $R_1 = h(C) + R_2$, where the 'errors' R_1 and R_2 are again randomly distributed about $g(P)$ and $h(C)$ respectively. Combining these results, we obtain the composite law $H = f(D) + g(P) + h(C) + R_2$.

In this example, the total "effect" *can* be divided into components analogously to the way in which a resultant vector north-east *can* be resolved into components due north and due east. But, in both cases, the problem is to say *why* it should be so resolved. After all, the choice of the variables D , P , and C seems somewhat arbitrary. Surely, we could replace this set with some inter-definable set of variables that would lead to a different decomposition of the "effect". Undoubtedly, yes. Or, alternatively, P and C should be seen as "aspects" of the one multi-dimensional variable, namely the vector *position* of the moon relative to the earth. The parallax

²⁹ *Ibid*, pp. 223-224.

and declination of the moon could be seen as merely *coordinates* of the moon's position. In such cases, we might expect that no particular decomposition, such as effected by a particular coordinatization of the law, has any physically significant above the rest. No such component laws are *fundamental* in any sense. So how do we tell when a composite phenomenological law should be divided into fundamental component laws?

The problem is to identify the criterion by which we recognize when a decomposition is of fundamental physical significance, and when it is merely a matter of mathematical convenience. It *is* fundamentally important to analyze the apparent motion of each planet as composed of two motions - the heliocentric motions around the sun together with the circumsolar motion of the earth. This is an example of a 'good' decomposition. An example of a 'bad' decomposition is the further division of these circumsolar motions into circular *epicyclic* components. What is it that distinguishes these two examples?

4.4: *The Consilience of Inductions*

To answer this question, we must first examine what Whewell says about the *tests of hypotheses*. These come in three distinct categories:

- (1) The Prediction of Untried Instances;
- (2) The Consilience of Inductions; and
- (3) The Convergence of a Theory towards Simplicity and Unity.

The first type of test occurs when new data are found to conform to a law already arrived at from previous observations. As Whewell puts it: "The prediction of results, even of the same kind as those which have been observed, in new cases, is a proof of real success in our inductive processes."³⁰ This notion of verification is standard in the philosophy of science, and need not detain us. We need only note that Whewell does not deny the importance of successful prediction (understood to encompass *postdiction* as well as *prediction*).

In fact, the second category of test, the consilience of inductions, is largely a special case of successful prediction. In his *Novum Organon Renovatum*, Whewell speaks of the consilience of inductions in the following terms:

We have here spoken of the prediction of facts *of the same kind* as those from which our rule was collected. But the evidence in favour of our induction is of a much higher and more forcible character when it enables us to *explain* and determine cases of a *kind different* from those which were contemplated in the formation of our hypothesis. The instances in which this has occurred, indeed, impress us with a conviction that the

³⁰ *Ibid*, p. 152.

truth of our hypothesis is certain. No accident could give rise to such an extraordinary coincidence. No false supposition could, after being adjusted to one class of phenomena, exactly represent a different class, where the agreement was unforeseen and un contemplated. That rules springing from remote and unconnected quarters should thus leap to the same point, can only arise from that being the point where truth resides.

Accordingly the cases in which inductions from classes of facts altogether different have thus *jumped together*, belong only to the best established theories which the history of science contains. And as I shall have occasion to refer to this peculiar feature of in their evidence, I will take the liberty of describing it by a particular phrase; and will term it the *Consilience of Inductions*.³¹

It is evident from this passage, and others, that Whewell understands the consilience of inductions as a feature of the evidence demonstrated by the successful application of magnitudes determined by the facts of one domain in predicting facts in a different domain. In this way, the conception invented in order to colligate the facts of one domain is shown to have a more general application. It is this “peculiar feature” in the *evidence*, which is adduced in favor of the best established theories which the history of science contains.

But how is this peculiar feature “in their evidence” shown to be present in the raw data? Here we must recall Whewell’s insistence that every colligation of facts adds a new *conception* to the facts. Now, if we consider two separate domains of inquiry, then the colligation of facts within one domain will introduce conceptions unconnected (or not thought to be connected) with the conceptions introduced in the second domain. When these different inductions, i.e. colligations of facts, “jump together”, Whewell means that the magnitudes independently measured within separate domains agree with one another, or are connected by some law-like regularity (i.e., connected by some formula). But this *higher-level* regularity will itself introduce a new formula, whose coefficients or magnitudes may be in turn be colligated with other coefficients, and so on. A proposition asserting the value of one coefficient, obtained from many lower-level facts, can be connected with other facts by a new law:

The One Fact, thus inductively obtained from several Facts, may be combined with other Facts, and colligated with them by a new act of Induction. This process may be indefinitely repeated: and these successive processes are the *Steps* of Induction, or of generalization, from the lowest to the highest.³²

³¹ *Ibid*, p. 153.

³² *Ibid*, p. 160.

The true generality of a law is not simply defined by its *immediate* domain of application, but also by the *generality of its coefficients*—as determined by the *higher-level* generalizations in which they figure. Thus, each successive step in the inductive *hierarchy* (Whewell uses the metaphor of a genealogical tree) serves to increase the generality of the lower-level laws in this sense.

Let us see how well my interpretation of Whewell's notion of consilience of inductions fits the examples he cites:

Thus it was found by Newton that the doctrine of the Attraction of the Sun varying according to the Inverse Square of this distance, which explained Kepler's *Third Law*, of the proportionality of the cubes of the distance to the squares of the periodic times of the planets, explained also his *First* and *Second Laws*, of the elliptical motion of each planet; although no connection between these laws had been visible before.³³

Newton actually *derives* instances of his inverse square law for each planet from Kepler's first and second laws. Given only Newton's definition of force, Kepler's area law implies that the force on each planet acts towards the sun, while the elliptic path (law 1) implies that the force is proportional to the inverse of the square of the distance from the sun. Kepler's third law—that the ratio of the major semi-axis squared to the period cubed is the same for each planet—then tells us that the constant of proportionality between the force and the inverse square of the distance is the same for each planet. This is interpreted as saying that each of the planetary motions has the very same “cause”, *viz.* the gravitational mass of the sun (as measured by the coefficients in each instance of Newton's inverse square law). That explanation places constraints on the magnitudes separately introduced in Kepler's first two laws in the precise sense that once the mass of the sun is specified, we can determine the area swept out by a radius drawn from the sun to each planet in a given time solely from the parameters determining the elliptic orbits (as introduced in Kepler's *first* law). Thus, Newton proved that the magnitudes introduced by Kepler's three laws are in fact interconnected, even though “no connexion of these laws had been visible before.” (I will treat Newton's discovery of gravitation more fully in the next section).

The reader should be aware that this interpretation of Whewell is not the standard one. Almost invariably, modern commentators have read Whewell as referring to the Newtonian *deduction* of Kepler's three laws from the laws of motion and gravitation. Thus, Butts claims that Whewell's theory of induction, in its full form, “expresses

³³ *Ibid*, p. 153.

what is now called the hypothetico-deductive character of well-developed sciences.”³⁴ And Blake *et al.* state more explicitly that:

The third test Whewell mentions is the capacity of a hypothesis to explain and predict cases of a different kind from those which were contemplated in the formation of the hypothesis. When this takes place, we have what he terms a ‘consilience of inductions’: that is, two laws obtained by independent inductions and concerning patently heterogeneous classes of phenomena turn out to be, both of them, *deducible from one and the same hypothesis*.³⁵

While I don’t want to deny that Kepler’s three laws can be derived from Newton’s laws, there is a simple and fatal objection to *equating* ‘deducibility’ with Whewell’s consilience of inductions. For under this equation, nothing rules out the trivial deduction of Kepler’s three laws from the conjunction of those same laws as being a genuine case of consilience. Clearly it is not. Admittedly, Whewell does say that “Deduction is a necessary part of Induction,”³⁶ but nowhere does he claim that it is *sufficient* for induction (or the *consilience* of inductions). Even if we defend the strict deducibility of lower-level laws from higher-level laws (against the many arguments in Cartwright, 1983) as being a necessary part of the story, it is clearly not the essential part. Rather, *the essential part of the consilience of inductions is the demonstration of a law-like connection between magnitudes determined by different colligations of facts—the “over-determination” of the coefficients.*

The point is that the hypothetico-deductive view of science takes the identity of coefficients in different applications of the theory as given right from the start (in the ‘hypothetico’ part of the process).³⁷ For instance, Newton presented his *Principia* in an axiomatic format in which it is taken for granted, for example, that the coefficient representing the mass of the sun has the same value for different applications of the theory. The consilience of the coefficients then appears as an observed correlation of different estimations of that value under *independent modes of measurements*.³⁸ However, Whewell’s emphasis is on the philosophy of scientific *discovery*, whereas

³⁴ *Ibid*, p.17.

³⁵ Blake et al (1960), p. 212, my emphasis added.

³⁶ Butts ed. (1968), p. 175.

