Synthese
DOI 10.1007/s11229-011-9895-7

# Verisimilitude: a causal approach

**Robert Northcott**

Received: 3 December 2008 / Accepted: 24 September 2009


**Abstract** I present a new definition of verisimilitude, framed in terms of causes. Roughly speaking, according to it a scientific model is approximately true if it captures accurately the strengths of the causes present in any given situation. Against much of the literature, I argue that any satisfactory account of verisimilitude must inevitably restrict its judgments to context-specific models rather than general theories. We may still endorse—and only need—a relativized notion of scientific progress, understood now not as global advance but rather as the mastering of particular problems. This also sheds new light on longstanding difficulties surrounding language-dependence and models committed to false ontologies.

**Keywords** Verisimilitude · Approximate truth · Causal strength · Causation · Scientific progress · Language dependence · Scientific realism

## 1 Introduction

No fact could seem more obvious than the progress of science, in the sense not only of technical and applied achievements but also of getting closer to the truth. Yet philosophers have had great difficulty in giving a plausible and consistent account of the latter concept. In this paper, I present an approach largely new to the literature, namely defining closeness to the truth in terms of *causes*. In particular, approximate truth will be understood as a matter of capturing accurately the strengths of the causes present in

R. Northcott (✉)
University of Missouri-St Louis, St Louis, MO, USA
e-mail: northcottr@umsl.edu

Published online: 27 February 2011

any given situation.[1] In order to make this idea precise, degree of causal importance, or causal strength (I shall use the two terms synonymously), will have to be defined. After that, we shall then see the many advantages of conceiving of verisimilitude in this way, including that it enables us to underwrite just those technical and applied achievements we began with.

## 2 A definition of causal strength

Let X be a cause variable and Y an effect variable. Y is a function of the state of the world, i.e. of X and W, where W is background conditions (i.e. formally a set of variables representing the state of the world just excluding X).[2] Let $x_A$ denote the actual value of X, and $x_C$ a baseline counterfactual value of X. And let $y_A$ and $y_C$ denote the values that Y takes given $x_A$ and $x_C$ respectively.[3] Then define the *strength* of a cause variable X with respect to an effect variable Y, to be:

$$y_A - y_C \quad [\mathrm{CS}]$$

The actual and counterfactual values of Y are real numbers. A causal strength is the difference between them, and may be positive, zero or negative. The unit of any causal strength is the unit of the effect variable Y. The intuition behind [CS] is straightforward—we are interested in the quantity of effect for which a cause is responsible, and this is just the level of effect with it compared to what that level would have been without it. In particular, the formula captures how much extra effect results from switching to the given level of cause from the neutral level (see below). To do this, we must compare the two levels of effect while keeping all else equal. For this reason, in [CS] the background conditions W must be constant across the two terms.[4]

In the simplest case, the counterfactual $x_C$ will just be the cause's absence. (Strictly speaking, of course, 'absence' should be thought of here merely as convenient shorthand for some contextually salient alternative event.) More generally, often it will be useful to set $x_C$ to be at what Humphreys (1990, p. 38) calls the *neutral level* of X. This he defines, in the case of a variable, as "the level of the variable at which the property corresponding to that variable is completely absent." A key point is that while this neutral level depends on the exact context of interest, thereafter it—and hence the causal strength—can be defined entirely objectively.

1 In the literature, 'approximate truth' is sometimes taken to denote a specific model's accuracy, as opposed to a general theory's overall closeness to the truth, which is denoted instead by 'verisimilitude'. But since I shall define verisimilitude only with respect to specific models anyway, in this paper the two terms are used synonymously.

2 In causal graph terms, there are arrows into Y from both X and W.

3 For ease of exposition, as well as using $y_A$, $x_C$ etc to denote particular values of a variable, throughout I will also use them to denote particular events that instantiate those values.

4 Strictly speaking, in fact the background conditions *do* vary across the formula, because as well as impacting Y, in general the switch from $x_A$ to $x_C$ will also change W too. But for our purposes we may ignore that technical wrinkle.

Notice that any value for the strength of X will be relativized to the levels of other causes of Y, as reflected in the background conditions W. For example, striking a match may have maximum strength with respect to causing a flame if background conditions include sufficient oxygen in the atmosphere, but not otherwise. Thus the strength associated with any particular X–Y pair is context-specific and should be understood as a token rather than type value.[5]

Such a definition of causal strength reflects the common emphasis on causation's difference-making aspect—a cause is something that makes a difference to its effect. Thus, naturally, the strength of a cause is how *much* difference it makes. The form of [CS] can be incorporated into the contemporary Bayes net and causal modeling literatures (Spirtes et al. 2000; Pearl 2000). It is also consistent with Woodward (2003) manipulationist theory of causation. As Woodward highlights at length, there is an intimate connection between this understanding of causal strength and scientific practice. For example, the method of controlled experiment is designed precisely to assess the impact on an effect of changing a causal input (i.e. of $x_A$ relative to $x_C$). The whole point of controlling for other causes is to ensure that, as far as possible, our actual data are a good proxy for the data that *would* have been generated in the ideal (counterfactual) situation of exactly identical background conditions, that is to say the right-hand $x_C$-term in [CS].

