

## **A Bayesian Account of the Virtue of Unification**

Forthcoming in *Philosophy of Science*

Wayne C. Myrvold†‡

University of Western Ontario

---

† Send requests for reprints to the author, Department of Philosophy, University of Western Ontario, London, ON, Canada N6A 3K7; e-mail: [wmyrvold@uwo.ca](mailto:wmyrvold@uwo.ca).

‡ The author is grateful to Bill Harper, Sona Ghosh, and three anonymous referees for *Philosophy of Science* for helpful comments and suggestions.

## ABSTRACT

A Bayesian account of the virtue of unification is given. On this account, the ability of a theory to unify disparate phenomena consists in the ability of the theory to render such phenomena informationally relevant to each other. It is shown that such ability contributes to the evidential support of the theory, and hence that preference for theories that unify the phenomena need not, on a Bayesian account, be built into the prior probabilities of theories.

## **1. Introduction**

An account of unification of phenomena by theory should make it clear what reasons, if any, we might have for preferring a theory that provides a unified account of what otherwise might seem to be independent phenomena. Is the ability of a theory to unify phenomena an epistemic virtue, relevant to the degree of confidence we can justifiably place in the theory, or is it merely a pragmatic one? If epistemic, is this based on a priori knowledge that Nature is simple, or does the ability of the theory to provide a unified account of disparate phenomena contribute to the evidential support these phenomena lend to the theory?

There is a surprisingly persistent tradition in the philosophy of science that locates the whole of the empirical support of a theory in its ability to save the phenomena. On such an account, if two theories both account for all available empirical evidence, then the choice between them, if a choice is to be made at all, must rest on extra-empirical considerations. Simplicity and unification are popular candidates for such extra-empirical considerations; these are often held to be pragmatic, or perhaps even aesthetic, virtues, rather than epistemic virtues. We should ask whether simplicity and unification can be defended from the charge of being extra-empirical virtues and regarded instead as contributory to the empirical support of the theory. It will be argued below that these qualities can play a role in the degree to which empirical evidence supports a theory.

It is the purpose of this paper to give a Bayesian account of unification that captures one interesting sense in which a theory can unify phenomena. On this account, the ability of a theory to unify phenomena consists in its ability to render what on prior grounds appear to

be independent phenomena informationally relevant to each other. It will be shown that it is a consequence of Bayes' theorem that this ability does, indeed, contribute to the degree of support provided to the theory by the phenomena unified by the theory; one need not base a preference for theories that unify on a prior belief in the simplicity of Nature. No claim is being made that every case that one can reasonably regard exhibiting unification will be captured by this account. Nor is it claimed that the contribution made by unification to the degree of support that the theory enjoys exhausts what is valuable in unification; as William Harper (1989, 2002*b*) has argued, unification can also result in greater resiliency in the face of apparently disconfirming evidence.

A few words in are in order about what is meant by a Bayesian account, in the context of this paper. Bayesianism, as it is construed here, is an approach to the construction of canons of scientific inference that takes as its central notion *degree of justified belief*, represented by a real-valued function  $\text{Pr}(\cdot | \cdot)$  that takes propositions as its arguments and is assumed to satisfy the axioms of the probability calculus. It will not be assumed that there is a unique correct probability function; this feature of Bayesianism is sometimes misleadingly called *personalism*.

The canons of inference that result are meant to have a status similar to that of the canons of deductive inference. Neither are descriptive of the way in which people actually think. No human being has a deductively closed set of beliefs, and it is likely that no human being has a consistent set of beliefs. These facts are, however, to be regarded as departures from ideal rationality; if one comes to realize that one believes a set  $\{p_1, \dots, p_n\}$  of

propositions but not some logical consequence  $q$  of this set, then one ought not to rest content with this state of affairs but ought to consider whether to accept  $q$  or to reject one or more of  $\{p_1, \dots, p_n\}$ . Similarly, it is doubtful that anyone has numerical degrees of belief satisfying the axioms of probability. If, however, one becomes aware that one's rankings of propositions (perhaps only qualitative) as highly probable, somewhat probable, highly improbable, etc., are inconsistent with the existence of such numerical degrees of belief, then such rankings should be regarded as failing to meet the standards of rationality. Moreover, it will not be assumed that conformity with the axioms of probability will be all that matters; some assignments of probability, on the basis of a given body of evidence, are more reasonable than others, and we will be appealing to the reader's judgments about such reasonableness. The Bayesianism of this paper is, to borrow Shimony (1970)'s term, of the tempered sort.

We will want to consider probabilities of the form  $\Pr(h | e \& b)$ , where  $h$  is a hypothesis,  $e$  some body of evidence being explicitly considered, and  $b$  an appropriate body of background information. The background  $b$  need not be the sum total of facts known to an agent at some time, and, in particular, should not include the evidence  $e$  being considered, as we will want to judge the relevance of  $e$  to  $h$  and hence will want to take  $b$  such that  $\Pr(h | e \& b)$  is not equal to  $\Pr(h | b)$ . It will be assumed that we have available to us judgments (which may be somewhat vague) about the reasonableness of assigning certain probabilities to hypotheses on the basis of bodies of knowledge that differ from the sum total of our own knowledge.

## 2. Examples

Two examples will be used to illustrate the notion of unification to be discussed in this paper, and to guide our Bayesian account of the epistemic value of such unification. The examples we will use are the choice between geocentric and heliocentric world-systems, and Newton's inference to the inverse-square law of gravitation. It will become clear that examples of this sort could be multiplied, and that the features exhibited by these examples that are relevant to our account are quite common in science.

### 2.1 *Copernicus v. Ptolemy*

As is well-known, Earth-based observations concerning the positions (if not the phases) of planets can be predicted, with good accuracy, both by systems such as Ptolemy's, on which the planets orbit the earth, and systems such as the Copernican system or the Tychonic system, on which the planets orbit the sun.<sup>1</sup> This has led to the suggestion that a body of data consisting only of geocentric positions of the planets leaves the decision between the systems a decision to be made on non-empirical grounds; simplicity is frequently invoked as a non-empirical advantage possessed by the Copernican system.

The phenomena to be saved are the apparent motions of the planets with respect to the fixed stars, as seen from Earth. Each planet traverses its path through the zodiac, with its own period. This motion is not uniform, however; the planets proceed at varying rates, at times stopping and reversing direction. In the case of the "superior" planets, Mars, Jupiter

and Saturn, these retrogressions occur only when the planet is near opposition to the sun, with the greatest retrograde angular velocity occurring at opposition. The period between successive oppositions, and hence between successive maximally retrograde motions, is called the *synodic period*.

A geocentric system can recover these qualitative features of the motion of the planets, and achieve a fairly good quantitative fit, by means of a system consisting of a deferent and one epicycle for each planet.<sup>2</sup> The deferent is a circle centered near the Earth. The epicycle is a circle centered on a point on the deferent which, in Ptolemy's system, moves along the deferent at an angular rate that appears constant when viewed, not from the center of the circle, but from an equant point placed on the diameter joining the Earth with the center of the deferent, at the same distance from the center as the Earth but on the opposite side of the center. The planet moves on the epicycle at a constant angular rate. Retrograde motion occurs when the swifter motion of the planet on its epicycle is directed contrary to the motion of the center of the epicycle along the deferent.

The main features of these motions can be recovered by a single circle, centered near the Sun, for each planet. The chief deviations of the apparent motion of the planets from a uniform traversal of the ecliptic, which, on the Ptolemaic system, are accounted for by the epicyclic motion, are, on the Copernican system, due to the motion of the earth about the sun, and, on the Tychonic system, to the motion of the sun, carrying the planets' orbits with it, about the earth. On the Copernican system, apparent retrograde motion occurs, in the case of

the superior planets, when the Earth catches up with and passes the planet in its orbit and, in the case of the inferior planets, when the planet passes the earth.

