On The Evidential Import of Unification

Wayne C. Myrvold
Department of Philosophy
The University of Western Ontario
wmyrvold@uwo.ca

September 10, 2015

Abstract

This paper discusses two senses in which a hypothesis may be said to unify evidence. One is the ability of the hypothesis to increase the mutual information of a set of evidence statements; the other is the ability of the hypothesis to explain commonalities in observed phenomena by positing a common origin for them. On Bayesian updating, it is only mutual information unification that contributes to the incremental support of a hypothesis by the evidence unified. This poses a challenge for the view that explanation is a confirmatory virtue that contributes to such incremental support; its advocates must ground it in some relevant difference between humans and a Bayesian agent.

Keywords: Unification, explanation, confirmation, Bayesianism.

1 Introduction

Myrvold (2003) identified what was described therein as "one interesting sense" in which a theory can unify phenomena. This consists of the ability of the theory to render distinct phenomena informative (or more informative) about each other. Call this *Mutual Information Unification* (MIU). This sense lends itself nicely to a probabilistic explication, and it can be shown that unification in this sense contributes to incremental evidential support of the theory by the phenomena unified.

There is another sense of unification, having to do with hypotheses that posit a common origin for the phenomena in question, be it a common cause or some other type of explanation. Call this *Common Origin Unification* (COU). As emphasized by Lange (2004), the two senses are logically independent; neither is a necessary or a sufficient condition for the other, even though, in a number of interesting cases, they are concomitants of each other.

The question arises as to the respective roles of these two notions of unification in theory confirmation. On a Bayesian analysis, the answer is clear: Mutual Information Unification contributes to incremental evidential support, and there is no scope, within Bayesian updating, for Common Origin Unification to add to the evidential support of the theory (see §4, below).

However, we need not take consideration of a Bayesian agent updating via conditionalization as normative for those of us who are not such agents, and one might still take an explanationist line, that COU, instead of or in addition to MIU, ought to be taken into account in evaluating the bearing of a body of evidence on a hypothesis. If, however, it is rational, or reasonable, or otherwise well and good for us to do what is impossible for a Bayesian agent updating its credences via conditionalization, that is, to regard the explanatory capacity of a hypothesis as something that contributes to its evidential support, then this must be grounded in some relevant difference between us and Bayesian agents. It is incumbent on an explanationist to give an account of what that difference is. No suggestion about how this might go will be offered in the present paper; I leave that to proponents of the explanationist thesis.

In the following, these points are first illustrated by means of a simple example that, despite its artificiality, shares some salient features with cases of actual scientific interest. Next, in §3, are presented the probabilisitic measures of MIU introduced by Myrvold (2003), and in §4 their impact on evidential support is exhibited. In §5 the question is raised whether there is still a role for Common Origin Unification to play in hypothesis assessment, in assessing priors rather than in assessing incremental evidential support (the answer is no). Finally, in §6 it is shown how Reichenbachian common causes fit into the schema of Mutual Information Unification.

2 Two kinds of unification

Consider the following toy example, of no use except for introducing the issues at hand, though it does share some salient features with a multitude of real-world cases of genuine scientific interest. You are about to be presented with two data streams, S_1 and S_2, each of which will be sequences of ten Heads or Tails. You know that they have been produced by coin flipping, but you aren't sure of exactly what the procedure used was, or whether the coin or coins involved are fair.

Suppose that you have nonzero credences in both of the following hypotheses:

H_1: A fair coin was flipped ten times, and the results of this series of coin flips are reported in both data streams.

H_2: Two fair coins were flipped ten times each, and each data stream reports the results of one of these series of coin flips.

I invite you to consider the effect of the evidence on these two hypotheses. That evidence consists of specification of the two data streams:

S_1: HHHTTHTHHT

S_2: HHHTTHTHHT

Let E_1 be the proposition that S_1 is the string given above, and E_2 the corresponding proposition about S_2.

Now, if you have nonnegligible prior credence that the strings might have been produced by radically unfair coins, E_1 and E_2 might boost your confidence in the fairness of the coins, and hence conditionalizing on each of E_1 and E_2, separately, might boost your credence in both H_1 and H_2. But, when taken together, E_1 and E_2 strongly favor H_1 over H_2.

There are two features of this example that I would like to draw your attention to. The first feature is that H_1, if true, renders E_1 informative about what data stream S_2 will be. Conditional on H_1, knowing E_1 permits one to anticipate the truth of E_2. That is, H_1 exhibits Mutual Information Unification (*MIU*) with respect to the evidence set $\{E_1, E_2\}$. A hypothesis has this property, with respect to a set of evidential propositions, if conditionalizing on that hypotheses increases the mutual informativeness of the set. Obviously, this is the sort of thing that comes in degrees; probabilistic measures of this sort of unification will be introduced below.

The second feature is that H_1 posits a common origin of the two data streams, and thus is ripe to be the subject of what Janssen (2002) has called a *COI story*, for *Common Origin Inference*. In addition to MIU, H_1 also exhibits Common Origin Unification, *COU*.

The two concepts are of a manifestly very different sort. One belongs to a cluster of concepts involving information, states of knowledge, and the like; the other is related to concepts of cause and explanation.[1] As already mentioned, they are logically independent. A hypothesis can posit a common origin for two (or more) evidential propositions without making them mutually informative about each other, as the propositions could be about independent aspects of their posited common origin, so we can have *COU* without *MIU*. Furthermore, once two or more evidential propositions are known, that is, have been absorbed into one's background knowledge, they are no longer informative about each other, though any common origin they might have remains, and again we have *COU* without *MIU*. One can also trivially construct hypotheses that exhibit *MIU* without *COU*. With respect to our toy example, consider the hypothesis,

H_3: Two fair coins were flipped ten times each, each data stream reports the results of one of these series of coin flips, and the results of each series of flips just happened to be the same.

This hypothesis, if true, also renders one data stream informative about the other. Of course, prior to the evidence, one would expect credence in H_3 to be low, lower than credence in H_2 by a factor of $1,024$.