Not surprisingly, [CS] or something like it has a long history in several other literatures too. In the philosophy of history, the motivation behind [CS] is similar to that behind several classical views, for instance those in the nineteenth century of Yule and Weber. More recently, many measures in psychology, psychiatry, statistics, epidemiology, evolutionary biology, law and computer science are similar. Moreover, still other measures are closely related, being again essentially comparative of the effect with and without a cause. Much of the literature on probabilistic causality also explicitly endorses this quantitative difference-making understanding of causal strength (Hitchcock 1996). Moreover, so too do attributions of causation in ethics and in the law (Hart and Honore 1985, *passim*), not to mention in everyday life. The definition is formally compatible with other theories of causation as well. Now let us see how it can help with the issue of verisimilitude.

## 3 The seriousness of errors problem

What happens when, standing on the Earth's surface, we drop a ball? Assume a simplified Newtonian model—the ball a point with mass negligible compared to that of the Earth, a perfect vacuum, no other forces to take into account, and so on. Assume also that the gravitational attraction of the Earth on the ball can be represented as the sum of those of the Earth's component parts. Next suppose that there is a large mountain near where the ball is being dropped. Then the true pattern of gravitational pull on the

[5] It turns out that a full account of causal strength also involves a number of other technicalities (Northcott 2008a, b), although for our purposes we may gloss over them without loss. I shall typically quote causal weightings only somewhat schematically in order just to illustrate the philosophical points at hand. I omit any discussion of the semantics of counterfactuals on the ground that, at least with regard to the issues discussed in this paper, that is not the salient locus of philosophical dispute.

ball will be a large amount from the main body of the Earth, plus a little extra from that nearby mountain. But a Newtonian model typically assumes the Earth to be a perfect and uniform sphere and so would neglect the mountain's existence. Therefore there is a mismatch between the true state of affairs, which includes the gravitational pull from the mountain, and the model's postulated state of affairs, which does not. We can represent the truth schematically by (1, 1) and the model by (1, 0), where for (a, b) *a* is the gravitational pull due to the main body of the Earth and *b* the extra pull from the mountain.

Suppose we have a second model that captures the pull due to the mountain but which ignores that from the main body of the Earth. Using our previous notation, this model can be represented by (0, 1). Now suppose we want to compare the relative performance of the two models—(1, 0) and (0, 1)—in capturing the reality (1, 1). At this level of abstraction, we are vulnerable to the conclusion that (0, 1) and (1, 0) are equally close to (1, 1), and so that the two models are equally good. But this offends intuition badly. The gravitational pull from the mountain is only a tiny fraction of that from the rest of the Earth, therefore we want to say that of the two the model capturing the Earth's pull but not the mountain's is much nearer the truth. The key point is that although one of the errors is much more *serious* than the other, this asymmetry is not reflected in our formal scheme, which registers only that each model makes one error. There can be no purely syntactic solution here since the motivation for any asymmetric weighting scheme must come from semantic factors—here, to motivate why the weighting on the Earth's pull should be deemed much greater than that on the mountain's.

Thus the primary worry is *not* the exact specification of a metric (in this case of similarity between 2-vectors); rather, it is to find some way of weighting errors so as to reflect their asymmetric importance. As it were, why devote energy to perfecting measurements to within inches when reading from a map already distorted by miles? The first priority is surely to fix the map. Typically, what matters is not that a model captures a large number of factors but rather that it captures the important ones.

The seriousness of errors problem is widely acknowledged within the literature on verisimilitude (Niiniluoto 1987, p. 38; Oddie 1986, p. 184; Miller 1994, p. 200, 207). Yet in practice attention has nevertheless been paid more to developing syntactic measures of similarity, either between logical propositions (Niiniluoto 1987; Oddie 1986) or sometimes between structures (Kuipers 1987).[6] More recent developments have followed much the same path (see Niiniluoto 1998 for a survey).

[6] For similar criticism see also e.g. Adams (1990), Aronson et al. (1994). Niiniluoto, recognising the desirability of asymmetric weightings, suggests (1987, p. 314) that we might add a coefficient to each term, referring to his own (1978) as an example of how to do this. In that earlier paper he specifies two types of error, 'serious' and 'non-serious', a distinction which in effect does introduce a non-uniform weighting system into his basic distance function. Yet his definition of these two types of error (p. 447) still turns out to be determined by a priori criteria. Oddie, for his part, introduces weights almost purely as a technical normalizing device, with no suggestion of asymmetric weightings or of imputing to the weights any physical significance (1986, p. 45).

It is possible for allowance to be made for extra-logical factors via a quantitative treatment of theories' overall functional forms (e.g. Niiniluoto 1987, p. 368f). However, this move leads to its own problems since arguably verisimilitude can only be satisfactorily defined with respect to context-specific models rather than general theories (Sect. 5).