A heliocentric system is, in a certain sense, simpler than the Ptolemaic, in that “[t]he motion of the earth alone . . . suffices to explain so many apparent inequalities in the heavens” (Copernicus, *Commentariolus*, in Rosen 1959, 59). Rheticus, in his *Narratio Prima*, invokes the frugality of Nature in support of the Copernican hypothesis:

mathematicians as well as physicians must agree with the statements emphasized by Galen here and there: “Nature does nothing without purpose” and “So wise is our Maker that each of his works has not one use, but two or three or often more.” Since we see that this one motion of the earth satisfies an almost infinite number of appearances, should we not attribute to God, the creator of nature, that skill which we observe in the common makers of clocks? For they carefully avoid inserting in the mechanism any superfluous wheel or any whose function could be served better by another with a slight change of position. (Rosen 1959, 137–38)

Galileo has Salviati make an analogous claim, concerning the related point that the apparent diurnal motions of all the heavenly bodies are, on the Copernican system, attributed not to these bodies severally, but to the earth alone:



who is going to believe that nature (which by general agreement does not act by means of many things when it can do so by means by few) has chosen to make an immense number of extremely large bodies move with inconceivable velocities, to achieve what could have been done by a moderate movement of one single body around its own center? (Galileo [1632] 1953, 135)

Both Rheticus and Galileo's Salviati are making strong claims about the nature of the world. Neither gives any argument beyond authority, Rheticus citing Galen, Salviati the authority of "general agreement" and, a few pages later, in a shrewd rhetorical move, of Aristotle.<sup>3</sup> Nor can it be claimed that such a principle is justified on empirical grounds; the aspect Nature presents to our observation contains a bewildering mixture of parsimony and profligacy.

Simplicity is sometimes held to be a pragmatic virtue, rather than an epistemic one. Even if the simplicity of a theory is not grounds for believing in its approximate truth, it may be a pragmatic reason for preferring one theory over the other, because a simpler theory is easier to work with. Both the Copernican and Ptolemaic systems are capable of recovering the phenomena; to a first approximation, at least, the Copernican system does so with fewer circles, and for this reason has been regarded as simpler. That the Copernican system surpasses the Ptolemaic in this sort of simplicity, when all complications of the two systems are taken into account, has been disputed (see, *e.g.* Neugebauer 1957, 204; Dijksterhuis 1961, 294; Hoskin and Gingerich 1999, 88); to achieve the same precision as Ptolemy without the

use of Ptolemy's equants, Copernicus had to introduce complications into the simple system outlined above. Kuhn declared that, since the Copernican system fails to surpass the Ptolemaic on grounds of either simplicity (measured by circle-count) or accuracy of predictions, there were, at the time of Copernicus, not only no epistemic grounds for preferring the Copernican system over the Ptolemaic, but also no pragmatic grounds.

But this apparent economy of the Copernican system, though it is a propaganda victory that the proponents of the new astronomy rarely failed to emphasize, is largely an illusion . . . Copernicus, too, was forced to use minor epicycles and eccentrics. His full system was little if any less cumbersome than Ptolemy's had been. Both employed over thirty circles; there was little to choose between them in economy. Nor could the two systems be distinguished in accuracy. When Copernicus had finished adding circles, his cumbersome sun-centered system gave results as accurate as Ptolemy's but it did not give more accurate results. Copernicus did not solve the problem of the planets. (Kuhn 1959, 169)

The only superiority of the Copernican system over the Ptolemaic system, according to Kuhn, lay in its greater harmony and beauty.

as Copernicus himself recognized, the real appeal of sun-centered astronomy was aesthetic rather than pragmatic. To astronomers the initial choice between Copernicus' system and Ptolemy's could only be a matter of taste, and matters of taste are the most difficult of all to define or debate. Yet, as the Copernican Revolution itself indicates, matters of taste are not negligible. (1959, 172)

This conclusion is, I believe, premature. Let us consider a bit more carefully the *manner* in which each of the two systems saves the phenomena.

As mentioned, the Ptolemaic system recovers the bulk of the variation of the angular speed of each planet via an epicycle for each planet; the Copernican system accounts for this variation, for all of the planets, by the orbital motion of the earth. The separate epicycles of the planets are, on the Ptolemaic system, kinematically independent; there is no kinematically necessary relation between the motion of any two epicycles, or between the motion of one planet's epicycle and the motion of any other planet. By adjusting the speed of the planet on its epicycle, one could produce any number of episodes of retrograde motion per synodic period, in place of the single episode that is observed. It so happens, however, that the epicyclic motion required on the Ptolemaic system to recover the *actual* observed behaviour of the planets displays a curious correlation with the motion of the sun. The period of a planet's epicyclic motion is its synodic period, if calculated, as Ptolemy does, with respect to the line joining the earth with the center of the epicycle; this means that, with respect to a fixed reference direction, the epicyclic period of a superior planet is the same as that of the

sun, and, moreover, the radius drawn from the center of the epicycle to the planet remains parallel, at all times, to the radius drawn from the Earth to the mean sun. For the inferior planets, the period of the center of the epicycle in its journey around the deferent is equal to the sun's period, and the radius between the Earth and center of the epicycle passes through the mean sun at all times. This correlation between the motions of the planets and that of the sun was, of course, recognized by Ptolemy, who, in setting out his preliminary hypotheses for the planets, says that there are "two apparent anomalies for each planet: that anomaly which varies according to its position on the ecliptic, and that which varies according to its position relative to the sun" (Ptolemy 1984, 442). Michael Hoskin, in *The Cambridge Concise History of Astronomy*, remarks that "[t]his unexplained involvement of the Sun in the geometry of the other planets was to puzzle later astronomers" (Hoskin 1999, 47); Georg Peurbach (1423–61), for one, was no heliocentrist but nevertheless observed, "It is clear that each of the six planets in its motion shares something with the Sun, and the Sun is, so to speak, the common mirror and measure for their motions" (quoted in Hoskin and Gingerich 1999, 88–89).

On the Ptolemaic system, then, the longitudinal motion of each planet (other than Mercury, for which Ptolemy had to introduce a more complicated hypothesis) is composed of two circular motions: one whose period is peculiar to the planet, and one whose period is equal to that of the sun about the earth. On the Tychonic system, the orbits of the planets are centered on the sun, and so this second motion simply *is* the sun's motion. On the Copernican system, the apparent motion of each planet is again composed of two circular motions: the

planet's own motion, and the motion of the observer's vantage point on the Earth. The component of motion recognized by Ptolemy to be related to the sun, is, on either the Tychonic or the Copernican system, the relative motion of the sun and earth.

As Kepler clearly perceived, the Copernican system explains what on the Ptolemaic system is a mysterious connection between the sun and the motion of the other planets:

For in the first place one might ask of Ptolemy how it comes about that the three eccentrics of the Sun, Venus, and Mercury have equal times of revolution? . . . Why do the five planets make retrogressions, whereas the luminous stars do not? . . . Similarly the ancients rightly wondered why the three superior planets are always in opposition to the Sun when they are at the bottom of their epicycles, but in conjunction when they are at the top. (Kepler [1596] 1981, 81)

On the Ptolemaic system, the periods of the planets and the intervals between episodes of retrograde motion are independent parameters. It so happens that the maximal retrogressions of the superior planets occur at intervals equal to their synodic periods (which are calculable from the periods of the planet and that of the sun). On the heliocentric explanation of retrograde motion, however, fixing the periods of the planets (including the Earth) fixes also the intervals between episodes of retrogression. A system, such as Tycho's or Copernicus', which centers the orbits of the planets on the sun, makes one set of

phenomena—the mean apparent motions of the planets, and the sun, along the ecliptic—carry information about what, on the Ptolemaic system, are independent phenomena, the deviations of the planets from their mean apparent motions.