Though artificial, our toy example has a multitude of parallels in actual science. One is the case of heliocentric v geocentric world systems, discussed by Janssen (2002) and Myrvold (2003). The analog of H_1 is what is called h_C by Myrvold (2003), that is, the heliocentric hypothesis that all planets have circular or nearly circular orbits centred at or near the sun, and the analog of H_2 is the bare-bones geocentric hypothesis h_P, which posits that, for each planet, there is a deferent circle centered near the earth, and that the planet travels on an epicycle whose center travels on the deferent, with no assumption made about any connections between the motions of different planets or between planetary motions and the motion of the sun. The analog of H_3 is the geocentric hypothesis with the added condition that Myr-

[1] As Glymour (2015) has argued, it would be a grave mistake to take *MIU* or any similar notion as an explication of explanatory unification.

vold calls the *sun-planet parallelism condition*; this is the hypothesis that Myrvold calls h_{SP}, or the *strengthened Ptolemaic* hypothesis.

Another analog, with a wider body of evidence, can be found in Perrin's argument for the existence of atoms. Perrin (1913, §119; 1916, §120) adduces 13 distinct phenomena that, on the atomic hypothesis, count as measurements of Avogadro's number. The analog of H_1 is that atoms exist, and hence there is a common origin explanation of the agreement of these measurements; the analog of H_2 would be the hypothesis that matter is continuously divisible, and the analog of H_3 would be the hypothesis that adds to H_2 the stipulation that Perrin's 13 phenomena yield values that just happen to agree within experimental error, even though they are not agreeing measurements of any physically meaningful parameter. Clearly, one could adduce a multitude of examples, from many domains of science, with the same sort of structure.

3 Probabilistic measures of unification

Consider a Bayesian agent whose credences are given by a probability function Cr. We define the mutual information of a pair of propositions, $\{p_1, p_2\}$, relative to background b, by[2]

$$I(p_1, p_2|b) = \log_2 \left(\frac{Cr(p_2|p_1 b)}{Cr(p_2|b)} \right) = \log_2 \left(\frac{Cr(p_1\, p_2|b)}{Cr(p_1|b) Cr(p_2|b)} \right). \quad (1)$$

If p_1 and p_2 are probabilistically independent on b, then $I(p_1, p_2|b)$ is zero; it is positive if conditionalizing on one boosts credence in the other, negative, if conditionalizing on one lowers credence in the other.

For a larger set, $\mathbf{p} = \{p_1, p_2, \ldots, p_n\}$, we add up the information yielded by p_1 about p_2, the information yielded by $p_1 p_2$ about p_3, and so on, up to the information about p_n yielded by the conjunction of

[2] A note on notation. We will use concatenation for conjunction, and the overbar $\bar{p}$ for the negation of p. We use boldface letters to denote sets of propositions. Note that these are *sets* and are not replaceable by a single proposition that is their conjunction. Thus, $\{p_1, p_2\}$ is not the same set as $\{p_1 p_2, \mathbb{T}\}$. where $\mathbb{T}$ is the logically true proposition, though the conjunction of their members is the same. This matters because we will be concerned with the mutual informativeness of members of a set of propositions; p_1 and p_2 may be mutually informative though the logically true proposition is not informative about their conjunction or anything else.

all the others.

$$\begin{aligned} I(p_1, \ldots, p_n|b) &= I(p_1, p_2|b) + I(p_1 p_2, p_3|b) + \ldots + I(p_1 \ldots p_{n-1}, p_n|b) \\ &= \sum_{k=1}^{n-1} I\left(\bigwedge_{i=1}^{k} p_i, p_{k+1}|b\right). \end{aligned} \tag{2}$$

Although the form of (2) does not make this obvious, this quantity is independent of the order in which the elements of the set $\mathbf{p}$ are taken, and we have,

$$\begin{aligned} I(p_1, \ldots, p_n|b) &= \log_2\left(\frac{Cr(p_1\, p_2 \ldots\, p_n|b)}{Cr(p_1|b)Cr(p_2|b)\ldots Cr(p_n|b)}\right) \\ &= \log_2\left(\frac{Cr(\bigwedge_{i=1}^{n} p_i|b)}{\Pi_{i=1}^{n} Cr(p_i|b)}\right). \end{aligned} \tag{3}$$

This is the logarithm of the quantity called a measure of similarity by Myrvold (1996) and taken by Shogenji (1999) as a measure of coherence of a set of propositions; $I(\mathbf{e}|hb)$ is the logarithm of the quantity that McGrew (2003) took as a measure of the degree of consilience of a hypothesis h with respect to a body of evidence $\mathbf{e}$. With a slight abuse of notation, we will write $I(\mathbf{p}|b)$ for $I(p_1, \ldots, p_n|b)$. We will also drop, as irrelevant, the base of the logarithm, since changing base is only a matter of a constant multiplicative factor.

We will say that a hypothesis h *MIUnifies* a set $\mathbf{e} = \{e_1, \ldots, e_n\}$, relative to background b, if and only if

$$I(\mathbf{e}|h\, b) > I(\mathbf{e}|b). \tag{4}$$

This suggests a way to measure the degree to which a hypothesis MIUnifies a set of evidential propositions.[3]

$$MIU_1(\mathbf{e}; h|b) = I(\mathbf{e}|h\, b) - I(\mathbf{e}|b). \tag{5}$$

We might also be interested in whether a hypothesis does a better job of unifying a set of propositions than its negation. Define

$$\begin{aligned} MIU_2(\mathbf{e}; h|b) &= MIU_1(\mathbf{e}; h|b) - MIU_1(\mathbf{e}; \bar{h}|b) \\ &= I(\mathbf{e}|h\, b) - I(\mathbf{e}|\bar{h}\, b). \end{aligned} \tag{6}$$

[3] This quantity is the logarithm of a quantity that was refereed to as an "interaction term" in Myrvold (1996), and is called *focussed correlation* in Wheeler (2009), Schlosshauer and Wheeler (2011), and Wheeler and Scheines (2013). What we are calling MIU_1 was called U (for unification) in Myrvold (2003). MIU_2 was discussed therein, though not given its own name.

The two are not ordinally equivalent, and, indeed, need not agree as to sign. Suppose a hypothesis h unifies a body of evidence, relative to background b. That is, suppose the evidence is more mutually informative conditional on hb than on b alone. Then $MIU_1(\mathbf{e}; h|b)$ is positive. But whether or not $MIU_2(\mathbf{e}|b)$ is negative or positive depends on whether or not $\bar{h}$ unifies the evidence more. If $I(\mathbf{e}; \bar{h}|b)$ is greater than $I(\mathbf{e}; h|b)$, then, even if $MIU_1(\mathbf{e}; h|b)$ is positive, $MIU_2(\mathbf{e}; h|b)$ is negative. In fact, all four combinations of signs of MIU_1 and MIU_2 are possible, though it is easy to show that, unless e_1 and e_2 are, when taken individually, oppositely relevant to h (that is, unless one of them is positively relevant and the other negatively relevant), if $MIU_1(e_1, e_2; h|b)$ is positive, $MIU_2(e_1, e_2; h|b)$ is also positive. See Appendix for details.