³⁷ Sneed’s (1970) development of Suppes’ model-theoretic approach to the reconstruction of theories, although fundamentally axiomatic in its approach, at least recognized the existence of *constraints* between different theory applications. But neither he, nor the more recent advocates of the so-called ‘semantic’ view of theories, seem to have recognized the proper significance of this feature.

³⁸ For a very interesting discussion of “robustness as a criterion of reality” within the realm of evolution and evolutionary theory, see Wimsatt (1980). See also Skyrms (1980, 1984) for some important insights into the role of the invariance, or *resiliency*, of subjective probabilities across contexts for the proper understanding of the *objectivity* of statistical laws and physical propensities.

the top-down emphasis of hypothetico-deductivism obscures the most important component of this process—the *discovery* that two, or more, coefficients invented to explain diverse phenomena actually represent a common “cause”. This feature is given its proper emphasis in the ‘bottom-up’ approaches to the philosophy of science.

Whewell has provided roughly the following picture of science. At the phenomenological level a colligation of facts will generally introduce the means of measuring formula coefficients, which may be instances of variable quantities when viewed more globally. These ‘theoretical’ variables may themselves enter into higher-level colligations, which may then introduce further coefficients, and so on. This hierarchical procedure will eventually end when the coefficients introduced are no longer variable - as in the case of fundamental physical constants - or when the coefficients introduced are easily interpreted as conversion factors between the different scales of measurement. Examples are everywhere; *e.g.*, consider the magnitude of ‘mass’ as measured from balance phenomena or from spring phenomena.³⁹ In cases like this, the observed correlation between independent measurements is *explained* in terms of an *identity* relation: The independently determined values are approximately the same (after being converted to a common scale) *because* they are measurements of the *same* thing, *e.g.*, of ‘mass’, rather than measurements of the different properties of ‘spring-mass’ and ‘balance-mass’. Their importance of these explanations is that they give rise to a simplification and unification of our previous theoretical commitments (*e.g.* by replacing ‘spring-mass’ and ‘balance-mass’ with ‘mass’). In terms of the hypothetico-deductive viewpoint, the unity or simplicity of a theory is due to the (successful) over-determination of its coefficients.

Let me introduce one final example to illustrate the utility of the bottom-up approach. Suppose we take a heterogeneous collection of N coins (with different physical properties) and toss each one M times, first using one method, and then another M times using a different method of tossing—maybe two different tossing machines are used. Because the methods of tossing are different, we have no prior guarantee that relative frequencies for each method will measure the same property of each coin. In fact, suppose that we are not even convinced that the relative frequencies measure any property at all! Let us denote the data set obtained using the first method by E_1 , and that obtained for the second by E_2 . We “explain” each set of data as arising from independent trials of a Bernoulli process with different probabilities of landing heads for different methods of tossing. Thence, we assign theoretical probabilities (propensities) $p_1(x_k)$ to each of N coins x_k , $1 \leq k \leq N$, to

³⁹ See Forster (1986) for details.

explain the facts in E_1 , and $p_2(x_k)$ to x_k , for $1 \leq k \leq N$, to explain E_2 . In Whewell's terminology, we have two different colligations of facts, each introducing a different set of Bernoulli probabilities; $p_1(x_k)$ for E_1 and $p_2(x_k)$ for E_2 . We now estimate the values of these theoretical probabilities as the relative frequencies $\hat{p}_1(x_k)$ and $\hat{p}_2(x_k)$, for $1 \leq k \leq N$, in the data sets E_1 and E_2 respectively. These two sets of estimates for $\{\hat{p}_1(x_k): 1 \leq k \leq N\}$ and $\{\hat{p}_2(x_k): 1 \leq k \leq N\}$, will not assign the same values to each coin, but suppose that we do find a significant statistical correlation between the two sets of values obtained by the two different methods of tossing (N is large). That is, we find that when we plot all the pairs $(\hat{p}_1(x_k), \hat{p}_2(x_k))$, for $k = 1, 2, \dots, N$, on a scatter diagram, they are mostly clustered close to the line $p_1 = p_2$. We have, in other words, observed a higher-level regularity in the data. Inductions from different classes of facts have thereby "jumped together" to produce a consilience of two inductions. This consilience provides evidence that the relative frequencies measure the same physical variable (the propensities of the coins), and *ipso facto* evidence that the propensities are real.

In this example, a consilience of inductions has *led us to a simplification* in our ontology, but we should not make the mistake of conflating the two steps. In the passage immediately following the last quotation, Blake *et al.* make exactly this error:

Such a [consilient] hypothesis may equally be described as *rendering consilient* separate inductions, *or as tending to simplicity and harmony*, or again as constituting an ascent to a higher order of *generality*. This third test of Whewell's is thus in its essence identical with that constituted by the principle that Sir William Hamilton has called the "principle of parsimony" and with the well-known maxim connected with the name of William of Occam.⁴⁰

In actual fact, the consilience of inductions is Whewell's *second* test and the convergence towards simplicity is his *third* test of a hypothesis, and the two are *not* "identical in their essence," even though both features are often "exemplified by the same cases." There is the conflation we might expect from someone trying to force Whewell into a hypothetico-deductive mould. But here is Whewell's denial:

The last two sections of this chapter direct our attention to *two* circumstances, which tend to prove, in a manner which we may term irresistible, the truth of the theories which they characterize:- the *Consilience of Inductions* from different and separate classes of facts;— and the progressive *Simplification of the Theory* as it is extended to new cases. These two Characters are, in fact, hardly different; they are exemplified by the same cases....The

⁴⁰ Blake et al. (1960), p. 212.

Consilience of our Inductions give rise to a constant Convergence of our Theory towards Simplicity and Unity.⁴¹

Consilience is not the same as simplicity; the former merely gives rise to a *convergence towards* the latter. The goal—of simplicity and ontological unification—is only *achieved* once the observed consilience is deemed to be non-accidental and is subsequently *explained* in terms of an identity. Consilience is the evidence for accepting the identity.⁴² After that, the unified theory can be further tested by its deductive consequences, but it is the *discovery* of the identities and unity in nature that is the admirable part of the process.⁴³

Finally, let us return to the earlier question of how a *principled* decomposition of the total data into component “effects” might be given? I think that the answer should go something like this: True “effects” are those that arise from true “causes”. How do we recognize what the true “causes” are? We must start at the top of the inductive hierarchy and work down. The true “causes” of the phenomena are represented by those consilient coefficients that are independently inferred from different phenomena. The “cause” of the correlation between the distances of the marks on the tape and the times they were punched is the acceleration due to gravity at that place. This “cause” is recognized as the true “cause” when that same “cause” is inferred from other phenomena (*e.g.*, the behavior of pendulums).

This idea goes some way towards explicating Newton’s first rule of reasoning in philosophy stated in Book III of his *Principia*: “We are to admit no more causes of natural things than such as are both true and sufficient to explain their appearances.” Of this rule, Whewell states that:

When the explanation of two kinds of phenomena, distinct, and not apparently connected, leads us to the same cause, such a coincidence does give a reality to the cause, which it has not while it merely accounts for those appearances which suggested the supposition. This coincidence of propositions inferred from separate classes of facts, is exactly what we

⁴¹ Butts ed. (1960), p. 159.

⁴² The ontological unification is achieved by an identity statement, and the observed consilience is the evidence for the identity. As Harper (1983) points out, this feature of Whewell agrees with the view, usually attributed to Kripke and Putnam, that some necessary truths can be discovered *a posteriori*.

⁴³ As Whewell himself says; “...although the Inductive Step, the Invention of the Conception, is the most important, yet since, when once made, it occupies a familiar place in men’s minds; and since the Deductive Demonstration is of considerable length and requires intellectual effort to follow at every step, men often admire the deductive part of the proposition... far more than that part in which the philosophical merit really resides”. (Butts ed., 1968, p.174).

noticed in the *Novum Organon Renovatum*, as one of the most decisive characteristics of a true theory, under the name of *Consilience of Inductions*.

That Newton's First Rule of Philosophizing, so understood, authorizes the inferences which he himself made, is really the ground on which they are so firmly believed by philosophers. Thus when the doctrine of a gravity varying inversely as the square of the distance from the body, accounted at the same time for the relations of times and distances in the planetary orbits and for the amount of the moon's deflection from the tangent of her orbit, such a doctrine became most convincing...⁴⁴

I want to make some more careful distinctions than Whewell does in interpreting Newton's rule. As Whewell says, the coincidence of propositions inferred from separate classes of facts is what constitutes a consilience of inductions, but note that such a consilience will never be perfect. That is, the values of the coefficient variables will only correlate *approximately*, and so the consilience will not *prove* that both values are of the same physical property. The *further* inference that the causes *are the same* is licensed by Newton's imperative that we are to admit *no more causes than are sufficient* to explain their phenomena; that is, sufficient to explain the correlation.

Apart from a little laziness with these distinctions on Whewell's part, I think that this interpretation agrees with what he says. Whewell's reading of Newton's "true causes" (*verae causae*) as those "causes" that are found to be identical with the "causes" of other phenomena corresponds with our understanding of Newton's rule 1 and reinforces the previous interpretation of what Whewell meant by consilience.