## 2.2 Newton and the inverse-square law of gravitation

Proposition 2 of Book 3 of Newton's *Principia* states:

*The forces by which the primary planets are continually drawn away from rectilinear motions and are maintained in their respective orbits are directed to the sun and are inversely as the squares of their distances from its center.*

(Newton [1726] 1999, 802)

For the first part of this proposition, the sun-directedness of the force, Newton cites Phenomenon 5 of Book 3, namely, Kepler's Area Law. For the second part, the inverse-square dependence of the force on distance from the sun, he cites two phenomena. One is Phenomenon 4, Kepler's Harmonic Law: the periods of the planets are proportional to the  $3/2$  power of their mean distances from the sun. He then adds, "But this second part of the proposition is proved with the greatest exactness from the fact that the apheia are at rest. For the slightest departure from the ratio of the square would (by book 1, prop. 45, corol. 1) necessarily result in a noticeable motion of the apsides in a single revolution and an immense such motion in many revolutions."

The relevance of these two phenomena to the force law are provided, respectively, by Proposition 4 of Book 1, and Proposition 45 of the same book. The argument from the Harmonic Law to the inverse-square law is the familiar one found in many physics textbooks. Proposition 4 concerns the centripetal acceleration of uniform circular motion. If the planets move in orbits that approximate uniform circular motion about the sun, and their periods are some function  $T(r)$  of their mean distances from the sun, their accelerations will satisfy

$$a(r) \propto r / T(r)^2. \quad (1)$$

This is Newton's Corollary 2 to Proposition 4. Putting  $T(r) \propto r^{3/2}$  yields Corollary 6,

$$a(r) \propto 1 / r^2. \quad (2)$$

That is, the accelerations of the planets towards the sun vary inversely as the square of their distances from the sun.

The argument from the quiescence of the apsides to the inverse-square dependence of the force law is less familiar (see Harper 2002a for a clear exposition). Newton was able to show (Bk. 1, Props. 43, 44) that, if a body moves in an orbit under the action of a centripetal force  $f(r)$ , adding an inverse-cube term to the force produces a second orbit, whose distance from the center of force at any time is the same as that of the first orbit ( $r_2(t) = r_1(t)$ ), and

whose angular displacement is a constant multiple of that of the first ( $\theta_2(t) = \alpha \theta_1(t)$ ) (this is posed as an exercise problem by Goldstein 1980, 123; see also Whittaker 1944, 83; Chandrasekhar 1995, 184). Since Newton had shown that an inverse-square force law produces a quiescent elliptical orbit, he is able to conclude that a force-law that takes the form of an inverse-square term plus an inverse-cube term produces a precessing ellipse. Newton then proceeds (Prop. 45) to approximate, over a small range of  $r$ , an arbitrary force  $f(r)$  by a sum of an inverse-square term and an inverse-cube term by taking (what we now call) a Taylor series expansion, to the first power in  $r$ , of  $g(r) = r^3 f(r)$  around the aphelion distance  $a$  of the orbit. This approximation yields the result that for each revolution there will be an advance of the aphelion by an amount (measured in degrees) equal to

$$p = 360 \left( \sqrt{g(a)/ag'(a)} - 1 \right). \quad (3)$$

Therefore, for orbits that are approximately circular, so that this Taylor series approximation is valid, a measurement of the aphelion advance of a planet yields information about the distance-dependence of the force on the planet, in the small range of distances explored by the planet.<sup>4</sup>

Suppose we make the hypothesis  $h_{\text{PL}}$  that, to a high degree of approximation, the acceleration of all of the planets is due to a single field of force centered on the sun, obeying some power-law:

$$f(r) = k r^\lambda. \quad (4)$$



Application of Newton's Proposition 45 to such a law gives, for an approximately circular orbit,

$$p = 360 \left( \frac{1}{\sqrt{\lambda + 3}} - 1 \right), \quad (5)$$

or,

$$\lambda = 1 / (p/360 + 1)^2 - 3. \quad (6)$$

A quiescent orbit, for which  $p = 0$ , gives us  $-2$  as an estimate of the parameter  $\lambda$ , and this in turn entails Kepler's Harmonic Law, that the periods of the planets vary as the three-halves power of their distances from the sun. On the hypothesis  $h_{PL}$  of a single power-law force, a measurement of the perihelion precession (or lack thereof) of each planet yields information about the exponent of the power law and thereby yields information about the dependence of their relative periods on their distances, and vice versa. What a priori are independent phenomena, namely, the approximate quiescence of the aphelia of the planets, and Kepler's Harmonic Law, are no longer independent on the supposition  $h_{PL}$  of a single heliocentric power-law force; if such a hypothesis is true or approximately true, each of the two phenomena strongly constrains the other.

The hypothesized form of the force law can be replaced by more general families of functions. In order for the hypothesis to make our two a priori independent phenomena yield

information about each other, all that is required of the hypothesized form of the force law is that the behaviour of the force law over the range of distances explored by one planet yield information about its behaviour at other planetary distances. If, on the other hand, one merely hypothesized that the force law is some function, with no preference given to any function over another—that is, if one had a prior probability measure that was sufficiently uniform over the space of possible force laws—then such a hypothesis would leave the a priori independent phenomena independent.

These two examples suggest an interesting way in which a hypothesis can unify disparate phenomena: the hypothesis can make two phenomena that in the absence of the hypothesis seem to be independent phenomena yield information about each other. In the next section we will tender an explication, in Bayesian terms, of information yielded by one proposition about another, and of unification of a body of evidence by a hypothesis. Such an account seems to capture at least part of the intuition invoked by Michael Friedman:

this is the essence of scientific explanation—science increases our understanding of the world by reducing the total number of independent phenomena that we have to accept as ultimate or given. A world with fewer independent phenomena is, other things equal, more comprehensible than one with more. (Friedman 1974, 15)

The Bayesian account of such unification will differ greatly in its details from the account given by Friedman; in particular, Friedman's notion of independent acceptability will be replaced by the notion of probabilistic independence; a unifying hypothesis will reduce, not the number of independently acceptable phenomena, but the degree of informational independence of a body of phenomena.

### **3. Informational relevance and evidential support**

We wish to give a Bayesian account of this notion of unification of disparate phenomena, which consists of the ability of the theory to make the phenomena yield information about each other. For this, we will need a Bayesian notion of the information one proposition  $p$  yields about another proposition  $q$ . We will assume a body of background knowledge  $b$ , and that we have available a probability function  $\Pr(\cdot | b)$ . Learning that a proposition  $p$  is true can be either positively or negatively relevant to another proposition  $q$ ; if it is positively relevant to  $q$ , then learning  $p$  furnishes a certain amount of information about whether or not  $q$  is true. Moreover, the relevance of  $p$  to  $q$ —that is, how much we learn about whether or not  $q$  is true when we learn that  $p$  is true—is the sort of thing that admits degrees. Accordingly, we will want to define a measure of the informational relevance of  $p$  to  $q$ , on background  $b$ . We will call the degree of informational relevance of  $p$  to  $q$ , on background  $b$ ,  $I(q, p | b)$ .  $I(q, p | b)$  will be assumed to be definable in terms of the probability function  $\Pr(\cdot | b)$  (and, in fact, continuously definable, so that small changes in probabilities yield small changes in information). By convention, we will take this quantity to be positive when

$p$  is positively relevant to  $q$ , negative when  $p$  is negatively relevant to  $q$ , and zero when  $p$  is irrelevant to  $q$ . Furthermore, we will define our measure of informational relevance so that independent evidence is additive—when  $p_1$  and  $p_2$  are independent items of evidence, the information yielded about  $q$  by the conjunction of  $p_1$  and  $p_2$  will be simply the sum of the information yielded by  $p_1$  and the information yielded by  $p_2$  (see the appendix for a precise statement of this condition). Finally, we will assume a normalization convention that agrees with the convention adopted in information theory for the special case when our information about  $q$  amounts to certainty that  $q$  obtains. If  $q$  is one of  $2^N$  equiprobable, mutually exclusive, and jointly exhaustive alternatives, then the information that  $q$  obtains amounts to  $N$  bits of information, and, in general, information that  $q$  obtains will count as  $-\text{Log}_2(\text{Pr}(q | b))$  bits of information.