Readers are asked to kindly refrain from engaging in a battle of the intuitions over which of these is the One True Measure of degree of unification. They are simply measuring different things, and if you have intuitions that are incompatible with properties that one or another of these quantities possesses, then your intuitions are about some other concept.[4]

4 The evidential value of unification

To some readers, it might seem obvious that what counts when it comes to confirmation is Common Origin Unification, with Mutual Information Unification being a poor cousin that hardly merits the illustrious family surname. This view is expressed by Lange (2004), who writes,

> the examples I have given suggest that insofar as theories that unify in the stronger,[5] ontological-explanatory sense derive greater support in virtue of the unification they achieve, they do so not solely in virtue of their achieving unification in the weaker, creating-mutual-positive relevance sense. The stronger sense of unification is epistemically significant. In the case of the light-quantum hypothesis, h_C and h_L both supply unity in the weaker sense, but Einstein took h_L to receive greater support from the

[4] And if your intuitions find it repugnant to use the word "unification" in connection with either of these, then feel free to use a different word.

[5] This is a slip; the two senses are, as Lange emphasizes, logically independent.

phenomena than h_C by virtue of h_L's unifying those phenomena in an ontological-explanatory sense (p. 212).

Here h_L is Einstein's light quantum hypothesis, and h_C is the hypothesis that h_L is false but nevertheless, by sheer coincidence, light behaves as if it were quantized. According to Lange, h_L receives greater support from the phenomena unified than does h_C.

It is not entirely clear whether incremental or absolute support is meant, where incremental support has to do with an increase in credibility lent to a hypothesis by the evidence, and absolute support with the credibility of the hypothesis, taking all known considerations into account. If absolute, this suggests that the case of h_C is analogous to that of our toy example's H_3, which is accorded a low prior because it posits an improbable coincidence. One the other hand, if the claim is to be a counterexample to anything in Myrvold (2003), incremental support must be what is meant. Let us therefore consider the position that, when it comes to incremental support, it is COUnification, not MIUnification, that counts.

A Bayesian analysis renders the opposite verdict: when it comes to incremental support of a hypothesis, it is MIUnification, rather than COUnification, that matters.

One popular measure of the degree to which an evidential proposition e lends incremental confirmation to a hypothesis h, relative to background b, is the ratio of posterior probability of h to its prior probability. This is, of course, ordinally equivalent to its logarithm. Let us define

$$R(h; e|b) = \log\left(\frac{Cr(h|e\,b)}{Cr(h|b)}\right). \tag{7}$$

Another is the ratio of the posterior odds of h to its prior odds, or, equivalently, the logarithm of this. Define

$$W(h; e|b) = \log\left(\frac{Cr(h|eb)/Cr(\bar{h}|eb)}{Cr(h|b)/Cr(\bar{h}|b)}\right) = \log\left(\frac{Cr(e|h)}{Cr(e|\bar{h})}\right). \tag{8}$$

As Myrvold (2003) pointed out, on either way of measuring incremental confirmation, we have a contribution of unification to confirmation.[6] The incremental support, as measured by R, of h by $\mathbf{e}$ can

[6] Equations (9) and (10) correspond to equations (12) and (13) of Myrvold (2003). In that paper, the displayed equations are for two items of evidence, with the generalization to larger evidence-sets stated in the prose beneath.

be decomposed into a sum of increments due to the individual members of $\mathbf{e}$, plus an additional term that is the degree of MIUnifcation (positive or negative) of $\mathbf{e}$ by h, as measured by MIU_1.

$$R(h; \bigwedge_{i=1}^{n} e_i|b) = \sum_{i=1}^{n} R(h; e_i|b) + MIU_1(\mathbf{e}; h|b). \qquad (9)$$

The result for W takes the same form, with MIU_2 in place of MIU_1.

$$W(h; \bigwedge_{i=1}^{n} e_i|b) = \sum_{i=1}^{n} W(h; e_i|b) + MIU_2(\mathbf{e}; h|b). \qquad (10)$$

These relations can be readily verified by the reader.

It would be incorrect to gloss these results as saying that hypotheses that are more unifying receive more confirmation; the MIU-term is not the only contribution to the increment of confirmation. Although it would not be incorrect to say that *ceteris paribus*, a hypothesis that achieves a higher degree of MIUnification of the evidence is accorded greater incremental support, this is strictly weaker than what is conveyed in equations (9) and (10), and there is no advantage in making the *ceteris paribus* claim when it is a trivial matter to say how things stand when all else is *not* equal.

Imagine, now, a Bayesian agent that had numerical credences, which it[7] updated by conditionalizing on new items of evidence. Then, depending on how we measured degree of incremental confirmation, the confirmational boost accorded to h by a set $\mathbf{e}$ of evidential propositions would be given by either (9) or (10). In each case the additional confirmational boost, beyond that attributable to the items of evidence taken singly, is given by the MIUnification term.

Applied to our toy example: The fact that H_1 and H_3 make E_1 and E_2 informative about each other is reflected in the likelihoods, $Cr(E_1E_2|H_1)$ and $Cr(E_1E_2|H_3)$, which are higher than $Cr(E_1E_2|H_2)$ by a factor of $1,024$. Thus, relative to H_2, credence in H_1 and H_3 is boosted:

$$\frac{Cr(H_1|E_1E_2)}{Cr(H_1)} = \frac{Cr(H_3|E_1E_2)}{Cr(H_3)} = 2^{10}\,\frac{Cr(H_2|E_1E_2)}{Cr(H_2)}. \qquad (11)$$

It doesn't follow, of course, that H_3 gets final credence comparable to that of H_1. Since H_3 posits an improbable coincidence, it is accorded

[7] I say "it," because a being with precise numerical credences would be far from human.

a lower prior probability, lower than that of H_2 by a factor of $1,024$; the additional confirmational boost it receives is just enough to bring it up to posterior credence equal to that of H_2 (which, of course, must be the case, since, given the evidence, H_3 is true if and only H_2 is).

There is a close parallel between this case and the case of geocentric v heliocentric world systems, and also the case of the light quantum, considered by Lange.