In the final analysis, it may be that my explication of Whewell's philosophy of science does not stand up to a closer scrutiny of the text, but I would still argue that this is what Whewell ought to have meant. The next section contains the reply to Cartwright and Ellis that follows from the ideas of this section. It examines the explanation of gravitational phenomena in the more realistic and interesting cases in which two, or more, "causes" combine to produce one composite "effect".

V. The Causes of Gravitational Phenomena

Cartwright has stressed the importance of the non-redundancy requirement for inference to the best explanation. She argued that explanations of single observable events, such as the acceleration of a planet at a particular time in terms of component forces, do not seem to satisfy this requirement. So, we cannot infer the existence of gravitational forces. Mill's approach of dividing accelerations into components does not help - it simply transfers to problem to that of asking what determines this decomposition in a non-redundant way. Even the problem of inferring the compo-

⁴⁴ Butts (ed.) (1968), p. 330.

ment *motion* of Mars, say, around the Sun is under attack. For all we can observe from Earth is the *resultant* motion of these bodies, being the combined result of the motion of Mars around the Sun and the apparent revolution of the Sun around us. It was Copernicus who first attempted to place this decomposition of “effects” on a sound footing, and our story will start with his contribution.

5.1: *The Copernican Contribution*

Most probably, Copernicus was initially impressed by the fact that there was an epicyclic component in each of Ptolemy’s planetary models whose period of revolution was (approximately) one year—viz. the time it takes the sun to circle the ecliptic. His mathematical demonstration that the earth’s motion *could* be viewed as an component of the motion of every planet did prove that the earth’s motion around the sun *could* be seen as the common “cause” of each of these separate “effects”. In Copernicus’s own words:

We thus follow Nature, who producing nothing in vain or superfluous often prefers to endow one cause with many effects. Though these views are difficult, contrary to expectation, and certainly unusual, yet in the sequel we shall, God willing, make them abundantly clear at least to the mathematicians.⁴⁵

Copernicus’s introduction of the heliocentric conception of the solar system then provided the means of measuring the three-dimensional configuration of planets. For once we suppose that the earth moves, we interpret observations at different times as instantaneous snapshots of the planets *from different angles*, which allows us to analyze the raw data in terms of parallax calculations - analogous to the way we judge the relative distances of the trees, houses and background hills from a moving train. This was a very important innovation, for although Ptolemy’s planetary models were understood as three-dimensional, their relative scales were undetermined - the model for each planet was unconstrained by the others. Ptolemy certainly tried to fit his planetary hypotheses into a unified system, but it is generally agreed that his attempt was unsuccessful. Ptolemaic astronomy was only successful in modeling the (two-dimensional) position of the planets against the fixed stars.

Copernicus’s inter-planetary constraint—that the epicycle representing the earth’s circumsolar motion is the same (and, *ipso facto*, has the same radius) for each planetary model—automatically fixed the ordering of the planets and proved that it does not change. Moreover, the ordering—so determined—agreed with ancient determinations of planet ordering on the basis of the periodicity of planetary motions. This had some *heuristic* value, and was eventually quantified in terms of Kepler’s third law. It is therefore plausible to suppose that the next generation of astronomers

⁴⁵ *De Revolutionibus*, Book I, Chapter 10.

worked on Copernicus's theory, not because it gave better *predictions* (Whewell's first test), but because it already demonstrated some consilience and simplicity (Whewell's second and third tests) and hinted at the possibility of more. Ptolemy's impoverished conceptualization of the same data, on the other hand, had not uncovered the regularities that astronomers felt sure were there to be discovered.

Copernicus succeeded in finding some (albeit weak) evidence for the decomposition of apparent planetary motions into two circumsolar components, but he did not stop there. In the tradition of his predecessors, he further divided these motions into circular epicyclic components. But he went too far. Copernicus did not discover any consilience among the new coefficients introduced by these epicyclic components in order to justify the further decomposition. Kepler's third law—that the ratio of period squared to the radius of the deferent (the largest epicycle) cubed is the same for all planets—does *appear* to justify the decomposition of each circumsolar motion into deferent plus residue, but Kepler recognized that this evidence should be interpreted differently. He was able to prove that deferent radii are equal to the semi-major axes of his ellipses. Hence this law need not mention the deferent radii at all, and the one apparent success of epicyclic decomposition is dissolved. Moreover, the continuing need to add more and more epicycles to the Copernican system was *ad hoc* in the bad sense of the word. These versions of Copernican astronomy failed Whewell's second and third tests - the new coefficients introduced failed to lead to any consiliences of inductions, or unification. The periods of these additional epicycles were simple fractions of the period of the deferent, and the epicycle radii showed no regularity what-so-ever, either within each planet or between the planets. This is a nice example of how the *absence* of consilience will eventually lead us to *abandon* a conceptual device in spite of its obvious advantages in terms of familiarity and mathematical convenience.

5.2: *The Content of Kepler's Laws*

Kepler's approach was more parsimonious (in the ontological sense of Occam's razor) in that it did away with many of the Copernican coefficients (such as numerous epicycle radii) that did no real work in identifying higher-level regularities. Moreover, Kepler's new conceptualization did some *new* work in identifying an internal regularity in the motion of each planet in the guise of his second law (that equal areas are swept out in equal times). Here was a lawful connection between the angular velocity at one time and the angular velocity at other times even though the speed and the length of the radius varied over time. The significance of this discovery was eventually clarified by Newton.

In order to set the stage for Newton's subsequent discoveries I wish to recount Kepler's discovery of the elliptic orbit of Mars in more detail.⁴⁶ The first step in Kepler's inference was the determination of the earth's orbit around the sun. Copernicus's theory allowed for the determination of the period of the Martian orbit as 687 days, which is a little under two years. Tycho Brahe's observations from earth at E (in opposition with Mars), and at E_1 687 days later, Kepler obtained the angle SE_1M directly, and obtained ESE_1 from well known tabulations of the (angular) motion of the sun across the fixed stars. (See Fig. 2.)

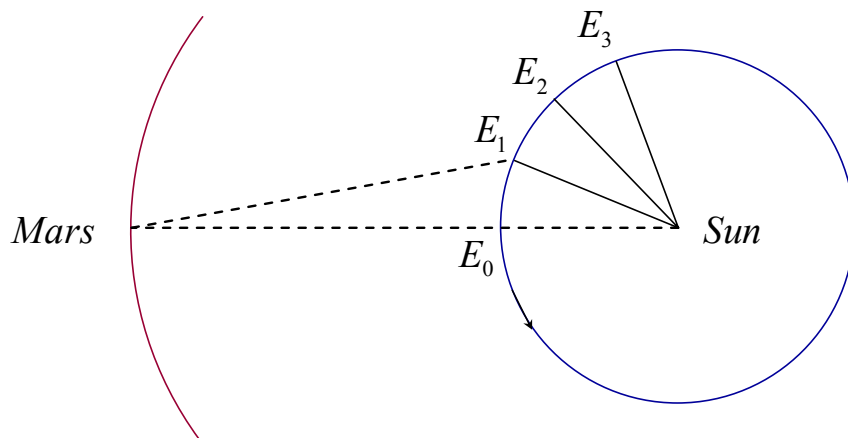


Fig. 2. The first step in Kepler's determination of Mars orbit was the calculation of the earth's orbital motion. S denotes the sun, and M refers to Mars.

Kepler might also have compared the two apparent positions of Mars relative to the fixed stars to obtain the angle SME_1 (the idea being that Mars returns to the same position M after 687 days), and this would provide a check on the other two measurements. So, the *shape* of the triangle SE_1M is thereby given, and this determines the distance SE_1 as a ratio of the (unknown) distance SM . Similar calculations for triangles SE_2M , etc, obtained when Mars had returned to the point M again, then gave the distances SE_2 , etc, as a ratio of SM also. By then fitting a smooth elliptic orbit to these discrete data points, Kepler determined the motion of the Earth around the sun.

Kepler was now able to turn his attention to measuring the distance of Mars from the sun at different stages of its orbit. Consider some other observation of Mars at M' in opposition with the earth at E_0' and 687 days later at E_1' . (See Fig. 3.)

Again, the shape of the triangle $SE_1'M'$ is determined from the knowledge of its angles, and this gives the distance SM' as a ratio of SE_1' . But the distances SE_1' are known (as a ratio of SM) from the previous colligation of the facts concerning the

⁴⁶ I follow Hanson's (1970) account, pp. 277-282.

orbit of earth. Therefore, SM' , SM'' , etc, are determined as ratios of SM . Kepler then fitted another elliptic function to obtain the orbit of Mars around the sun as a continuous function of time, which he described in his first law (elliptic path) and second law (equal areas swept out in equal times). Here Kepler is adding a *new* conception to the Copernican facts by applying his elliptical *formula* to the heliocentric motions, and its introduction is not proven by the mere predictive success of his laws anymore than the epicycle was proven by its predictive success. As already explained, the justification of Kepler's conceptual innovation lay in the intra-planetary consiliences identified by his area law and the inter-planetary consilience of his harmonic law. Although these discoveries were suggestive, Kepler himself did not succeed in *fully explaining* his results. It was left to Newton to properly identify the scope and significance of Kepler's laws.