It is shown in the appendix that these conditions uniquely determine a measure of degree of informational relevance:<sup>5</sup>

$$I(q, p | b) = \text{Log}_2(\text{Pr}(q | p \& b) / \text{Pr}(q | b)). \quad (7)$$

It is worth noting that the informational relevance function  $I(q, p | b)$  so obtained is symmetric in its two arguments:  $I(q, p | b) = I(p, q | b)$ .

Suppose that  $p_1, p_2$  are independent phenomena; that is,  $I(p_1, p_2 | b) = 0$ . A hypothesis  $h$  makes  $p_2$  yield information about  $p_1$  if  $I(p_1, p_2 | h \& b) > 0$ , and, in such a case, we can take  $I(p_1, p_2 | h \& b)$  as a measure of the extent to which  $h$  makes  $p_2$  yield information

about  $p_1$ . If  $I(p_1, p_2 | b) \neq 0$ , the excess of  $I(p_1, p_2 | h \& b)$  over  $I(p_1, p_2 | b)$  measures the extent to which  $h$  makes  $p_2$  yield information about  $p_1$ . Let us call this quantity  $U$ , as it is a measure of the extent to which  $h$  unifies the set of phenomena  $\{p_1, p_2\}$ .<sup>6</sup>

$$U(p_1, p_2; h | b) = I(p_1, p_2 | h \& b) - I(p_1, p_2 | b). \quad (8)$$

We will also want to consider bodies of phenomena  $\{p_1, p_2, \dots, p_n\}$  consisting of more than two elements. We define the  $n$ -place function  $I^{(n)}$ , which is a measure of the degree of mutual dependence to be found in the set  $\{p_1, p_2, \dots, p_n\}$ , as the information about  $p_2$  yielded by  $p_1$ , plus the information about  $p_3$  yielded by the conjunction of  $p_1$  and  $p_2$ , and so on, up to the information about  $p_n$  yielded by the conjunction of all the other members of the set. This gives

$$I^{(n)}(p_1, p_2, \dots, p_n | b) = \text{Log}_2(\text{Pr}(p_1 \& p_2 \& \dots \& p_n | b) / \text{Pr}(p_1 | b) \text{Pr}(p_2 | b) \dots \text{Pr}(p_n | b)). \quad (9)$$

It can easily be seen that this quantity is also independent of the order in which the elements are taken. We now define the quantity  $U^{(n)}$ , which is the degree to which  $h$  unifies the set  $\{p_1, p_2, \dots, p_n\}$ :

$$U^{(n)}(p_1, p_2, \dots, p_n; h | b) = I^{(n)}(p_1, p_2, \dots, p_n | h \& b) - I^{(n)}(p_1, p_2, \dots, p_n | b). \quad (10)$$

The question to be asked now is whether the ability of a hypothesis to unify a set of phenomena contributes to the degree to which the phenomena lend evidential support to the hypothesis. To ask this we will need a measure of the degree to which evidence  $e$  supports hypothesis  $h$  on background  $b$ . The quantity  $\Pr(h | e \& b) / \Pr(h | b)$ , or its logarithm, is a popular choice for such a measure, and is often referred to as the “degree of confirmation” of  $h$  by  $e$  on background  $b$ , confirmation being taken in the relative, or incremental sense, rather than the absolute—that is, to say that  $e$  confirms  $h$ , in this sense, is to say that  $e$  is positively relevant to  $h$ , and is not to say that  $e$  provides sufficient grounds for accepting  $h$  (for discussions of this distinction, see Carnap 1962, xv–xix; Salmon 1975, 3–36). The logarithm of  $\Pr(h | e \& b) / \Pr(h | b)$  is the informational relevance  $I(h, e | b)$ . This quantity, is, therefore, a candidate for a measure of the degree to which a piece of evidence supports a hypothesis. Another candidate is what I. J. Good (1950; 1983) has called the “weight of evidence,”

$$\begin{aligned}
 W(h, e | b) &= I(h, e | b) - I(\sim h, e | b) \\
 &= \text{Log}(\Pr(h | e \& b) / \Pr(\sim h | e \& b)) - \text{Log}(\Pr(h | b) / \Pr(\sim h | b)) \\
 &= \text{Log}(\Pr(e | h \& b) / \Pr(e | \sim h \& b)). \tag{11}
 \end{aligned}$$

It will not be necessary to adjudicate between these two candidates, as there is an interesting relation between the power of a hypothesis to unify a body of evidence and its degree of confirmation on either choice of *explicatum* for the latter. On either choice of measure of degree of evidential support, it can be shown that the ability of a hypothesis to

unify a body of evidence contributes in a direct way to the support provided to  $h$  by the body of evidence. It follows from Bayes' theorem that<sup>7</sup>

$$I(h, e_1 \& e_2 | b) = I(h, e_1 | b) + I(h, e_2 | b) + U(e_1, e_2; h | b). \quad (12)$$

That is: the degree of support provided to  $h$  by  $e_1$  and  $e_2$  taken together is the sum of three terms: the degree of support of  $h$  by  $e_1$  alone, the degree of support of  $h$  by  $e_2$  alone, and an additional term which is simply the degree of unification of the set  $\{e_1, e_2\}$  by  $h$ . An analogous result holds for larger bodies of evidence; the degree of support provided to  $h$  by the set  $\{e_1, e_2, \dots, e_n\}$  taken together is simply the sum of the degrees of support provided to  $h$  by each of the evidence items taken individually, plus an additional term which is the degree to which  $h$  unifies the body of evidence.

To understand what equation (12) means, suppose that two phenomena,  $e_1$  and  $e_2$ , which on prior grounds are regarded as having little to do with each other, both occur. If  $h_1$  makes  $e_1$  informationally relevant to  $e_2$ —that is, if the truth of  $h_1$  would render it more probable that  $e_2$  occurs if  $e_1$  does—then the joint occurrence of  $e_1$  and  $e_2$  is more probable on the supposition of  $h_1$  than it would be if  $h_1$  didn't unify  $e_1$  and  $e_2$  in this sense. Even if the likelihoods<sup>8</sup> of two hypotheses  $h_1$  and  $h_2$  are equal to each other on evidence  $e_1$  taken alone, and on  $e_2$  taken alone, if  $h_1$  unifies the pair  $\{e_1, e_2\}$  by making them informationally relevant to each other, and  $h_2$  doesn't, then the likelihood of  $h_1$  on the evidence  $e_1 \& e_2$  is higher than that of  $h_2$ , and consequently  $h_1$  is better supported by  $e_1 \& e_2$  than  $h_2$  is. Furthermore, if  $h_1$

unifies the pair  $\{e_1, e_2\}$  *more* than  $h_2$  does by making them *more* relevant to each other, then  $h_1$  is again better supported by  $e_1 \& e_2$  than  $h_2$  is.

For Good's weight of evidence, we have an analogous result,

$$W(h, e_1 \& e_2 | b) = W(h, e_1 | b) + W(h, e_2 | b) + U(e_1, e_2; h | b) - U(e_1, e_2; \sim h | b). \quad (13)$$

Here what counts is the *excess* of the degree of unification of the evidence by  $h$  over its degree of unification by  $\sim h$ . In this case also the corresponding result holds for larger bodies of evidence.