In the case of planetary motion, on both the heliocentric hypothesis and the strengthened Ptolemaic hypothesis, features of one planet's orbit are informative about features of others (see Janssen 2002 and Myrvold 2003 for discussion). In the case of the heliocentric hypothesis, H_C, these have a common origin in the motion of our vantage point as observers on earth; for H_{SP}, they are the consequence of the posited sun-planet parallelism. Against a background that includes little or no information about observed planetary motions, both of these get a confirmational boost from the celestial phenomena, due to the MIU-component of incremental confirmation. It doesn't follow that they end up with equal posterior credence. Arguably, H_{SP}, on that background, should be accorded markedly lower prior credence than H_P, as it posits a relation that H_P by itself would not lead one to anticipate. H_C and H_{SP} get the same incremental confirmation on the evidence. Therefore, posterior credence in H_{SP} will be comparable to posterior credence in H_C only if prior credence in H_C is markedly lower than prior credence in H_P.

Something similar can be said in regards to Lange's case of the light quantum hypothesis. Let us grant that the light quantum hypothesis plays a unificatory role. Lange asserts that Einstein took the observed phenomena to lend greater support to the light quantum hypothesis than the hypothesis that, by sheer coincidence, all observable phenomena are as if the light quantum hypothesis is true. In order for this assertion to be relevant to the issue at hand, this must mean that the phenomena lend greater *incremental* support to the light-quantum hypothesis than to the coincidence hypothesis. On the historical point, it seems that the documentary evidence is silent; there is no textual evidence that Einstein even considered Lange's h_C. One would suspect that he would have regarded it as so implausible as to be dismissed out of hand. But this would mean according it a low prior, which is consistent with the Bayesian account of the virtue of unification. There is no reason to suspect that Einstein's failure to seriously consider the coincidence hypothesis ought to be ascribed to his taking the phenomena to lend greater incremental support to theories that achieve MIUnification via positing a common origin than to theories that merely posit unexplained connections between phenomena.

Bayesian updating leaves no room for an additional confirmatory boost to be attached to hypotheses with greater explanatory power; the contribution to incremental support comes via the MIUnification term. We, however, are not Bayesian agents, and perhaps we should not take consideration of a Bayesian agent that updated by conditionalization to be normative for the judgments of human scientists about incremental support.

If this is right—if one accepts that for the credences of a Bayesian agent it is MIU and not COU that is relevant to incremental support, but holds that we humans ought to judge hypotheses that posit a common origin to receive greater support from the phenomena COUnified by the hypothesis—then this difference must be grounded in some difference between us and Bayesian agents. We are certainly different from Bayesian agents in a number of ways. We do not have precise numerical degrees of belief; our judgments about how likely or unlikely a hypothesis is tend to be vague. Moreover, as an abundance of empirical evidence shows, routinely our qualitative judgments of the relative credibility of various propositions are not even compatible with the existence of numerical credences satisfying the axioms of probability, and our changes in credences are often not in accord with Bayesian conditionalization. There is also experimental evidence that, it is claimed, "strongly suggest that judgments of the explanatory goodness of hypotheses are of crucial importance in updating ... in a way that is incompatible with the Bayesian doctrine of Strict Conditionalization" (Douven and Schupbach, 2015, p. 309).[8]

The usual understanding of facts of this sort is that they are due to cognitive limitations, and that some of them can be understood as resulting from usually reliable heuristics, of the sort that any agent with limited cognitive capacities would be well-advised to employ, as an alternative to spending excessive time on cogitation. Moreover, deployment of such heuristics can be regarded as rational, from a

[8]It would not be surprising if this turned out to be a robust psychological phenomenon. Though the work reported in Douven and Schupbach (2015) is an interesting first step towards this, one that will hopefully stimulate further research, further experiments would be required to make a case for the thesis that explanatory considerations play an important role in updating beliefs.

decision-theoretic point of view, when the cost of deliberation is taken into account—an instance of what I. J. Good (1971, 1976) called "Type II Rationality."

If it is not an error for us cognitively limited humans to take Common Origin Unification to be what counts in regards to incremental support of a hypothesis by evidence, then this must be due to some relevant difference between us and a Bayesian agent that updates by conditionalization. It is hard for me to see how such an account would go. But I do think that it is incumbent on those who hold an explanationist view, according to which it is explanatory capacity, rather than (or in addition to) mutual information unification, that is relevant in updating credences, to provide such an account.

5 A prior preference for unifying hypotheses?

We have considered cases (in the toy example, H_1 and H_3, in the case of planetary motion, H_C and H_{SP}, and in the light quantum case, h_L and h_C), in which each of a pair of hypothesis possess the same ability to render items of evidence informationally relevant to one another, but do so in different ways. In each of these cases one does it by virtue of positing a common origin for *prima facie* unrelated phenomena, the other, by brute fiat, in positing an unexplained correlation between the phenomena. In each of these cases, the hypothesis that involves a common origin is, arguably, less implausible than the one that posits brute coincidence.

One might be tempted to generalize, positing, that, whenever we have *MIU* without *COU*, there will be a corresponding hypothesis that achieves precisely the same MIUnification via COUnification, and we should accord much less prior credence to the hypothesis that exhibits *MIU* without *COU* than to the one that achieves it via *COU*. This would mean that there is a role for *COU*, not in incremental confirmation, but in setting priors.

Anything so sweeping would be a mistake, I think. There are patterns in the world of all sorts, some due to some sort of common origin, some not, and it is easy to cook up examples of hypotheses with *MIU* but not *COU*. Nature is unified in some respects, disunified in others, and it is the task of science to find out in *which* respects nature is unified.

Perhaps, then, the generalization should be that, when we *do* have a pair of hypotheses that both induce the same informational relevance relations among a body of phenomena, one doing it via COUnification and the other by brute fiat, we should attach higher prior credence to the COUnifying hypothesis. One should be cautious, however, about overgeneralizing. When we have a case of two hypotheses h_1 and h_2 of roughly equal prior credibility, and create a third h_3 by tacking on to h_2 some conjunct with low prior plausibility, then, indeed, in such a case, we should place lower credence in h_3 than in h_1. But not all cases will be like that, and a COUnifying hypothesis might be deemed implausible on other grounds. Take, for example, Ptolemy's attitude towards heliocentric hypotheses. Since Ptolemy recognized that in the observed phenomena there was a connection between the apparent motion of the sun and that of the other planets, he was in a position to appreciate the COUnifying power of geocentrism. But, since he accepted Aristotelian physics for terrestrial phenomena, he thought that terrestrial phenomena ruled out a diurnal rotation of the earth (see Ptolemy 1984, Bk. I, §7); for him, it was reasonable to place low credence in heliocentric theories that posited such a rotation.