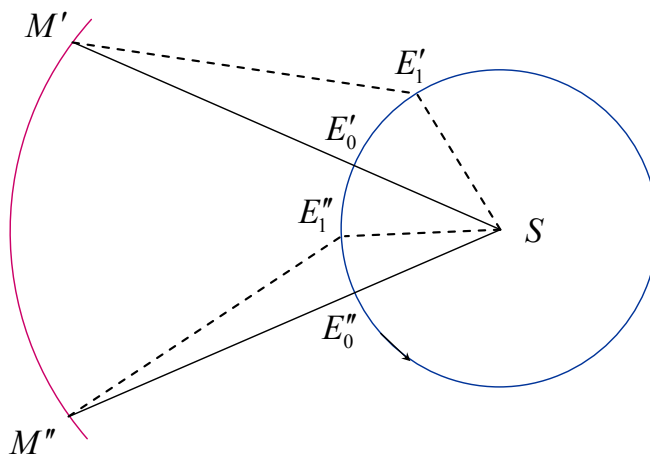


Fig. 3. The second step in Kepler's calculation of the Martian orbit.

Before passing to Newton's *Principia*, there is one *very important* lesson we can learn about the nature Kepler's laws, and of laws in general. Kepler builds up a picture of Mar's motion by sampling at most one value of its position from any single circumsolar revolution. Some commentators might say that Kepler's method of inference makes the *false* assumption that Mars repeats *exactly* the same orbit in each of its many journeys around the sun. But I suggest that this is the wrong way of looking at the situation. It is true that this assumption is false, but Kepler's laws only make this (false) assumption if they pretends to describe the *exact* motion of the planet. But it is clear from Kepler's method of inference that it is capable of "averaging out" small deviations, or residues, of Mars from its elliptic orbit provided that those residues are *random* with respect to the independent variable used (*e.g.*, the time elapsed from its perihelion position). Clearly, there were 'errors' in the data, but Kepler had no way of being sure that they were all observational errors, or

whether ‘errors’ due to the operation of other laws were also present. Throughout the history of planetary astronomy, the laws of motion of the sun, moon or planets, were resolved into a series of *partial* motions, or inequalities, each obtained by the method of residues described earlier. The laws of astronomy had *always* described the *mean* motions of the celestial bodies with respect to the independent variables so far discovered. The mere existence of a residue for each law did not prove it to be false, it only proved that the law was not complete. This was clearly regarded by Whewell as uncontroversial when he caricatured the history of the discoveries made about the motion of our moon in the following terms:

The *Equation of the Center*, for the Moon, was obtained out of the *Residue* of the Longitude, which remained when the mean anomaly was taken away. This Equation being applied and disposed of, the *Second Residue* thus obtained, gave to Ptolemy the *Evection*. The *Third Residue*, left by the Equation of the Center and the *Evection*, supplied to Tycho the *Variation* and the *Annual Equation*. And the Residue, remaining from these, has been exhausted by other Equations, of Various arguments, suggested by theory or by observation. In this case, the successive generations of astronomers have gone on, each in its turn executing some step in this Method of Residues.⁴⁷

So why assume that Kepler and his scientific descendants rejected that tradition?! Like all previous astronomical discoveries, the truth of Kepler’s first and second laws did not rest on a claim to describe the *exact* motion (even if Kepler hoped the non-observational ‘error’ would be zero). On this reading, Kepler’s method of inference made no obviously false presuppositions. His laws simply describe the *mean* elliptic motions of planets around the sun with respect to the independent variables used. And the presence of a discrepancy between the ‘exact’ values and those ‘predicted’ by the law does not imply the falsity of the law any more than does the presence of observational errors.

This fact goes most of the way towards countering Cartwright’s charge that the fundamental laws of physics fail to describe how bodies behave—what Cartwright calls the *facticity requirement*. Kepler’s laws, for instance, do correctly predict *aspects* of the behavior of planets that manifested by the statistical *regularities* in a large number of observations. Cartwright’s view is that fundamental laws, if true, must be qualified by a *ceteris paribus* clause, which for her means that each instance of the law must be *complete* in the sense of leaving no unexplained residue. In the case of planetary motion, a *believable* astronomical law must describe the *exact* positions of the planets at any time. But this requirement of completeness is too strict, and so Cartwright fails to establish that we must trade-off generality for facticity. Kepler’s laws achieve facticity without qualification. Moreover, this

⁴⁷ Butts ed. (1968), p. 235.

interpretation of laws, I claim, leads to a better understanding of Newton's arguments for universal gravitation.

5.3: *The Newtonian Argument for Gravitation*

Let us examine a part of Newton's argument concerning the lunar motions in order to make sense of the idea that Newton is really treating Kepler's laws as *true descriptions of components* of the motions. My reconstruction is as follows. In Prop. IV, Book III, Newton argues that if we take the mean distance of the moon from the earth to be 60 (mean) earth radii, then from the known (mean) period of the lunar orbit we can calculate the (mean) acceleration of the moon towards the earth. This calculation reveals that constants of proportionality between the acceleration of the moon and the inverse square of its mean distance, and that between the acceleration of terrestrial bodies and the inverse square of one earth radius, *are the same*. Therefore, concludes Newton in the following Scholium, "since both these forces, that is, the gravity of heavy bodies, and the centripetal forces of the moons, are directed to the centre of the earth, and are similar and equal between themselves, they will (by Rule 1 and 2) have one and the same cause."⁴⁸

This calculation is only approximate: For one thing, it uses an undervalued estimate of the mean lunar distance of 60 earth radii, when a better estimate would be $60^{1/2}$ earth radii. So, Newton must account for the increase of the corrected estimate of the moon's acceleration. In doing this,⁴⁹ he refers to the fact that the calculated acceleration is actually the *relative* acceleration of the moon from the earth. But what should be compared is the *absolute* acceleration of the moon towards the earth, and this should be taken from their common center of gravity. Here is what happens. The constant of proportionality between the *relative* acceleration of the moon from the earth and the inverse of the distance squared is equal to the mass of the earth *plus* the mass of the moon; whereas the similar constant of proportionality describing the relative motion of terrestrial bodies is the sum of the earth's mass plus the mass of that body (a total which differs insensibly from the mass of the earth by itself). So the two coefficients are not equal. But if we subtract the mass of the moon from the first constant then the two are equal. This 'subtraction' is effected by taking the accelerations to be relative to the center of mass of the earth-moon system.

This brings us back to the anti-realist polemics of Cartwright and Ellis, and shows how their arguments are related. For we can only maintain that the "cause" of moon's motion is exactly the earth's gravity if we divide the observed relative acceleration into *components* defined relative to the centre of mass of the system.

⁴⁸ Newton, as translated in Cajori (1960), p. 409.

⁴⁹ *Ibid*, p. 410.

And this move of decomposing the “effect” into parts seems purely *conventional* because when we try to find a justification of it we tend to go round in circles. If we *define* the celestial gravity of the earth to be the constant of proportionality in the inverse square law governing the moon’s acceleration *relative to the center of mass frame*, we are just defining it to be equal to the known terrestrial value. This begs the issue at hand (*viz.* the identity of celestial and terrestrial gravity).

So, we must deny that the earth’s mass is *defined* solely in terms of the motion of the moon. We need independent confirmation of its value from other celestial phenomena. The earth’s gravitational mass should be defined as the *common* “cause” of a number of other independent phenomena, in the same way that we have independent confirmation of the terrestrial value from different terrestrial phenomena such as the motions of pendulums and projectiles.

We will not be concerned with the evidence *Newton* had for the identification of celestial and terrestrial gravity, but with the general nature of the argument for universal gravitation available from the wealth of evidence we have today, e.g. from accurate information about the moon’s motion or from data collected from artificial satellites circling the earth, and (occasionally) the moon. We will examine the argument for gravitation in the simple case of the earth-moon-sun system under the assumption that our data is sufficiently accurate to yield fairly precise estimates for all the parameters involved. This simple example will adequately illustrate the full argument.

Consider the three-body system consisting of the sun, earth, and the moon, denoted by x_0 , x_1 , and x_2 , respectively. Our data set consists of many measurements of the relative displacement vectors $\mathbf{r}_{10}(t)$, $\mathbf{r}_{12}(t)$, and the vector accelerations $\ddot{\mathbf{r}}_{10}(t)$, and $\ddot{\mathbf{r}}_{12}(t)$, at time t . $\mathbf{r}_{10}(t)$ is the vector from the earth to the sun, and $\ddot{\mathbf{r}}_{01}(t)$ is the acceleration of the sun relative to the earth (taking the fixed stars to be non-rotating). (See Fig. 4.)