On this Bayesian account, therefore, the power of a theory to unify a body of evidence is *not* an extra-empirical virtue but contributes directly to the degree to which the evidence supports the theory. Note that these results do not depend on a special form of the prior probabilities and will hold for *any* probability assignment; in particular, it is not necessary to build a preference for unification or simplicity into the assignment of prior probabilities. Nor do we have to invoke a preference for unification as an extra, supplementary rule, beyond the usual Bayesian updating rules, or add an extra boost of confirmation to the unifying theory beyond the degree of confirmation it receives from simple Bayesian conditionalization.<sup>9</sup>



#### 4. Application to our examples

Let  $b$  be a body of background knowledge including qualitative information about the apparent course of the planets (including, perhaps, the fact that they undergo retrograde motion), but not including precise values of the periods of the planets, or detailed information about their retrogressions. This body of knowledge is, of course, not the body of knowledge possessed by any of the readers of this paper; the motivation for considering such a body of knowledge will be made clear shortly. Let  $h_C$  be the hypothesis that some system following the outlines of the Copernican system, as sketched in section 2.1 above, is true. Note that  $h_C$  by itself makes no specific predictions as to the observed location of any planet at any time, as it contains a number of parameters—the size of the planetary orbits, their periods, and the location of the planets at some initial epoch—that must be filled in from empirical data. Similarly, let  $h_P$  be the hypothesis that some Ptolemaic system is correct, again, with parameters concerning the periods of the planets, and the frequency, time, and length of their retrogressions, left unspecified.

We suppose that we have some probability distributions, conditional on background  $b$ , over the periods of the planets and of the times of their retrogressions. Let  $p_m$  be the statement that Mars traverses the sphere of fixed stars with a period that, within a small observational error, is equal to 1.88 years, and let  $r_m$  be the statement that it retrogresses when it is near opposition to the sun, within a period equal to 2.14 years (plus or minus a small observational error). Now, the Copernican hypothesis  $h_C$  entails that  $p_m$  holds if and only  $r_m$

does; hence, these two propositions, conditional on  $h_C$ , have a maximal degree of informational relevance to each other:

$$I(r_m, p_m | h_C \& b) = -\text{Log}_2(\text{Pr}(r_m | h_C \& b)). \quad (14)$$

Since, by hypothesis, the background knowledge  $b$  contains only qualitative information about the motions of the planets, it is not to be expected that one could anticipate, in advance, the precise values of these parameters; hence, the quantity  $\text{Pr}(r_m | h_C \& b)$  ought to be quite small, and  $I(r_m, p_m | h_C \& b)$  will be quite large. If, conditional on all serious rivals to  $h_C$ ,  $p_m$  and  $r_m$  are independent, then  $I(r_m, p_m | b)$  will be fairly small, and so  $U(r_m, p_m; h_C | b)$  will be positive— $h_C$  unifies  $\{p_m, r_m\}$ , relative to background  $b$ .

On the other hand, if only the bare-bones Ptolemaic hypothesis  $h_P$  is assumed, then  $p_m$  affords little or no information about whether or not  $r_m$  is true, and so a reasonable probability assignment will have them informationally independent, or nearly so, conditional on  $h_P$ :

$$\text{Pr}(p_m \& r_m | h_P \& b) \approx \text{Pr}(p_m | h_P \& b) \text{Pr}(r_m | h_P \& b), \quad (15)$$

$$I(p_m, r_m | h_P \& b) \approx 0. \quad (16)$$

Let us assume that the evidential support lent to  $h_C$  by  $p_m$  on background  $b$  is approximately the same as the evidential support lent to  $h_P$  by  $p_m$  on the same background (and, indeed, since  $p_m$  merely tells us what the period of Mars is, it would seem that the support lent to *either* hypothesis alone by  $p_m$  is nil), and similarly, that the degree of support lent to  $h_C$  by  $r_m$  on background  $b$  is approximately the same as the evidential support lent to  $h_P$  by  $r_m$ . Then the degree of support lent to  $h_C$  by the conjunction of  $p_m$  and  $r_m$  will be considerably greater than the degree of support lent to  $h_P$  by this conjunction — since  $h_C$  unifies  $p_m$  and  $r_m$  in the sense of making them informationally relevant to each other, they work together to support  $h_C$ .

Now, of course, it doesn't follow from this that every probability assignment will have  $\Pr(h_C | p_m \& r_m \& b)$  greater than  $\Pr(h_P | p_m \& r_m \& b)$ . What does follow is that any probability assignment satisfying the conditions outlined above will have  $\Pr(h_C | p_m \& r_m \& b)$  greater than  $\Pr(h_P | p_m \& r_m \& b)$  *unless* it also has the prior probability of  $h_C$  considerably less than that of  $h_P$ . An agent who attaches low prior probability to the Copernican hypothesis (perhaps on the basis of considerations of terrestrial dynamics;<sup>10</sup> see Ptolemy 1984, 44–45) should nevertheless acknowledge that the ability of the Copernican hypothesis to explain the otherwise puzzling correlation between the motions of the planets and that of the sun counts in favor of the hypothesis. Note that the procedure here is to assess the reasonableness of probabilities conditional on  $p_m \& r_m \& b$  by considering what sorts of probability assignments are reasonable on  $b$  alone—which was the motivation for

introducing  $b$ . Such considerations should be available even to agents who already know  $p_m$  and  $r_m$ .

Since a parallel discussion can be applied as well to the other planets, the degree of support lent to the Copernican hypothesis by planetary phenomena becomes considerably stronger when the other planets are taken into account.

So far we have contrasted a bare-bones Ptolemaic hypothesis with a bare-bones Copernican hypothesis; in particular, no relation between the motion of the planet on its epicycle and the sun was built into the Ptolemaic hypothesis  $h_p$ . The system presented by Ptolemy in the *Almagest*, however, does contain restrictions of just this sort: for the superior planets Mars, Jupiter, and Saturn, the line drawn between the planet and the center of its epicycle remains at all times parallel to the earth-sun radius, whereas for the inferior planets Mercury and Venus, the line from the Earth to the center of the planet's epicycle passes through the sun. Call this set of restrictions the condition of sun-planet parallelism. Although it is plausible that the bare-bones Ptolemaic hypothesis was considered by some astronomer at some point, it must be admitted that we have no historical record of such an episode. It seems only fair, therefore, that we consider what happens when the Copernican hypothesis  $h_C$  faces as a rival, not the minimal Ptolemaic hypothesis  $h_p$ , but a strengthened Ptolemaic hypothesis that includes the sun-planet parallelism condition.<sup>11</sup>

Call the strengthened Ptolemaic hypothesis  $h_{SP}$ . This hypothesis shares with the Copernican hypothesis the feature of entailing that  $r_m$  obtains if and only  $p_m$  does, and hence

enjoys the same evidential boost from the conjunction of  $r_m$  and  $p_m$  enjoyed by  $h_C$ ; the ratio of posterior to prior should be equal, or at least approximately equal, for the two hypotheses.

$$\Pr(h_{SP} | r_m \& p_m \& b) / \Pr(h_{SP} | b) \approx \Pr(h_C | r_m \& p_m \& b) / \Pr(h_C | b) \quad (17)$$

Suppose, now, that we have

$$\Pr(h_C | b) \approx \Pr(h_P | b), \quad (18)$$

where  $h_P$  is, as before, the minimal Ptolemaic hypothesis. Suppose, further, that we also have

$$\Pr(h_{SP} | b) \ll \Pr(h_P | b). \quad (19)$$

This seems eminently reasonable; since  $h_{SP}$  is a strengthening of  $h_P$ , we must have  $\Pr(h_{SP} | b) \leq \Pr(h_P | b)$ , and to have  $\Pr(h_{SP} | b)$  equal to  $\Pr(h_P | b)$  is to attach zero probability, on background  $b$ , to the possibility that some Ptolemaic hypothesis not satisfying the sun-planet parallelism condition is true. Moreover, it seems unreasonable to assume that the background knowledge  $b$  suffices to permit one to expect the sun-planet parallelism condition to be true; if this is right, we should take  $\Pr(h_{SP} | b)$  to be substantially less than  $\Pr(h_P | b)$ .