One can exhibit plenty of hypothesis pairs in which the less unifying, less explanatory hypothesis has less prior credibility, because the less explanatory hypothesis posits an implausible coincidence. But the emphasis should be on the credibility-diminishing role of coincidence, rather than any prior conviction that nature is unified. What H_3, the strengthened Ptolemaic hypothesis, and Lange's h_C have in common is that, in each case, we have a hypothesis to which is tacked on some additional condition that one would not expect to hold, in the absence of evidence that it does, and hence we have a hypothesis that ought to be accorded low prior credence. Rather than a sweeping preference for COUnification, I suggest that the methodological adage that underwrites low prior credence in such hypotheses is:

> Place little prior credence in things you take to be improbable.

This is, I hope, unobjectionable! It is, of course, utterly empty, but I am skeptical that anything stronger could be defended as a maxim of more than *very* limited scope.

It would be a mistake to raise this bland but unobjectionable maxim into a global rejection of hypotheses that posit coincidences. Improbable things do happen, after all. Moreover, in some cases it is reasonable to accept hypotheses that posit an improbable coincidence. The evidence available to you in the toy example strongly suggests a common cause. But, if you were to obtain strong evidence that the two data streams were the results of independent tosses of two fair coins, then it would be reasonable to accord high credence to H_3. Similarly, Ptolemy propounded a geocentric system with an unexplained sun-planet parallelism, because he thought he had strong evidence to rule out hypotheses that involved a moving earth.

The case of Einstein and the light-quantum hypothesis is interesting, and complex. When Einstein proposed the light-quantum hypothesis in 1905, the suggestion was rather tentative. What he concluded at the time was that

> Monochromatic radiation of low density (within the range of validity of Wien's radiation formula) behaves thermodynamically as if it consisted of mutually independent energy quanta of size $R\beta\nu/N$ (Einstein 1989, p. 97, from Einstein 1905, p. 143).

For this restricted conclusion Einstein gave a very strong argument (see Norton 2006). His next step is to propose an investigation: "it seems reasonable to investigate whether the laws of generation and conversion of light are also so constituted as if light consisted of such energy quanta" (Einstein 1989, p. 143–144, from Einstein 1905, p. 98). The next sections, on Stokes's Rule, on the photoelectric effect, and on ionization of gases, argue that the conception is consistent with the observed phenomena, but make it clear that further experimental work is needed. Pais (1982, p. 377) rightly calls Einstein's step a bold one, but it is boldness tempered by caution.

Einstein was not alone in his hesitation about the light quantum hypothesis, and other physicists were even more sceptical (see Hendry 1980 for some of the history). It was not until a little over a decade after his first proposal of the light-quantum hypothesis that Einstein was willing to declare that the basic principles of the light-quantum theory were "rather certainly proven" and called for, as "almost unavoidable," the development of a proper quantum theory of radiation (Einstein 1997, p. 232, from Einstein 1916). As Kao (2015) argues, a case can be made for thinking of the accumulating evidential support for the quantum hypothesis during the first decades of the twentieth century in terms of unification in the mutual information sense: since the quantum hypothesis turns *prima facie* independent phenomena into agreeing measurements of Planck's constant, it renders these phenomena informative about each other, much as the atomic hypothesis turns disparate phenomena into agreeing measurements of Avogadro's number.

6 Unification and Reichenbachian Common Causes

Among unifying hypotheses are those that posit a Reichenbachian common cause to explain some observed statistical correlation (Reichenbach, 1956, §19). This type of hypothesis fits well within the schema of the Bayesian account of unification, but, since this might not be obvious, it is worth showing how it fits.

Consider two sequences of propositions, $\{A_i, i = 1, \ldots, n\}$, and $\{B_i, i = 1 \ldots, n\}$. Let $f(A)$ be the relative frequency of true instances of the A_i, that is, the number of values of i for which A_i is true, divided by n, and similarly for $f(B)$, and $f(AB)$. A statistically significant difference between $f(AB)$ and the product $f(A)f(B)$ is thought to call for explanation. A *Reichenbachian Common Cause* of an observed correlation between A and B is a third sequence C_i that screens off their correlation. That is,

$$\begin{aligned} Pr(A_i B_i|C_i) &= Pr(A_i|C_i)Pr(B_i|C_i); \\ Pr(A_i B_i|\bar{C}_i) &= Pr(A_i|\bar{C}_i)Pr(\bar{B}_i|\bar{C}_i). \end{aligned} \tag{12}$$

On the assumption that distinct A_is are independent and identically distributed, and that the same holds for $\{B_i\}$ and $\{A_i B_i\}$, we can take the C_is to be independent and identically distributed also.

A hypothesis that there is a common cause of this sort can be characterized by six parameters:

$$\begin{aligned} &p = Pr(C_i), \\ &a_1 = Pr(A_i|C_i), \qquad a_0 = Pr(A_i|\bar{C}_i), \\ &b_1 = Pr(B_i|C_i), \qquad b_0 = Pr(B_i|\bar{C}_i). \end{aligned} \tag{13}$$

On the supposition of such a hypothesis, we have

$$Pr(A_i|H_{cc}) = pa_1 + (1-p)a_0, \quad Pr(B_i|H_{cc}) = pb_1 + (1-p)b_0, \tag{14}$$

and

$$\begin{aligned} Cov(A_i, B_i|H_{cc}) &= Pr(A_iB_i|H_{cc}) - Pr(A_i|H_{cc})Pr(B_i|H_{cc}) \\ &= p(1-p)(a_1 - a_0)(b_1 - b_0). \end{aligned} \tag{15}$$

Thus, conditional on the hypothesis, the A_is are positively correlated with the B_is if $a_1 - a_0$ and $b_1 - b_0$ have the same sign; negatively correlated if they have opposite sign, and uncorrelated if the C_is are irrelevant to either the A_is or the B_is, that is, if $a_1 = a_0$ or $b_1 = b_0$.

We consider a family of such hypotheses, characterized by varying values of the parameters listed above in (13). Let E_1 be a proposition expressing which of the A_is are true, and which are false. For example, in our toy example, A_i could be the proposition that the ith element of S_1 is *Heads*, and E_1 would be

$$A_1 A_2 A_3 \bar{A}_4 \bar{A}_5 A_6 \bar{A}_7 A_8 A_9 \bar{A}_{10}.$$

Let E_2 be the evidence statement specifying the B-sequence. We inquire into the degree of support lent to common-cause hypotheses, with various values of the parameters, by the pair $\{E_1 E_2\}$.