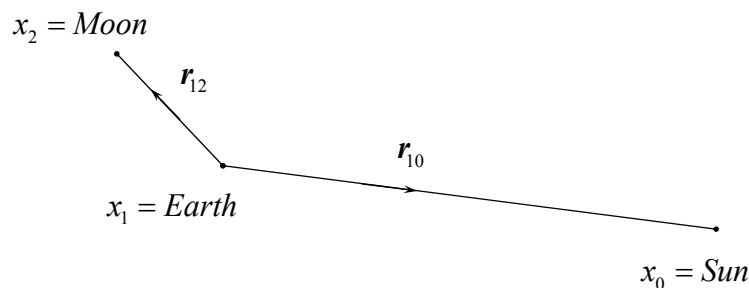


Fig. 4. The three-body problem in Newton’s theory of gravitation.

Notice that $\mathbf{r}_{10} = -\mathbf{r}_{01}$, and $\ddot{\mathbf{r}}_{10}(t) = -\ddot{\mathbf{r}}_{01}(t)$, etc. As suggested by this notation, the acceleration $\ddot{\mathbf{r}}_{01}(t)$ is *in principle* obtainable from the (twice differentiable) vector function $\mathbf{r}_{01}(t)$ by *twice* differentiating with respect to time (as indicated by the *two* dots). But this fact of modern calculus is not relevant to Newton's reasoning, and so I will simplify this notation in order to emphasize the essential features of the argument. Let us therefore denote the *dependent* variables of our colligations, the relative accelerations, by the vector variables $\mathbf{Y}_{10}(t)$ and $\mathbf{Y}_{12}(t)$ such that

$$\mathbf{Y}_{10}(t) = \ddot{\mathbf{r}}_{10}(t) \text{ and } \mathbf{Y}_{12}(t) = \ddot{\mathbf{r}}_{12}(t). \quad (1)$$

The facts at hand are data concerning the values of these relative accelerations at different times. The first step in the colligation of facts, we may recall, is the selection of an independent variable, while the second step consists in the selection of the *form* of the law that will connect them. These two steps are not unrelated, so to simplify the second step we shall take the independent variables for \mathbf{Y}_{10} and \mathbf{Y}_{12} to be chosen as the vector variables \mathbf{X}_{10} , \mathbf{X}_{12} , and \mathbf{X}_{20} , respectively, where these are defined in a rather complicated way as:

$$\mathbf{X}_{10} \equiv \frac{\mathbf{r}_{10}}{|\mathbf{r}_{10}|^3}, \mathbf{X}_{12} \equiv \frac{\mathbf{r}_{12}}{|\mathbf{r}_{12}|^3}, \text{ and } \mathbf{X}_{20} \equiv \frac{\mathbf{r}_{20}}{|\mathbf{r}_{20}|^3}, \quad (2)$$

where $|\mathbf{r}_{10}|$ denotes the length of the vector \mathbf{r}_{10} , and so on. Thus, $\mathbf{X}_{10}(t)$ is a vector directed from the earth to the sun whose magnitude is inversely proportional to the square of the earth-sun distance. Again note that $\mathbf{X}_{10} = -\mathbf{X}_{01}$, etc. As data, suppose we are given two (finite) sets of quadruples of vectors $\{(\mathbf{Y}_{10}(t), \mathbf{X}_{10}(t), \mathbf{X}_{12}(t), \mathbf{X}_{20}(t))\}$ and $\{\mathbf{Y}_{12}(t), \mathbf{X}_{10}(t), \mathbf{X}_{12}(t), \mathbf{X}_{20}(t)\}$ for many different values of t .

Suppose that we first investigate the dependence of \mathbf{Y}_{10} on \mathbf{X}_{10} and of \mathbf{Y}_{12} on \mathbf{X}_{12} . What is the form of this dependence? This is implicitly the problem that Kepler solved, and Newton proves in his *Principia* that Kepler's three laws for the earth's motion around the sun or the moon's motion around the earth can each be expressed as instances of his inverse square law, respectively, as follows;

$$\mathbf{Y}_{10} = -\mu_{10}\mathbf{X}_{10}, \quad (3)$$

$$\mathbf{Y}_{12} = -\mu_{12}\mathbf{X}_{12}, \quad (4)$$

where the coefficients μ_{10} and μ_{12} are to be estimated from the data. Each equation represents the *Keplerian component* of the motion as described by Kepler's three laws. The solutions of this form of equation are elliptic orbits satisfying the area law and the harmonic law. And conversely, Kepler's area law implies that the accelera-

tion of each body is towards the other, and the ellipticity of the paths proves that the acceleration is inversely proportional to square of the distance between the two bodies. Kepler's laws entail instances of the inverse square law, and *Newton's theory of gravitation entails that Kepler's laws are true*. This conclusion is diametrically opposed to the received view of philosophers and historians of science. Most modern commentators have tended to agree with Pierre Duhem's famous statement on the matter:

The principle of universal gravity, very far from being derivable by generalization and induction from the observational laws of Kepler, *formally contradict these laws. If Newton's theory is correct, Kepler's laws are necessarily false.*⁵⁰

But Newton himself was quite explicit in denying this opinion. He says that Kepler's third law - as applied to the planets - holds *exactly* because all the observed deviations from that law can be ascribed to other causes (other independent variables) which he can afterwards explain (as arising from the effects of *other* gravitating masses):

*Kepler and Boulliau have, with great care, determined the distances of the planets from the sun; and hence it is that their tables agree best with the heavens. And in all the planets, in Jupiter and in Mars, in Saturn and the earth, as well as in Venus and Mercury, the cubes of their distances are as the squares of their periodic times; and therefore (by Cor. VI, Prop. IV, Book I) the centripetal circumsolar force throughout all planetary regions decreases as the inverse square of the distances from the sun. In examining this proposition, we are to use the mean distances, or the transverse semiaxes of the orbits (by Prop. XV, Book I), and to neglect those little fractions, which, in defining the orbits, may have arisen from the insensible errors of observation, or may be ascribed to other causes which we shall afterwards explain. And thus we shall always find the said proportion to hold exactly;...*⁵¹

It is reasonable for Newton to hold the same view for the motion of the moon, or for modern-day artificial satellites circling the earth. On this view Kepler's laws do not provide a *complete* description of the phenomena, but they do give a description in full agreement with Newton's theory. They describe only *part* of the motions, and do not, therefore, explain *all* the facts of celestial mechanics.

However, Kepler's laws play an *essential* part in *all* of Newton's reasoning, since they are used to define the *residues* that lead inductively to a more robust determination of the gravitational masses. Newton's success in explaining these residues argues for the *truth*—not falsity—of Kepler's laws. Let us now turn to the *explana-*

⁵⁰ Duhem (1970), p. 193, my emphasis.

⁵¹ Cajori (1960), p. 559, my emphasis.

tion of these residues. The fact that Kepler's laws do not describe the exact motion is especially noticeable for the moon's motion. As the previous quote from Whewell indicates, the problem of finding a *complete* set of laws governing the moon's motion had long been the object of some despair among astronomers before Newton. The traditional approach was to discover new independent variables in an effort to reduce the size of the residue left unexplained. Newton's approach is basically the same except that he has a better idea as to what the *variables* should be. The example is very important as an illustration of how a very complex "effect" is *successfully explained by the composition of several "causes"*—viz. the combined gravitational attraction of the earth and the sun.

First, Newton was armed with the idea of searching for the laws governing the residue from Kepler's laws (equations (3) and (4)). Kepler's third law then says that the coefficients in instances of the inverse square law are the same for the same attracting body. Each planetary motion, therefore, provides a different measurement of the sun's gravitational mass, and this gravitational mass is then the common "cause" of these Keplerian "effects". And a similar story is provided for the motions of satellites of Jupiter and Saturn.

So, if the differences, or *residues*, of the exact motions from their Keplerian components could be *explained* as well, then this would provide a sort of confirmation of Kepler's laws in virtue of the essential role they play in defining these residues. In fact, Newton not only succeeds in explaining the residue motions, *but he does so in terms of the same "causes" (the gravitational masses) already introduced to explain the Keplerian motions.* That is, he strengthens his initial explanation of Kepler's laws by providing more independent ways of measuring the magnitude of the "causes", thereby strengthening the evidence for the reality of the "causes" already discovered. That the explanation of this consilience *should* be that the properties measured are identical is stated by *Newton's second rule of reasoning in philosophy*: "Therefore to the same natural effects we must, as far as possible, assign the same causes."