From (17), (18), and (19) it follows that

$$\Pr(h_{SP} | p_m \& r_m \& b) \ll \Pr(h_C | p_m \& r_m \& b). \quad (20)$$

It seems, therefore, that the only way that a reasonable agent can avoid having  $\Pr(h_C | p_m \& r_m \& b)$  be much larger than  $\Pr(h_{SP} | p_m \& r_m \& b)$  is to have the prior probability of the Copernican hypothesis  $h_C$  be much smaller than the prior probability of the bare-bones Ptolemaic hypothesis  $h_P$ .

Now, what Ptolemy presented in the *Almagest*, and Copernicus in *De Revolutionibus*, were fully specified models of the heavens, with all parameters filled in. From such models precise predictions can be deduced. If we let  $H_P$  and  $H_C$  be these fully specified Ptolemaic and Copernican hypotheses, respectively, then these hypotheses actually entail the observed phenomena, and the likelihoods that appear in Bayes' theorem will be equal to unity. But any two items of evidence  $e_1$  and  $e_2$ , if entailed by a hypothesis  $H$ , are probabilistically independent conditional on  $H$ , and hence not informationally relevant to each other, conditional on  $H$ .

Considerations similar to those discussed above will apply. Since the unsaturated Copernican hypothesis  $h_C$  has a greater ability to unify the celestial phenomena than does the unsaturated Ptolemaic hypothesis  $h_P$ , then, if  $e$  is a body of evidence containing facts about the mean motions of planets and their retrogressions, we will have  $\Pr(H_C | e \& b) \gg \Pr(H_P | e \& b)$  unless  $\Pr(h_C | b) \ll \Pr(h_P | b)$ .

The analysis of our second example is similar. It is not as clear in this case what rival hypotheses are to be considered, but let us contrast, by way of example, a theory (call it  $h_1$ ) that posits a single power-law force affecting all the planets, with a theory (call it  $h_2$ ) that posits independent power-law forces acting on each of the planets. Let  $b$  be a body of knowledge containing qualitative facts about the motion of the planets but not containing numerical information sufficient to fix the values of the parameters appearing in  $h_1$  and  $h_2$  corresponding to the exponents of the power laws and the strengths of the force fields. According to  $h_1$ , measurement of the rate of precession of any planet furnishes information about the precession rates of the other planets and also about the relation between the periods and distances of the planets. On the assumption of a single power-law force, once we have obtained the exponent of the power-law from a measurement of precession rates, measurement of the period and orbital radius of one planet permits the orbital radii of the others to be predicted from their periods (or vice versa).

On  $h_2$ , these quantities could be regarded as informationally independent, but they need not be—it is, after all, compatible with  $h_2$  that the exponents of the separate power laws for each of the planets be all equal, or approximately so, and one might even attach high prior probability to this being the case. However, one who takes  $h_2$  as a serious rival to  $h_1$  ought to attach non-negligible probability to the possibility that the several distinct forces do not perfectly mimic a single acceleration field—if one regards it as inevitable that these distinct forces perfectly mimic the action of a single force, it is hard to understand what is meant by calling them “distinct.” Suppose, then, that one does assign a non-negligible probability to

the several force laws having different values of their parameters, either in the exponent of the power law or in the strength of the force. On such a probability assignment,  $h_1$  will do a better job of unifying the phenomena than  $h_2$ , and will have a correspondingly higher degree of support.

The inference that results from an application of Newton's Rule 1, that "No more causes of natural things should be admitted than are both true and sufficient to explain their phenomena," is justified in this case without any commitment to a principle that "Nature does nothing in vain" (Newton [1726] 1999, 784). What counts, instead, is the fact that the supposition that these forces are the same makes inevitable a relation that would be a puzzling coincidence on the supposition of independent forces.

## **5. Relation to some other views**

### *5.1. Whewell on consilience.*

The sorts of cases considered in this paper as examples of unification, are, at the very least, reminiscent of what Whewell called *consilience of inductions*.

the evidence in favour of our induction is of a much higher and 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 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, when 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 particular feature in their evidence, I will take the liberty of describing it by a particular phrase; and will term it the *Consilience of Inductions*. (Whewell 1847, 65; Butts, ed., 153)

On a common reading of Whewell's account of consilience, the unifying hypothesis is a fully specified hypothesis that *entails* evidence from disparate domains. If this were the whole story according to Whewell, there would little room for a close parallel between our account and Whewell's, as two evidence statements that are both entailed by a hypothesis *h* are informationally irrelevant to each other, conditional upon *h*.<sup>12</sup> We should bear in mind, however, that for Whewell induction is a multi-stage process. The three steps of induction, according to Whewell, are the selection of the idea, the construction of the conception, and the determination of the magnitudes (Whewell 1847, 380; Butts, ed., 211). For scientific theories formulated mathematically, these steps become the selection of the independent variable, the construction of the formula, and the determination of the coefficients (382; 213).

Thus, according to Whewell, the induction that leads to a law such as Newton's law of gravitation includes a stage in which one is considering a formula containing parameters, called by Whewell "coefficients," to be filled in empirically. In such a case, a consilience of inductions would occur when the values of certain parameters can be determined from two different sorts of phenomena, and the values determined from one class of phenomena agree with those determined from another. Note that Whewell speaks of a hypothesis being "adjusted to one class of phenomena" and then found to represent another class; he is clearly contemplating situations in which initially unsaturated hypotheses are considered, and the values of their parameters filled in empirically. Malcolm Forster's interpretation of Whewell's views on consilience seems to be correct, at least where mathematically formulated theories are concerned: "*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*" (Forster 1988, 76). A hypothesis that entails such a law-like connection will render the facts in the disparate domains connected by the law informationally relevant to each other. The account of unification offered in this paper, therefore, seems to mesh nicely with Whewell's conception of consilience.

### *5.2 Forster and Sober.*

Much of what has been said in this paper is also reminiscent of Forster and Sober's (1994) account of the epistemic virtue of simpler, more unified theories. Here, however, the

resemblance goes less deep. According to Forster and Sober, the virtue of simplicity and unification is to be found in the resistance of such theories to the phenomenon of overfitting, which is the tendency for the best-fit curve from an overly capacious family of functions to track observational errors instead of the systematic dependencies one is trying to capture. Provided that the statistical distribution of the parameters estimated from the data around their optimal values is at least approximately normal, the tendency towards overfitting will be roughly linear in the number of degrees of freedom of the family of functions considered, and this is why a best-fit curve from a family of functions with fewer degrees of freedom may yield a better fit to future data than the best-fit curve from a family with more. For a fixed family of functions, overfitting is diminished if observational errors are diminished, or if the body of data is enlarged. Therefore, the weight to be attached to simplicity and unification, according to Forster and Sober, should decrease as observations are made more precise or more data points are added.

Let us consider two hypotheses  $h_1$  and  $h_2$ , which assert that the force law for the influence of the sun on the planets is to be found in families of functions  $F_1$  and  $F_2$ , respectively. It may happen that, if  $F_1$  is a family of fewer degrees of freedom, then  $h_1$  makes the quiescence of the apsides yield more information about the harmonic law than does  $h_2$ . This may seem to provide an important point of connection between the account of the virtue of unification offered in this paper and that given by Forster and Sober. There is an important difference, however; on the account given here, what counts is, not the number of degrees of freedom of a family of functions, but the extent to which a hypothesis makes

one set of phenomena constrain another, which in our example consists of the extent to which the rate of change of the force over the distances explored by a single planet constrains the relative strength of the force at other planetary distances. On Forster and Sober's account, the virtue of "simpler, more unified" theories lies solely in the ability of such theories to resist overfitting the data, and hence is diminished when the data are made more accurate or when the number of data points is increased. On the account given here, the ability of a theory to unify a body of phenomena lends support to the theory that goes well beyond such resistance to overfitting and persists when random errors in the data are diminished. This is not to say that the statistical considerations invoked by Forster and Sober will play no role—they will, in estimation of parameters from the data, and in error analysis. But the chief importance of theoretical unification is not to be found in such considerations.