Given E_1, E_2, let $n(A)$ be the number of values of i for which A_i is true, $n(B)$ the number for which B_i is true, $n(AB)$ the number of values of i for which A_iB_i is true, and similarly for $n(A\bar{B})$, and the other combinations. Let H_{cc} be some Reichenbachian common-cause hypothesis. We have, from (9),

$$\begin{aligned} R(H_{cc}; E_1E_2) = R(H_{cc}; E_1) + R(H_{cc}|E_2) \\ + MIU_1(\{E_1, E_2\}; H_{cc}). \end{aligned} \tag{16}$$

Since we're interested in comparing degrees of support for different hypotheses on a fixed body of evidence, it is useful to compare log-likelihoods, as, for two different hypotheses, the differences between their R-values will be the same as the differences between the respective log-likelihoods. The log-likelihoods can be partitioned in a manner parallel to our partitioning of R:

$$\begin{aligned} \log Pr(E_1E_2|H_{cc}) = \log Pr(E_1|H_{cc}) + \log Pr(E_2|H_{cc}) \\ + I(E_1, E_2|H_{cc}). \end{aligned} \tag{17}$$

The first two terms of this are

$$\log Pr(E_1|H_{cc}) = n(A) \log Pr(A_i|H_{cc}) + n(\bar{A}) \log Pr(\bar{A}_i|H_{cc});$$

$$\log Pr(E_2|H_{cc}) = n(B) \log Pr(B_i|H_{cc}) + n(\bar{B}) \log Pr(\bar{B}_i|H_{cc}). \tag{18}$$

These are maximized by a hypothesis H_{cc} that has $Pr(A_i|H_{cc}) = n(A)/n$ and $Pr(B_i|H_{cc}) = n(B)/n$. That is, these terms are largest for hypotheses that posit probabilities for the A_is and B_is that are equal to the observed relative frequencies.

The mutual information of E_1 and E_2, conditional on a hypothesis H_{cc}, is

$$I(E_1, E_2|H_{cc}) = n(A, B)I(A_i, B_i|H_{cc}) + n(A, \bar{B})I(A_i, \bar{B}_i|H_{cc}) + n(\bar{A}, B)I(\bar{A}_i, B_i|H_{cc}) + n(\bar{A}, \bar{B})I(\bar{A}_i, \bar{B}_i|H_{cc}). \quad (19)$$

Once $Pr(A_i|H_{cc})$ and $Pr(B_i|H_{cc})$ are fixed, this is maximized by taking

$$Pr(A_iB_i) = n(AB)/n. \quad (20)$$

Thus, in the expression (17) for the log-likelihood, we see that the first two terms reward hypotheses whose probabilities for A_i and B_i are close to the observed relative frequencies of these, and the last term, which corresponds to unification in the Mutual Information sense, rewards hypotheses with theoretical correlations close to the observed statistical correlations. What goes for log-likelihoods goes also for the evidential support R. Thus, when there is a difference between $f(AB)$ and $f(A)f(B)$, a common-cause hypothesis on which this difference is expected, by virtue of appropriate values of the parameters, counts as a MIUnifyng hypothesis, and thereby achieves greater support.

This might seem paradoxical. A common cause screens off the correlations between the A_is and B_is; how can it be that, at the same time, there is a confirmational boost associated with rendering them informative about each other?

The answer to this is: the hypothesized common causes C_i screen off the correlations, but the hypothesis *that there is some such common cause* renders the A_is and B_is informative about each other, relative to a hypothesis that posits no correlation. There's no tension between the Bayesian account of the virtue of unification and Reichenbach's thesis that common causes screen off correlations.

The key is that, for appropriate values of the parameters, the hypothesis H_{cc} affords MIUnification to the evidence statements E_1, E_2, even though, in individual cases, the supposition C_i does not render A_i informative about B_i. This does not prevent C_i from being regarded as a *common origin* of A_i and B_i. To take an example used by Lange in §3 of his paper, suppose that we take the clinical evidence to establish that some disease C is a cause of symptoms A and B. Then, if we observe A and B in some patient, this will raise our credence that C also occurs in that patient, even if the symptoms A and B are independent, conditional on C. In such a case, the support provided by the symptoms A and B to the hypothesis that the patient has disease C is just the sum of the supports given to the hypothesis by the individual items by themselves.

Lange raises the question of whether we should place more credence in a hypothesis that posits a single disease than in one that posits two independent origins of the symptoms A and B. Suppose there are two other diseases D_1 and D_2, such that A but not B is a symptom of D_1, and B but not A is a symptom of D_2, and suppose further that the chance of a patient with D_1 exhibiting symptom A is the same as that of a patient with C, and that the chance of a patient with D_1 exhibiting symptom B is the same as that of a patient with C. Then, upon observation of both symptoms, the confirmational boost afforded to the hypothesis that the patient has C is the same as the boost afforded to the hypothesis that the patient has both D_1 and D_2, and the issue comes down to priors. Is the joint occurrence of D_1 and D_2 much rarer than the occurrence of C? If the answer is yes—as would be the case if the three diseases are equally rare, and D_1 and D_2 uncorrelated—then we should place more credence in the hypothesis that the patient has C. If not—if the disease C is so rare, and D_1 and D_2 so common that more patients contract both D_1 and D_2 than C—then our credences should favor the two-disease hypothesis. It would seem to be clearly a mistake to take the C-hypothesis to receive greater support from the evidence merely on a preference for common origin explanations.

7 Conclusion

Mutual Information Unification is not the same as common origin explanation, and is neither a necessary nor sufficient condition for a hypothesis to play an explanatory role. Nevertheless, in a host of interesting cases, MIUnification is a concomitant of common origin explanation. Moreover, when a hypothesis that renders an otherwise puzzling coincidence comprehensible by providing a common origin explanation *does* receive an incremental confirmational boost from a body of evidence, beyond that provided by the individual items of evidence, that boost stems from MIUnification.

So, at least, is the verdict delivered by a Bayesian analysis; there is no room in Bayesian conditionalization for an extra confirmatory boost that is due to Common Origin Unification. A proponent of an explanationist thesis, to the effect that we ought to take hypotheses that involve common origin explanations to receive greater incremental support than hypotheses that achieve the same degree of Mutual Information Unification without explanation, should be in a position to explain why what is impossible for a Bayesian agent is rational for us. I know of no argument that this cannot be done, but, so far, it has (as far as I know) not been done.