Newton's idea is to account for the residues in equations (3) and (4) by the independent variables defined in equation (2). This move leads to the following phenomenological equations;

$$Y_{10} = -\mu_{10}X_{10} + m_2X_{21} + (-m_2')X_{20}, \quad (5)$$

$$Y_{12} = -\mu_{12}X_{12} + m_0X_{01} + (-m_0')X_{02}, \quad (6)$$

where the values of the new coefficients m_2 , m_2' , m_0 , and m_0' are also estimated from the data. Because these two equations, (5) and (6), are *vector* equations, each of them might be further divided into three scalar equations giving six scalar equations

altogether. So, in our ‘bottom-up’ approach to the argument, we should label each of the three appearances of each coefficient, μ_{10} , m_2 , etc., *differently* so as not to assume *a priori* that all three instances are appearances of the *same* coefficient. But once we estimate the values of these coefficients by fitting each scalar equation to the data by the method of least squares, we find that all three instances yield values that are approximately the same. So, we explain that consilience by assuming that they are identical, to arrive at equations (5), and (6) above. We will label the estimated values of the coefficients as $\hat{\mu}_{10}$, \hat{m}_2 , \hat{m}'_2 , $\hat{\mu}_{12}$, \hat{m}_0 and \hat{m}'_0 . Then we notice a still further consilience amongst our coefficients, which justifies the following identities:

$$m_2 = m'_2, \quad m_0 = m'_0, \quad \mu_{10} - m_0 = \mu_{12} - m_2 = m_1. \quad (7)$$

We now have a robust estimate of the earth’s gravitational mass *alone*, denoted by m_1 , from the celestial phenomena alone, and this value turns out to be the same *as* the terrestrial values. Facts such as these are what substantiates Newton’s conclusion that the celestial gravity is the same as terrestrial gravity. This fits into the general inference pattern in which we are able to *explain* the consilience of coefficients by supposing that each value measures a common “cause” of the phenomena in question.

With these identities taken as given, we can re-write our descriptive equations as:

$$Y_{10} = -(m_0 + m_1) X_{10} + m_2 [X_{21} - X_{20}] \quad (8)$$

$$Y_{12} = -(m_2 + m_1) X_{12} + m_0 [X_{01} - X_{02}] \quad (9)$$

As we increase the number of bodies included in the system, we get more independent occurrences of each mass coefficient and we thereby obtain stronger consiliences (given that Newton’s theory is approximately true), and greater justification for the identities which are explicitly stated in (7) and implicit in equations (8) and (9).⁵²

The gravitational masses are interpreted as the “causes” of the complex phenomena described in equations (8) and (9). Since each equation has terms with coefficients m_0 , m_1 , and m_2 , each composite “effect” has three “causes”; the gravity of the sun, earth and moon respectively. This is now a typical case of *composite* “causa-

⁵² To generalize to present discussion to deal with the mutual gravitational interaction of n particles, we should use the equations:

$$Y_{ij} = -(m_i + m_j) X_{ij} + \sum_{k \neq i, j}^n m_k (X_{ki} - X_{kj}),$$

where the measurement scale of ‘mass’ coefficients has (again) been ‘chosen’ so that the universal gravitational constant, G , has unit value.

tion”, where the magnitude of each “cause” is measured by its partial “effect”, much along the lines suggested by Mill for singular causation. The crucial difference is that the “effects” are no longer *events* occurring at a particular time and location. It is this difference that *allows* us to defend the view that “effects” have parts. The empirical determination of the decomposition is based on the consilience exhibited by the coefficients of the phenomenological laws. Since the consilience is a feature of the *evidence*, viewed globally, the decomposition is determined by the *empirical* facts in this general sense, and we satisfy thereby the demands of a moderate empiricism.

Once given, the values of the masses allow us to define the center of mass frame of reference, relative to which equations (8), and (9) can be rewritten in a more perspicuous form. First of all, notice that the *relative* accelerations Y_{10} and Y_{12} are defined with respect to an *earth-centered* frame of reference, which is non-rotating with respect to the fixed stars. There is nothing problematic about this in itself. But the next step is to introduce ‘absolute’ accelerations Y_0 , Y_1 , and Y_2 of the sun, earth, and moon respectively, defined relative to a *center of mass frame* such that the previous relative accelerations Y_{10} and Y_{12} are definable in terms of them as:

$$Y_{10} = Y_0 - Y_1 \text{ and } Y_{12} = Y_2 - Y_1. \quad (10)$$

But we must realize that the equalities here are not sufficient to uniquely determine the vectors Y_0 , Y_1 , and Y_2 from Y_{10} and Y_{12} . So, the question is: How are Y_0 , Y_1 and Y_2 determined by the *observed* facts?

Or, more usefully, how do we determine the acceleration of the origin of the frame of reference from which Y_0 , Y_1 and Y_2 are measured? Let us denote this acceleration by Y^* . Once we have determined this, we can define Y_0 , Y_1 and Y_2 by:

$$Y_0 = Y_{10} - Y^*, \quad Y_1 = -Y^*, \quad Y_2 = Y_{12} - Y^*. \quad (11)$$

The problem is to say how Y^* is determined by the facts. The solution is that Y^* is determined by the robust ‘mass’ coefficients in equations (8) and (9). In particular, we define Y^* by:

$$(m_0 + m_1 + m_2) \cdot Y^* = m_0 Y_{10} + m_1 Y_{11} + m_2 Y_{12}. \quad (12)$$

It is now easy to check that

$$Y^* = m_0 X_{01} + m_2 X_{21}. \quad (13)$$

When we substitute this expression into the equations in (11), using (8) and (9), we arrive at:

$$\mathbf{Y}_0 = m_1 \mathbf{X}_{01} + m_2 \mathbf{X}_{02} \quad (14)$$

$$\mathbf{Y}_1 = m_0 \mathbf{X}_{10} + m_2 \mathbf{X}_{12} \quad (15)$$

$$\mathbf{Y}_2 = m_1 \mathbf{X}_{21} + m_0 \mathbf{X}_{20} \quad (16)$$

The essential feature of these equations is that each independent variable has a *single* coefficient. This is the key to understanding what is achieved by the transformation. Of course, our primitive empirical formulae, (5) and (6), also have this property - each independent variable has at most one coefficient associated with it in any given equation. But once we discover the identities expressed in (7), this property is lost; as is seen by the first terms on the right hand sides of equations (8) and (9), which have compounded coefficients of the form (m_0+m_1) and (m_2+m_1) . The transformation, defined by (11), serves to restore the situation of at most one coefficient to one independent variable in any single equation. The center of mass transformation is a necessary step towards the pairing of “causes” (*verae causae*) with their “effects”. But the above equations still describe complex phenomena composed of two component “effects”, which we aim to represent separately.

This is achieved in the same way that the two partial dependencies in $Y = a.X_1 + b.X_2$ can be rewritten as $Y_1 = a.X_1$ and $Y_2 = b.X_2$ if we allow that $Y = Y_1 + Y_2$. So, Newton writes the resultant accelerative forces \mathbf{Y}_0 , \mathbf{Y}_1 , and \mathbf{Y}_2 , according to the equations:

$$\mathbf{Y}_0 = \mathbf{A}_{01} + \mathbf{A}_{02}, \quad \mathbf{Y}_1 = \mathbf{A}_{10} + \mathbf{A}_{12}, \quad \text{and} \quad \mathbf{Y}_2 = \mathbf{A}_{20} + \mathbf{A}_{21}. \quad (17)$$

We then obtain Newton’s equations in their more fundamental form as:

$$\mathbf{A}_{01} = m_1 \mathbf{X}_{01} = m_1 \frac{\mathbf{r}_{01}}{|\mathbf{r}_{01}|^3}, \quad \mathbf{A}_{10} = m_0 \mathbf{X}_{10} = m_0 \frac{\mathbf{r}_{10}}{|\mathbf{r}_{10}|^3}, \quad \text{etc.} \quad (18)$$

For instance, \mathbf{A}_{10} is the *component accelerative force* of the sun acting on matter at the earth’s position, and this vector quantity is directed from the earth towards the sun with magnitude inversely proportional to the square of the earth’s distance from the sun, where the constant of proportionality represents the sun’s gravitational mass. The equations in (18) have still not quite captured the full symmetry of the phenomena since pairs such as \mathbf{A}_{01} and \mathbf{A}_{10} are not independent of one another, because $m_0 \mathbf{A}_{01} = -m_1 \mathbf{A}_{10}$, etc. This is *Newton’s third law* of mechanics, which allows us to define

$$\mathbf{F}_{01} = G m_0 m_1 \frac{\mathbf{r}_{01}}{|\mathbf{r}_{01}|^3} = -\mathbf{F}_{10}, \text{ etc.}, \quad (19)$$

where I have changed the *scale* of measurement of the gravitational mass coefficients m_0 , m_1 , and m_2 , to introduce the universal constant of gravitation G so that (19) combines properly with the modern form of Newton's second law, $\mathbf{F}_{01} = m_0 \mathbf{A}_{01}$, etc., where m is the *inertial* mass of the body in question. This puts the Newton's special force law in the form familiar to all of us. (The reason for doing this is that it then combines properly with non-gravitational force laws, such as Hooke's spring law).

To complete this story, we should emphasize that, once discovered to hold within a certain domain, the law is automatically extended universally to apply in all future instances and *for all bodies whatsoever*. There is an inductive generalization as traditionally construed, but it is a generalization *constrained* by the results of experimentation. Newton's third rule of reasoning is as follows: "The qualities of bodies, which admit neither intensification nor remission of degrees, and which are found to belong to all bodies within the reach of our experiments, are to be esteemed the universal qualities of all bodies whatsoever". In particular, the *power* of gravity measured by the gravitational mass of a body is found to be invariant over all independent determinations for applications and for all times so far examined, so this property should be extended to other bodies and to other times. So, we should extend the previous measured value for the mass of the cup to apply to future times, and we assume that all as yet undiscovered bodies obey the same laws (as was instrumental in the discovery of Neptune). However, the preconditions placed on this inductive inference—conditions demanding the consilience of the coefficients so far measured—protect this form of inference from the usual counterexamples to enumerative induction (such as the chicken that is fed every day 99 times in a row and then killed on the 100th day). As Whewell says, it is the *intermediate* step of adding a new conception—in this case the 'mass' concept—that enables us to *verify* the truth of our law in terms of the consilience of inductions. It is this form of verification (Whewell's second and third tests of hypotheses) that the naive cases of enumerative induction lack.