## Appendix

The conditions that our measure of information relevance,  $I(q, p | b)$ , will be assumed to satisfy are the following:

- i) *Continuous definability in terms of probability.*  $I(q, p | b)$  is a continuous real-valued function of the values that  $\Pr(\cdot | b)$  takes on Boolean combinations of  $p$  and  $q$ .
- ii) *Zero point.* If  $q$  is probabilistically independent of  $p$  (that is, if  $\Pr(q | p \& b) = \Pr(q | b)$ ), then  $I(q, p | b) = 0$ .
- iii) *Additivity of independent information.* If  $p_1$  and  $p_2$  are probabilistically independent of each other, and remain so under conditionalization on  $q$  (that is, if  $\Pr(p_1 \& p_2 | b) = \Pr(p_1 | b) \Pr(p_2 | b)$  and  $\Pr(p_1 \& p_2 | q \& b) = \Pr(p_1 | q \& b) \Pr(p_2 | q \& b)$ ), then  $I(q, p_1 \& p_2 | b) = I(q, p_1 | b) + I(q, p_2 | b)$ .
- iv) *Normalization.* If  $p \& b$  entails  $q$ , then  $I(q, p | b) = -\text{Log}_2(\Pr(q | b))$ .

**Theorem.** The conditions (i)–(iv) entail that  $I(q, p | b) = \text{Log}_2(\Pr(q | p \& b)/\Pr(q | b))$ .

**Proof.**  $I(q, p | b)$  is to be determined by the values  $\Pr(\cdot | b)$  takes on Boolean combinations of  $p$  and  $q$ , and these values will, in turn, be determined by  $\Pr(p | b)$ ,  $\Pr(q | b)$ , and  $\Pr(q | p \& b)$ . There will, therefore, be a continuous function  $F$ , such that

$$I(q, p | b) = F(\Pr(p | b), \Pr(q | b), \Pr(q | p \& b)). \quad (\text{A1})$$

We will assume that this function is well-defined whenever its arguments are all nonzero.

The condition (ii) becomes the condition

$$\text{ii')} \text{ If } z = y \text{ then } F(x, y, z) = 0.$$

Since  $\Pr(p \mid q \ \& \ b) = \Pr(q \mid p \ \& \ b) \Pr(p)/\Pr(q)$ , condition (iii) entails that,

$$\text{iii')} \text{ If } x_3 = x_1 x_2 \text{ and } z_3 = z_1 z_2 / y, \text{ then } F(x_3, y, z_3) = F(x_1, y, z_1) + F(x_2, y, z_2).$$

The first step of the proof is to show that conditions (ii') and (iii') entail that  $F(x, y, z)$  is independent of  $x$ . By (iii'), for all  $x, x', y, z \in (0, 1]$  such that  $x \leq x'$ ,

$$F(x, y, z) = F(x' y, z) + F(x/x', y, y). \tag{A2}$$

By (ii'),  $F(x/x', y, y) = 0$ , and so

$$F(x, y, z) = F(x' y, z). \tag{A3}$$

$F$  is, therefore, independent of its first argument. We can therefore write  $F$  as a function of  $y$  and  $z$ , or, alternatively (and this will turn out to be more convenient), as a function of  $y$  and the new variable  $u = z/y$ :

$$F(x, y, z) = H(y, z / y). \quad (\text{A4})$$

Condition (iii') now becomes

$$H(y, u_1 u_2) = H(y, u_1) + H(y, u_2). \quad (\text{A5})$$

By mathematical induction, we can show that, for any integers  $n, m, m \neq 0$ ,

$$H(y, 2^{n/m}) = n/m H(y, 2). \quad (\text{A6})$$

Therefore, whenever  $u$  is a rational power of 2,

$$H(y, u) = H(y, 2) \text{Log}_2(u). \quad (\text{A7})$$

By continuity, this must hold for all  $u > 0$ , and hence

$$H(y, u) = K(y) \text{Log}_2(u). \quad (\text{A8})$$

for all  $y \in (0,1]$  and  $u > 0$ , where  $K(y) = H(y, 2)$ . The normalization condition (iv) fixes the function  $K(y)$ . If  $p \& b$  entails  $q$ ,  $\Pr(q | p \& b) = 1$ , and so we must have

$$H(y, 1/y) = K(y) \text{Log}_2(1/y) = -\text{Log}_2(y). \quad (\text{A9})$$

This gives us  $K(y) = 1$ , and so

$$H(y, u) = \text{Log}_2(u), \quad (\text{A10})$$

$$F(x, y, z) = \text{Log}_2(z/y). \quad (\text{A11})$$

Thus we have the desired result,

$$I(q, p | b) = \text{Log}_2(\Pr(q | p)/\Pr(q)). \quad (\text{A12})$$

In conclusion, it should be pointed out that the requirement that the function  $F$  be continuous can be replaced by the considerably weaker requirement that it be a measurable function.

The present treatment, however, seems conceptually simpler.



## References

Adams, Marilyn McCord (1987), *William Ockham*, Volume I. Notre Dame, IN: University of Notre Dame Press.

Butts, Robert E., ed. (1968), *William Whewell's Theory of Scientific Method*. Pittsburgh: University of Pittsburgh Press.

Carnap, Rudolf (1962), *Logical Foundations of Probability*, 2nd ed. Chicago: University of Chicago Press.

Chandrasekhar, Subrahmanyan (1995), *Newton's Principia for the Common Reader*. Oxford: Oxford University Press.

Copernicus, Nicolaus ([1543] 1976), *Copernicus: On the Revolutions of the Heavenly Spheres*. Tr. A.M. Duncan. New York: Barnes & Noble Books.

Dijksterhuis, Eduard Jan (1961), *The Mechanization of the World Picture*. Oxford: Oxford University Press.

Forster, Malcolm R. (1988), “Unification, Explanation, and the Composition of Causes in Newtonian Mechanics”, *Studies in History and Philosophy of Science* **19**, 55–101.

Forster, Malcolm, and Elliott Sober (1994), “How to Tell when Simpler, More Unified, or Less *Ad Hoc* Theories will Provide More Accurate Predictions”, *British Journal for the Philosophy of Science* **45**, 1–35.

Friedman, Michael (1974), “Explanation and Scientific Understanding”, *The Journal of Philosophy* **71**, 5–19.

Galilei, Galileo ([1632] 1953), *Dialogue Concerning the Two Chief World Systems*. Tr. Stillman Drake. Berkeley: University of California Press.

Goldstein, Herbert (1980), *Classical Mechanics*. Reading, MA: Addison-Wesley Publishing Company.

Good, Irving John (1950), *Probability and the Weighing of Evidence*. London: Charles Griffin & Company.

——— (1966), “A Derivation of the Probabilistic Explication of Information”, *Journal of the Royal Statistical Society B* **28**, 578–81.

——— (1983), *Good Thinking: The Foundations of Probability and Its Applications*.  
Minneapolis: University of Minnesota Press.

Harper, William (1989), “Consilience and Natural Kind Reasoning (in Newton’s Argument for Universal Gravitation)”, in James R. Brown and Jürgen Mittelstrass (eds.), *An Intimate Relation*, Dordrecht and Boston, Kluwer Academic Publishers, 115–152.

——— (2002a), “Newton’s argument for universal gravitation”, in I. Bernard Cohen and George E. Smith (eds.), *The Cambridge Companion to Newton*, Cambridge: Cambridge University Press, 174–201.

——— (2002b), “Howard Stein on Isaac Newton: Beyond Hypotheses?” in David B. Malament (ed.), *Reading Natural Philosophy: Essays in the History and Philosophy of Science and Mathematics*, Chicago and La Salle, IL: Open Court, 71–112.

Hoskin, Michael, ed. (1999), *The Cambridge Concise History of Astronomy*. Cambridge: Cambridge University Press.

Hoskin, Michael (1999), “Astronomy in Antiquity”, in Hoskin 1999, 18–47.