8 Appendix

Given a probability function Pr, and propositions h, e_1, e_2, define,

$$U_1 = \frac{Pr(e_1e_2|h)}{Pr(e_1|h)Pr(e_2|h)}\frac{Pr(e_1)Pr(e_2)}{Pr(e_1e_2)}; \tag{21}$$

$$U_2 = \frac{Pr(e_1e_2|\bar{h})}{Pr(e_1|\bar{h})Pr(e_2|\bar{h})}\frac{Pr(e_1)Pr(e_2)}{Pr(e_1e_2)}. \tag{22}$$

Then we have

$$MIU_1(e_1, e_2; h) = \log U_1; \tag{23}$$

$$MIU_2(e_1, e_2; h) = \log\left(U_1/U_2\right). \tag{24}$$

Thus, $MIU_1(e_1, e_2; h)$ is positive iff $U_1 > 1$, negative iff $U_1 < 1$, and zero iff $U_1 = 1$, and $MIU_2(e_1, e_2; h)$ is positive iff $U_1 > U_2$, negative iff $U_1 < U_2$, and zero iff $U_1 = U_2$.

We want to show that each of the following four alternatives can be realized by some probability function.

1. $MIU_1 > 0$ and $MIU_2 > 0$; that is, $U_1 > 1$ and $U_1 > U_2$.
2. $MIU_1 > 0$ and $MIU_2 < 0$; that is, $1 < U_1 < U_2$.
3. $MIU_1 < 0$ and $MIU_2 > 0$; that is, $U_2 < U_1 < 1$.
4. $MIU_1 < 0$ and $MIU_2 < 0$; that is, $U_1 < 1$ and $U_2 > U_1$.

It is easy to show (see Lemma 1, below), that, if either e_1 or e_2 is irrelevant to h, then, if $U_1 > 1$, $U_2 < 1$, and *vice versa*. Thus, it is easy to construct examples that satisfy conditions 1 and 4. Take $Pr(e_1|h) = Pr(e_1)$. Then, on an any probability function with $U_1 > 1$, we will have $U_2 < 1 < U_1$, and condition 1 will be satisfied. Similarly, if $Pr(e_1|h) = Pr(e_1)$, on any probability function with $U_1 < 1$, we will have $U_1 < 1 < U_2$, and condition 4 will be satisfied.

For condition 2, we need to have both U_1 and U_2 positive. As is shown in Lemma 1, below, this is possible only if e_1 and e_2 are relevant to h in opposite directions; that is, only if $R(h; e_1)$ and $R(h; e_2)$ have opposite sign. Here's one way to do it. Take, for simplicity, $Pr(h) = Pr(e_1) = Pr(e_2) = 1/2$, and take $Pr(e_1e_2) = 1/4$. Take $Pr(e_1|h) = 0.7$, $Pr(e_2|h) = 0.3$, and $Pr(e_1e_2|h) = 0.24$. The reader can readily verify that these are consistent, and that they determine the full probability function on boolean combinations of $\{h, e_1, e_2\}$. In particular, they entail that $Pr(e_1|\bar{h}) = 0.3$, $Pr(e_2|\bar{h}) = 0.7$, and $Pr(e_1e_2|\bar{h}) = 0.26$. We thus have $U_1 = 24/21$ and $U_2 = 26/21$, satisfying the desired conditions.

For condition 3, we can take the probability assignment described in the previous paragraph and create a new one by interchanging e_2 and $\bar{e}_2$. We have, once again, $Pr(h) = Pr(e_1) = Pr(e_2) = 1/2$, $Pr(e_1e_2) = 1/4$, $Pr(e_1|h) = 0.7$, and $Pr(e_1|\bar{h}) = 0.3$. We also have $Pr(e_2|h) = 0.7$, and $Pr(e_1e_2|h) = 0.46$. These further entail that $Pr(e_2|\bar{h}) = 0.3$, and $Pr(e_1e_2|\bar{h}) = 0.04$. We thus have $U_1 = 46/49$, and $U_2 = 4/9$, and so $U_2 < U_1 < 1$, and condition 4 is satisfied.

Having shown that all four alternatives are possible, we now prove the Lemma alluded to above.

Lemma 1. *Let $\{h, e_1, e_2\}$ be logically independent propositions, and let Pr be a probability function on the boolean algebra generated by this set. We assume that the denominators of the relevant fractions are nonzero, and define U_1 and U_2 as above.*

a) If $Pr(h|e_1) = Pr(h)$ or $Pr(h|e_2) = Pr(h)$, then, if $U_1 > 1$, $U_2 < 1$, and vice versa.

b) If U_1 and U_2 are both less than one, then either e_1 and e_2 are both positively relevant to h, or they are both negatively relevant to h.

c) If U_1 and U_2 are both greater than one, then one of $\{e_1, e_2\}$ is positively relevant to h, and the other negatively relevant.

Proof. Let

$$\begin{aligned} p &= Pr(h); & q &= Pr(\bar{h}) = 1 - p; \\ \alpha_1 &= Pr(h|e_1)/Pr(h); & \alpha_2 &= Pr(h|e_2)/Pr(h); \\ \beta_1 &= Pr(\bar{h}|e_1)/Pr(\bar{h}); & \beta_2 &= Pr(\bar{h}|e_1)/Pr(\bar{h}). \end{aligned} \tag{25}$$

This allows us to write

$$U_1 = \frac{1}{\alpha_1\alpha_2}\frac{Pr(e_2e_2|h)}{Pr(e_1e_2)}; \quad U_2 = \frac{1}{\beta_1\beta_2}\frac{Pr(e_2e_2|\bar{h})}{Pr(e_1e_2)}. \tag{26}$$

Once p, α_1, α_2, β_1, and β_2 are fixed, this yields a constraint on U_1 and U_2:

$$p\,\alpha_1\alpha_2\,U_1 + q\,\beta_1\beta_2\,U_2 = 1. \tag{27}$$

It is convenient to write this in terms of a weighted average of U_1 and U_2. Define

$$w_1 = \frac{p\,\alpha_1\alpha_2}{p\,\alpha_1\alpha_2 + q\,\beta_1\beta_2}; \quad w_2 = \frac{q\,\beta_1\beta_2}{p\,\alpha_1\alpha_2 + q\,\beta_1\beta_2}. \tag{28}$$