5.4: *The Reply to Cartwright and Ellis*

The 'bottom-up' perspective on the Newton's theory of gravitation points to the discovery of the separate "causes" of complex phenomena by the consilience of inductions. The "effects" represented by the various terms in the primitive phenomenological equations are identified as having separate or common "causes" by discovering the pattern of *consiliences* among their coefficients. Once the "causes" have been properly individuated in this way, the introduction of forces and a center

of mass frame of reference serve to represent the “effects” of each of these “causes” separately. We arrive, thereby, at the axiomatic form of Newton’s theory. This form of representation properly reflects the ontological simplicity discovered in the consilience of the mass coefficients.

The ‘naturalness’ of this frame of reference for the measurement of accelerations is based the *empirical* fact of consilience, and is not simply a matter of convention. The further introduction of component forces in Newton’s fundamental law of gravitation is also indirectly motivated by the same empirical evidence. Therefore, forces are not conventional, and there is no basis to the allegation that there is no evidence for their existence. The empiricist arguments of Cartwright and Ellis are faulty in that they fail to recognize that the consilience of inductions is a form of empirical evidence. They overlook the global features of the facts.

What about the alleged tautological status of Newton’s law of inertia?⁵³ The law is most commonly understood as asserting that: *Relative to an inertial frame of reference*, a body will not accelerate unless acted on by a force. What determines whether a body is moving along an *inertial* trajectory, asks Ellis, if not just the fact that there is no net force acting on it? ‘Nothing’, answers Ellis. But this reply overlooks that most primitive phenomenological laws in Newtonian mechanics are stated in terms of *relative* accelerations, and there is no problem in specifying their reference frames. The success in identifying the “causes” of the variations of this relative acceleration with various independent variables leads us to correctly identify *components* of this acceleration. This identification then determines a new ‘theoretical’ frame of reference, and the new frame determines the components. The correct decomposition is determined by the consilience of the coefficients in the phenomenological laws. There is no need to define an *inertial* frame of reference in Newtonian mechanics. Of course, it may well be that the centre of mass of the solar system is linearly accelerating with respect to the fixed stars. But that component of acceleration would be common to all bodies in our solar system, and thus does not have any effect on the *relative* accelerations. This indifference to the specification of inertial frames is part of the strength of Newtonian mechanics, and not its failure! So, why does Newton state his law of inertia at all? The existence of an inertial frame of reference is tied up with two claims; (1) that only dependent variables of the phenomenological equations are accelerations (rather than velocities), and (2) the assertion that Newton’s theory is *complete*. So, the law of inertia implies that any *acceleration* of the solar system with respect to the fixed stars (if discovered) must fall under the umbrella of Newton’s inverse square law; in which case the centre of

⁵³ It is instructive to note that Sober (1984) has reached similar conclusions about the alleged tautological status of the law of natural selection.

mass of the enlarged system will serve as an even better approximation to a truly inertial frame of reference. And it implies that any residue acceleration of the objects (relative to that frame of reference) beyond what is explained by gravitation must be explained by other forces (e.g. the force of the table acting on the cup). The empirical content of Newton's law of inertia, and his second law ("The change of motion is proportional to the motive force impressed"), is that all residue *accelerations* will be explained by special force laws. Certainly, this assertion has important *empirical* implications of great heuristic value, but its truth is not crucial to the wide success of Newtonian physics (as history has shown!).

Let us state what has been achieved in terms of the example of the cup sitting on the table. How are we justified in saying that there is a gravitational force of 2 newtons acting on it? The answer is that the *special force law* in (19) tells us what gravitational forces are acting at any time. That is how we arrive at the figure of 2 newtons rather than 20 newtons. The question of *justification* therefore reverts back to the empirical foundation of these special laws, which has been the subject of this section. The proper empirical foundation of this law lies in the consilience of the mass coefficients, which Newton demonstrated by correctly deriving the phenomena from his definitions, axioms, and laws.

However, we have only dealt with the *accelerative* force, whereas we also need to determine the mass of the cup ($F = m.A$). What about Ellis's possible charge that the measurement of the cup's mass must involve an element of conventionality? For if we measure the mass of the cup by its effect on a spring, say, then how is the effect non-conventionally determined? Again, we must look at the phenomena more globally. Suppose the cup oscillates up and down on the end of the spring. We treat the gravitational acceleration as given by the special law of gravitation and the mass of the earth (as measured by previous experiments), and then seek to explain the *residue* acceleration, of the exact acceleration, Y , minus the gravitational acceleration A_g . Suppose, we examine the *variation* of this quantity with respect to the independent variable X —the extension of the spring from its un-extended position—and fit the following equation to the data:

$$A_H \equiv Y - A_g = -a X + b . \quad (20)$$

When we measure the coefficient a in cases where we use different objects (on the same spring), and on different springs, we find that a enters into more general connections with properties of spring, stiffness k , and the mass properties of the bodies, m . This higher-level consilience is of the form:

$$a = k/m . \quad (21)$$

We finally arrive at Hooke's law for linear springs as:

$$A_H = - (k/m) X + b . \quad (22)$$

All that matters for our purposes is that we can measure the mass of the cup m , as a ratio of a standard mass m_0 , by the ratio a_0/a , where both of these a -coefficients are obtained from the same spring. The question is: Does this value depend on whether we take X to be measured from the 'natural' un-extended position of the spring, or not? *The answer is that it does not.* If we instead use the independent variable X' related to X by $X' = X - C$, for some arbitrary non-zero constant C , we still obtain the same numerical measurement of the mass. The upshot is that the concept of mass is introduced to explain a *global* phenomenon - the *variation* of A_H with respect to X (or with respect to X') and not just to explain the isolated instances of the X (or X') variable.

The mechanical properties of objects, such as 'mass' or 'stiffness', are introduced as the "causes" of the observed covariation of two, or more, variables. Once these "causes" are properly introduced, the explanatory role of 'forces' is clarified. 'Forces' are not introduced to explain isolated events; they are introduced as the dependent variables in the phenomenological laws that describe the "effects". Although the forces are not "causes" themselves, they do play an essential role in the explanatory success of the theory, and their magnitudes are uniquely determined by the service they perform.

VI. Concluding Remarks

Most modern anti-realist philosophers of science, who are also motivated to find some middle ground between austere positivism and cavalier realism, tend to drive some sort of epistemological wedge between different *levels* of theorizing. For example, van Fraassen (1980) condones the belief in propositions about observables, but not about unobservables.⁵⁴ Cartwright (1983) draws the line differently; between phenomenological laws and fundamental laws. But the use of simplicity and the evidence of consilience as the proper 'mark of reality' does not draw any such horizontal line. Consilience and unification occur at every level of the epistemic hierarchy. On the lowest level of fact-finding in science, assumptions that different observations made in different situations are of the *same* thing or of the *same* type are

⁵⁴ For criticisms of van Fraassen on this point, see Churchland (1985), Musgrave (1985) and Sober (1985). See also Smart (1985) for a reply to van Fraassen which appeals to simplicity. Hooker (1985) tackles van Fraassen on the thesis that pragmatic virtues are non-cognitive. The arguments of this essay, I think, tend to support Hooker's conclusions.

so common that they tend to go unnoticed. Why do we believe, for example, that the morning star and the evening star are the same object? Or, why do we *justly* believe that different channels of sensory information (e.g. simultaneous snapshots from different eyes) sometimes give us information about the *same* world? The justification of this sort of ‘pre-theoretical’ inference is also likely to appeal to the principles of consilience and simplicity. So, if the anti-realists find the higher-level inferences of science automatically suspect, then on pain of contradiction they should find themselves doubting the soundness of our more primitive judgments as well. The best way to avoid this road to total skepticism is not to step onto the slippery slope in the first place.

A better solution to the problem described by Friedman—the problem of finding middle ground between positivism and cavalier realism—is in terms of inference to the best common “cause” explanation. The inference must be from an explanation of observed *regularities*, and their inferred “causes” must be independently verified as the “causes” of other phenomena. In fact, it is exactly the proven *generality* of the “causes”—their consilience—that turns a ‘mere’ description into a genuine *explanation* of the phenomena.

This principle of common “cause” insists on the *empirical* evidence of consilience as a precondition for the simplification of our theoretical constructions. If there is no *empirical evidence* that two “causes” are consilient, then there is obviously no justification for assuming that they are one and the same “cause”. So, in areas of inquiry where nature is not simple, the simplicity postulate will not lead us to believe otherwise. Contrary to what is often alleged, the moderate application of Occam’s razor does not presuppose that nature is simple. There is absolutely no *a priori* limit placed on the complexity of nature by this realist methodology.