Hoskin, Michael, and Owen Gingerich (1999), “Medieval Latin astronomy”, in Hoskin 1999, 68–93.

Kepler, Johannes ([1596] 1981), *Mysterium Cosmographicum: The Secret of the Universe*. Tr. A. M. Duncan. New York: Abaris Books.

Kuhn, Thomas S. (1959), *The Copernican Revolution*. New York: Random House.

Milne, Peter (1996). “ $\log[P(h/eb)/P(h/b)]$  is the One True Measure of Confirmation”, *Philosophy of Science* **63**, 21–26.

Myrvold, Wayne C. (1996), “Bayesianism and Diverse Evidence: A Reply to Andrew Wayne”, *Philosophy of Science* **63**, 661–665.

Neugebauer, Otto (1957), *The Exact Sciences in Antiquity*. Providence, Rhode Island: Brown University Press.

Newton, Isaac ([1726] 1999), *The Principia: Mathematical Principles of Natural Philosophy*. Tr. I Bernard Cohen and Anne Whitman. Berkeley: University of California Press.

Niiniluoto, Ilkka, and Raimo Tuomela (1973), *Theoretical Concepts and Hypothetico-Inductive Inference*. Dordrecht and Boston: D. Reidel Publishing Company.

Ptolemy (1984), *Ptolemy's Almagest*. Tr. G. J. Toomer. London: Duckworth.

Rosen, Edward. (1959), *Three Copernican Treatises*, 2nd ed. New York: Dover Publications.

Salmon, Wesley C. (1975), "Confirmation and Relevance", in Grover Maxwell and Robert M. Anderson, Jr. (eds.), *Induction, Probability, and Confirmation. Minnesota Studies in the Philosophy of Science, Volume VI*, Minneapolis: University of Minnesota Press, 3–36.

Shimony, Abner (1970), "Scientific Inference", in R. G. Colodny (ed.), *Mind and Cosmos: Essays in Contemporary Science and Philosophy*, Pittsburgh: Pittsburgh University Press, 79–172. Reprinted in *Search for a Naturalistic World View, Vol. I: Scientific Method and Epistemology*, Cambridge: Cambridge University Press, 1993, 183–273..

Thorburn, W. M. (1918), "The Myth of Occam's Razor", *Mind* **27**, 345–353.

van Fraassen, Bas C. (1989), *Laws and Symmetry*. Oxford: Oxford University Press.

Whewell, William (1847), *The Philosophy of the Inductive Sciences, Founded Upon Their History*, Vol. II, 2nd edition. London: John W. Parker.

Whittaker, Edmund Taylor (1944), *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies*. New York: Dover Publications.

Wingate, Edmond (1653), *Maximes of Reason: or, the Reason of the Common Law of England*. London: R.&W.L.

## Notes

<sup>1</sup> The Copernican and Tychoonic systems are, as far as the relative motions of the planets are concerned, kinematically equivalent. For the purposes of this section, Tycho's and Copernicus' systems can both be regarded as heliocentric, in that they both center the planetary orbits near the sun, and hence both attribute part of the apparent motion of the planets to the relative motion of the earth and sun. If one expects (as did Kepler) that the sun will play a pre-eminent *dynamical* role in the motions of all the planets, then such a consideration will support the Copernican system over the Tychoonic.

<sup>2</sup> The following discussion ignores the variation in latitude of the planets. Also, for Mercury, Ptolemy was compelled to resort to a more complicated arrangement; the simple deferent/epicycle model mentioned here gives at best a first approximation of Mercury's motion.

<sup>3</sup> The Latin formulation given by Galileo (1953, 143), "*frustra fit per plura quod potest fieri per pauciora*," does not seem to be a translation of any dictum of Aristotle. It does, however, occur in Duns Scotus' commentary on Aristotle's *Physics*, as well as a number of other places in the writings of Scotus and of his student, William of Ockham, and, though not originating with Ockham, has come to be associated with him as one formulation of Ockham's razor. For a discussion, see Thorburn 1918; Adams 1987, 156–161. The maxim is invoked by Newton in the third edition of the *Principia* in support of Rule 1. As Thorburn points out, by the seventeenth century the maxim had gained considerable currency,

appearing even in a collection of legal maxims (Wingate 1653, 319, maxim 177). Newton's invocation of it need not be regarded as a reference to any specific source.

<sup>4</sup> Newton, therefore, not only showed that the force on the planets is given very nearly by the inverse-square law, but also showed how to detect small deviations from the inverse square law via an examination of the precession of the orbit. Such a method has since been put to good use.

<sup>5</sup> For a different argument that this quantity ought to be regarded as a measure of the amount of information yielded by one proposition about another, see Good 1966. For a related theorem, see Milne 1996.

<sup>6</sup> As an anonymous referee has pointed out, this notion of unification is related to the notion of “inductive systematization” introduced by Niiniluoto and Tuomela (1973, 11–12). On one of the versions of inductive systematization suggested by Niiniluoto and Tuomela,  $h$  establishes inductive systematization if and only if, for some  $p, q$ ,

a)  $\Pr(q \mid h \ \& \ p) > \Pr(q \mid h)$ .

b)  $\Pr(q \mid p) \leq \Pr(q)$ .

c)  $p \ \& \ h$  does not entail  $q$ .

Conditions (a) and (b) entail that  $U(p, q; h)$  is positive. Niiniluoto and Tuomela do not define a degree of inductive systematization and hence do not demonstrate a relation between such a degree and degree of confirmation.



Some readers may wonder whether the quantity  $U$  is subject to an “irrelevant conjunction” problem; does the mere conjunction of any  $p_1, p_2$  count as a hypothesis that unifies them? Conditional on  $p_1 \& p_2$ ,  $p_1$  and  $p_2$  both have probability one and hence are independent;  $I(p_1, p_2 | p_1 \& p_2 \& b) = 0$ . Therefore, if  $p_1$  and  $p_2$  are independent phenomena,  $U(p_1, p_2; p_1 \& p_2 | b) = 0$ . If  $p_1$  and  $p_2$  are not independent on  $b$ —if they are negatively (or positively) relevant to each other—then conditionalization on their conjunction brings them up (or down) to independence and hence in a sense “unifies” (or “disunifies”) them.

Although this is in some sense “unification,” in most cases the mere conjunction of  $p_1$  and  $p_2$  will not be a very interesting hypothesis, and in these cases this will not be a very interesting form of unification.

<sup>7</sup> This result is essentially the same as equation (5) of Myrvold 1996; the only difference is that we are here dealing with the logarithms of the quantities involved. It is not known to the present author whether an earlier appearance of this way of parsing the degree of confirmation of  $h$  by  $e_1 \& e_2$  exists in the literature.

<sup>8</sup> Recall that, in a misleading bit of terminology, the *likelihood* of  $h$  on evidence  $e$  is  $\Pr(e | h)$ , not  $\Pr(h | e)$ .

<sup>9</sup> The following conversation took place between an itinerant Preacher of Unification and van Fraassen’s Bayesian Peter (see van Fraassen 1989, 166–169).

“You should raise your credence in a hypothesis that unifies the evidence better than its rivals.”

“More than I would anyway?” asked Peter.

“No,” said the Preacher. “Exactly as much as you would anyway.”

<sup>10</sup> It is difficult for a modern reader to take such criticisms seriously. As Galileo plainly saw, such criticisms must be answered if the Copernican hypothesis is to count as a serious candidate for the truth; one who believes that it is a consequence of physical law that, if the Copernican hypothesis were true, “neither clouds nor other flying or thrown objects would ever be seen moving toward the east, since the earth’s motion toward the east would always outrun and overtake them” (Ptolemy 1984, 45), *ought* to regard the Copernican hypothesis as empirically untenable.

<sup>11</sup> I am grateful to Sona Ghosh for raising this point.

<sup>12</sup> I am grateful to an anonymous referee for raising this concern.