Then (27) becomes,

$$w_1\,U_1 + w_2\,U_2 = \frac{1}{p\alpha_1\alpha_2 + q\beta_1\beta_2}, \tag{29}$$

with

$$w_1 + w_2 = 1. \tag{30}$$

It is instructive to rewrite the right-hand side of (29), using the fact that $p\,\alpha_1 + q\,\beta_1 = p\,\alpha_2 + q\,\beta_2 = 1$. A bit of algebraic manipulation yields,

$$w_1\,U_1 + w_2\,U_2 = 1 - \frac{pq(\alpha_1 - \beta_1)(\alpha_2 - \beta_2)}{p\alpha_1\alpha_2 + q\beta_1\beta_2}. \tag{31}$$

From (31) it is readily apparent that, if either e_1 or e_2 is irrelevant to h—that is, if $\alpha_1 = \beta_1$ or $\alpha_2 = \beta_2$, then

$$w_1\,U_1 + w_2\,U_2 = 1, \tag{32}$$

and in such a case, if $U_1 > 1$, then $U_2 < 1$, and *vice versa.* If we want to construct a case in which U_1 and U_2 are both greater than one, this requires the right-hand side of (31) to be greater than one, which means that $\alpha_1 - \beta_1$ and $\alpha_2 - \beta_2$ must have opposite sign: one of $\{e_1, e_2\}$ must be positively relevant to h, and the other negatively relevant. If we want to construct a case in which U_1 and U_2 are both less than one, then $\alpha_1 - \beta_1$ and $\alpha_2 - \beta_2$ must have the same sign: e_1 and e_2 are either both positively relevant, or both negatively relevant, to h. □

9 Errata and comment

I would like to take this opportunity to correct two errors in the Appendix of Myrvold (2003), which are to blamed on proofreading lapses on the part of the author, and not on the editorial staff of *Philosophy of Science*. On p. 421, equation (A2) should read,

$$F(x, y, z) = F(xy, z) + F(x/x, y, y) \tag{33}$$

and equation (A9) should read,

$$H(y, 1/y) = K(y) \log_2(1/y) = -\log_2(y). \tag{34}$$

It should also be pointed out that the four conditions from which the measure of informational relevance are derived are not logically independent. Condition (ii) is entailed by the conjunction of conditions (iii) and (iv), and hence conditions (i), (iii), and (iv) suffice to uniquely determine the function $I(q, p|b)$.

10 Acknowledgments

I would like to thank Michel Janssen, Marc Lange, Bill Harper, and Molly Kao for helpful discussions. I am grateful to Clark Glymour for raising the question, addressed in §6, of how common-cause explanations fit into the framework. This work was supported, in part, by a grant from the Social Sciences and Humanities Research Council of Canada (SSHRC).

References

Douven, I. and J. N. Schupbach (2015). The role of explanatory considerations in updating. *Cognition 142*, 299–311.

Einstein, A. (1905). Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt. *Annalen der Physik 17*, 132–148.

Einstein, A. (1916). Zur Quantentheorie der Strahlung. *Physikalische Gesellschaft Zürich Mitteilungen 18*, 47–62.

Einstein, A. (1989). On a heuristic point of view concerning the production and transformation of light. In *The Collected Papers of Albert Einstein, Volume 2: The Swiss Years: Writings, 1900-1909 (English translation supplement)*, pp. 86–103. Princeton University Press. Tr. Anna Beck.

Einstein, A. (1997). On the quantum theory of radiation. In *The Collected Papers of Albert Einstein, Volume 6: The Berlin Years: Writings, 1914-1917. (English translation supplement)*, pp. 220–223. Princeton University Press. Tr. Alfred Engel.

Glymour, C. (2015). Probability and the explanatory virtues. *The British Journal for the Philosophy of Science 66*, 591–604.

Good, I. J. (1971). Twenty-seven principles of rationality. In V. P. Godambe and D. A. Sprott (Eds.), *Foundations of Satistical Inference*, pp. 123–127. Toronto: Holt, Rinehart and Winston of Canada. Reprinted in Good (1983), pp. 15–19.

Good, I. J. (1976). The Bayesian influence, or how to sweep subjectivism under the carpet. In W. L. Harper and C. Hooker (Eds.), *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, Volume II, pp. 125–174. Dordrecht: D. Reidel Publishing Company. Reprinted in Good (1983), pp. 22-55.

Good, I. J. (1983). *Good Thinking: The Foundations of Probability and its Applications*. Minneapolis: The University of Minnesota Press.

Hendry, J. (1980). The development of attitudes to the wave-particle duality of light and quantum theory, 1900–1920. *Annals of Science 37*, 59–79.

Janssen, M. (2002). *COI* stories: Explanation and evidence from Copernicus to Hockney. *Perspectives on Science 10*, 457–522.

Kao, M. (2015). Unification and the quantum hypothesis in 1900–1913. *Philosophy of Science*. http://www.jstor.org/stable/10.1086/683648.

Lange, M. (2004). Bayesianism and unification: A reply to Wayne Myrvold. *Philosophy of Science 71*, 205–215.

McGrew, T. (2003). Confirmation, heuristics, and explanatory reasoning. *The British Journal for the Philosophy of Science 54*, 553–567.

Myrvold, W. C. (1996). Bayesianism and diverse evidence: A reply to Andrew Wayne. *Philosophy of Science 63*, 661–665.

Myrvold, W. C. (2003). A Bayesian account of the virtue of unification. *Philosophy of Science 70*, 399–423.

Norton, J. D. (2006). Atoms, entropy, quanta: Einstein's miraculous argument of 1905. *Studies in History and Philosophy of Modern Physics 37*, 71–100.

Pais, A. (1982). *'Subtle is the Lord ...': The Science and Life of Albert Einstein.* Oxford University Press.

Perrin, J. (1913). *Les Atomes.* Librairie Félix Alcan.

Perrin, J. (1916). *Atoms.* New York: D. Van Nostrand Company. Tr. D. L. Hammick.

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

Reichenbach, H. (1956). *The Direction of Time.* Berkeley: University of Los Angeles Press.

Schlosshauer, M. and G. Wheeler (2011). Focused correlation, confirmation, and the jigsaw puzzle of variable evidence. *Philosophy of Science 78*, 376–392.

Shogenji, T. (1999). Is coherence truth conducive? *Analysis 59*, 338–345.

Wheeler, G. (2009). Focused correlation and confirmation. *The British Journal for the Philosophy of Science 60*, 79–100.

Wheeler, G. and R. Scheines (2013). Coherence and confirmation through causation. *Mind 122*, 135–170.