The increasing generality, robustness, or cross-situational invariance, of our theoretical concepts and the consequent *unification* of our theories should *not* be regarded as a merely *pragmatic* or aesthetic kind of utility. On the present view, simplicity is *fundamentally* a *realist* virtue because it is the mark of our *success* in gaining information *about the (external) world*. The coefficients of our phenomenological equations *refer* to the “causes” of the phenomenon, on this view, and the more fundamental laws governing these coefficients *describe* these “causes”. The judgment that an observed statistical correlation is non-accidental (or not) is a judgment about whether the correlation really has a “cause” (or not). So, the ‘stiffness’ coefficient of a spring, *k*, *refers* to a property of the spring, and this “cause” is *described* in terms of its functional dependency on factors such as temperature, spring shape, and the more general properties of the material. These higher-level laws are descriptions *of* an external “cause”, and their unifying power

serves as the mark of *successful reference*. Our best theories at least succeed in referring to the world (rather than idealizations of it) even if our *descriptions* are never beyond improvement.

The fact that our epistemological access to the deeper aspects of reality are ultimately built upon our experiences does not mean that the degree of confirmation, or reliability, of these deeper truths cannot exceed that of the observations on which they are based. If we view the evidence too locally, it appears that a theory is always *seriously* underdetermined by its evidence. But a more *global* examination of the data may uncover an overdetermination of the coefficients due to their increased generality, which may then compensate for the underdetermination of the conception (formula) used in the lower-level colligation of the facts.⁵⁵ The greater the depth and generality of our theoretical propositions, the wider the domain of data relevant to their confirmation becomes. Deep theoretical facts can have high confirmation because of their generality, not in spite of it.

We can ‘correct’ the errors of our observations in spite of the paradoxical fact that our knowledge of *how* to make the ‘correction’ ultimately comes from the same data. For example, it is well known that the effects of stellar aberration will displace the apparent positions of the stars and planets slightly from their ‘true’ positions. But once it is known how the effect arises from the rotation of the earth, the ‘error’ can be ‘corrected’. Yet, our understanding of the phenomenon itself is ultimately based on observations that have been tainted by those same ‘errors’. There are two reasons why there is no self-contradiction involved here. First, it is the *global* features of the data—obtained by statistical ‘averaging’—that (eventually) lead to the ‘correction’ of local data points, and these are *different aspects* of the evidence. The second (not unrelated) reason is because the ‘correction’ is not really a correction at all; but rather a *judgment* about which part, or component, of the observation is *caused* by the elements of reality under investigation (*e.g.*, the planets), and which part is due to irrelevant, extraneous, or ‘chance’ factors (the motion of the earth). Of course, as a report of how the stars *appear*, the observation may be 100% accurate. This point is especially obvious in the coin tossing example. The outcome of a coin toss *qua* observation of the coins propensity is grossly inaccurate, but as an observation of its position on the table it may be completely reliable. Conversely, the relative frequency of heads over a large number of tosses may be reliable as an estimate of the propensity, but fairly hopeless for predicting the outcome of the next toss. The point is that global facts are different from local facts in what they *refer* to. This is how Whewell must be understood when he says that the method of means provides us with something *more true* than the individual facts themselves. *We get rid of the*

⁵⁵ See section 4.

errors of observation only in the sense of being able to separate the wheat from the chaff.

Thus, the scientific realist acknowledges that we do have epistemic access to the inner reaches of reality, and this deeper knowledge teaches us to properly *understand* what we experience. Uneducated perceptual judgments can go horribly wrong in new situations. For example, it was recently speculated that fighter pilots were sometimes crashing because their naive perceptual judgments of the horizon can be distorted in the presence of high *g*-forces. And, the methods of induction that *naively* extrapolate the past into the future can be just as fatal; as in Russell's parable of the man counting windows as he falls from a very tall building, becoming ever more confident that nothing will happen. We can ill afford such mistakes. *Science*, fortunately, succeeds in making the world more controllable and the future more predictable by providing us with information about those deeper aspects of reality that *do* remain unchanged from one situation to the next.

Acknowledgements-- I am very grateful to Peter Forrest, Nicholas Jardine, Martin Leckey, Fritz Rohrlich, Jack Smart, and Elliott Sober for written comments on an earlier draft. The work was carried out under a University Post-Doctoral Fellowship in the Department of Mathematics and Statistics at Monash University, Australia.

References

- Blake, R. M., Ducasse, C. J., and Madden, E. H. (1960). *Theories of Scientific Method: The Renaissance through the Nineteenth Century*. University of Washington Press, Seattle.
- Butts, R. (1968) (ed.) *William Whewell's Theory of Scientific Method*. University of Pittsburgh Press.
- Cajori, F. (1960) (ed.). *Motte's translation of Sir Issac Newton's Principia Mathematica*, as revised by F. Cajori. University of California Press, Berkeley.
- Cartwright, N. (1983). *How the Laws of Physics Lie*. Clarendon Press, Oxford.
- Churchland, P. (1985). 'The Ontological Status of Observables: In Praise of the Superempirical Virtues,' in Churchland & Hooker (1985): 35-47.
- Churchland, P. & Hooker C. (1985). (eds.) *Images of Science*. The University of Chicago Press, Chicago and London.
- Creary, L. G. (1981). 'Causal Explanation and the Reality of Natural Component Forces.' *Pacific Philosophical Quarterly* 62: 148-157.
- Duhem, P. (1974). *The Aim and Structure of Physical Theory*. Atheneum, New York.
- Ellis, B. D. (1965). 'The Origin and Nature of Newton's Laws of Motion.' in R. G. Colodny (ed.), *Beyond the Edge of Certainty*. Prentice Hall.
- Ellis, B. D. (1976). 'The existence of forces.' *Studies in History and Philosophy of Science* 7: 171-186.
- Ellis, B. D. (1985). 'What Science Aims to Do,' in Churchland & Hooker (1985): 48-74.
- Ellis, B. D., Pargetter, R., and Bigelow, J. (1986). 'Forces.' Presented to the *Australasian Association of Philosophy* meetings held at Monash University, August 29, 1986.

- Forster, M. R. (1986). 'Unification and Scientific Realism Revisited,' in Arthur I. Fine and Peter K. Machamer (eds.) *Proceedings of the Philosophy of Science Association*, 1986, Volume 1, 394-405.
- Friedman, M. (1974). 'Explanation and Scientific Understanding.' *The Journal of Philosophy* LXXI 5-19.
- Friedman, M. (1981). 'Theoretical Explanation.,' in R. A. Healy (ed.), *Reduction, Time, and Reality*. Cambridge University Press, Cambridge.
- Friedman, M. (1983). *Foundations of Space-Time Theories*. Princeton University Press: Princeton.
- Hanson, N. R. (1973). *Constellations and Conjectures*, W. C. Humphreys, Jr. (ed.) D. Reidel: Dordrecht-Holland.
- Harper, W. (1983). 'Consilience and Natural Kinds' in Abstracts of Section 7, 7th International Congress of Logic, Methodology and Philosophy of Science, Volume 1: 216-219.
- Hooker, C. A. (1985). 'Surface Dazzle, Ghostly Depths,' in Churchland & Hooker (1985): 153-196.
- Hooker, C. A. (1980). 'Explanation, Generality and Understanding.' *Australasian Journal of Philosophy* 58: 284-290.
- Hunt, I. E. and W. A. Suchting, (1969). 'Force and Natural Motion.' *British Journal for the Philosophy of Science*, 36: 177-94.
- Mill, J. S. (1879). *A System of Logic*. Reprinted in Nagel (1950).
- Musgrave, A. (1985). 'Realism versus Constructive Empiricism,' in Churchland & Hooker (1985): 197-221.
- Nagel, E. (1950) (ed.) *John Stuart Mill's Philosophy of Scientific Method*. Hafner Publishing Company, New York.
- Skyrms, B. (1980). *Causal Necessity*, Yale University Press: New Haven and London.
- Skyrms, B. (1984). *Pragmatics and Empiricism*, Yale University Press: New Haven and London.
- Smart, J. J. C. (1985). 'Laws of Nature and Cosmic Coincidences.' *The Philosophical Quarterly* 35: 272-80.
- Sneed, J. (1970). *The Logical Structure of Mathematical Physics*, D. Reidel.
- Sober, Elliott (1984). *The Nature of Selection*. MIT Press: Cambridge, Mass.
- Sober, E. (1985). 'Constructive Empiricism and the Problem of Aboutness,' *British Journal for the Philosophy of Science*, 36: 11-18.
- van Fraassen, B. C. (1980). *The Scientific Image*. Oxford: Clarendon Press.
- Whewell, W. (1968), in R. Butts (ed.) *William Whewell's Theory of Scientific Method*. University of Pittsburgh Press.
- Wimsatt, W. (1980). 'Randomness and Perceived Randomness in Evolutionary Biology,' *Synthese* 43: 287-39.