The notion of causal strength now offers us a solution to the seriousness of errors problem. The Earth's gravitational causal strength (with respect to the effect of the ball's acceleration) really is much greater than the mountain's. This enables us to motivate a weighting function that represents the two causes' objectively asymmetric strengths. In this example the Earth might have, say, ten million times as much mass as the mountain. In both cases, the neutral level of mass is one that makes no difference to the acceleration of the ball, namely zero. Thus, for instance, according to formula [CS] the strength of the mountain's gravitational pull on a ball is yielded by the ball's actual acceleration compared to its acceleration if the mountain had been of mass zero. It follows that the causal strength of the Earth's gravity is 10 mn times that of the mountain's—since it leads to 10 mn times more units of effect, the left-hand term in its formula is 10 mn times higher. Thus we can represent reality not by (1, 1) as before, but instead by (10 mn, 1). The model capturing the Earth but not the mountain is (10 mn, 0), and the model of just the mountain is (0, 1).

## 4 A definition of verisimilitude

So we have a way of representing the asymmetric seriousness of different errors. How exactly should we now formulate verisimilitude? An obvious strategy is to interpret these representations as vectors defined on some abstract vector space of causes. Assessing closeness-of-fit between model and reality then becomes a geometrical matter of comparing the similarity of two vectors.

Suppose there are $n$ causes $X_1, \ldots, X_n$ under consideration, and suppose that, as defined by formula [CS], their true causal strengths with respect to the relevant effect are respectively $t_1, \ldots, t_n$. And suppose that a model ascribes them strengths respectively of $m_1, \ldots, m_n$. Then on my account the verisimilitude of this model is the similarity between the 'T-vector' of true causal strengths $(t_1, \ldots, t_n)$ and the 'M-vector' of the model's postulated ones $(m_1, \ldots, m_n)$. I propose the following definition of distance between the T- and M-vectors:

$$\sum_{i=1}^{n} |t_i - m_i| \quad [\mathrm{VE}]$$

The smaller this distance, corresponding to the model's distance from the truth, the better the model scores for verisimilitude. The units of the distance are the units of the effect term.[7]

Footnote 6 continued

An alternative suggestion is that we define verisimilitude by whether the possible worlds consistent with one theory are closer to the actual world than those consistent with another (Lewis 1986). In principle, the closeness relation between worlds might then incorporate an error's seriousness. But unfortunately with respect to this point 'closeness' is left a primitive and therefore no light is shed on verisimilitude beyond mere intuition. (Analyses of the closeness relation's possible *formal* structure, such as Hilpinen's (1976), still likewise shed no light on the crucial point.)

[7] I do not think that comparability of verisimilitude scores across contexts is an important desideratum because the only significance of the absolute scores is with respect to context-specific interventions (Sect. 7).

The definition is a so-called *Manhattan* distance. Intuitively, it is obtained by travelling from one point to another only in the directions of the orthogonal axes and never taking the direct diagonal, so to speak, of the Euclidean distance. In our earlier example, formula [VE] would tell us that the Manhattan distance between the truth (10 mn, 1) and the Earth-model (10 mn, 0) is 1. That between the truth (10 mn, 1) and the mountain-model (0, 1) is 10 mn. Thus the Earth-model is ranked closer to the truth, just as desired. Use of causal strengths solves the seriousness of errors problem because a model's verisimilitude score now automatically adjusts in proportion to its errors' importance.

It is true that a [VE] score, like any summary measure, inevitably implies some information loss. In particular, it is just the sum of all a model's causal strength errors, thereby blurring which specific causal strengths a model has got wrong. Is it arbitrary to take the simple sum in this way? There are reasons to think not.[8] But in any case, the point is that [VE]'s constituent weightings are sensitive to the seriousness of errors problem. The main critique of rival measures, as we shall see, is that in effect they are arbitrary about the weightings put on each error in the first place, which is surely a much more serious concern.

[VE] has some attractive technical features. Hindrances, or negative causes, are easily represented in the T- and M-vectors as negative causal strengths. And citing an infinity of causes also presents no problem just so long as the magnitude of the T-vector remains finite, which in any real case it presumably will. Finally, in practice false models often just omit mention of a relevant cause altogether rather than get that cause's strength slightly wrong. However, this too can be readily accommodated by [VE]: omitted causes, i.e. entries in the T-vector that a model has failed even to mention, just appear in the M-vector at their neutral levels.

Why the preference for Manhattan over Euclidean distance? Answer: because, in general, only the former is constant with respect to arbitrary subdivision of causal elements. An example will illustrate. Suppose we were now considering the pulls merely of two mountains, one twice as massive as the other. Then let:

(1) The T-vector of true causal strengths be (2, 1), and compare two models:
(2) A model of just the big mountain, with an M-vector of (2, 0), and
(3) A model of just the small mountain, with an M-vector of (0, 1).

On both the Manhattan and Euclidean measures, the models' distances from the truth are respectively 1 and 2. But now suppose that we arbitrarily divide the big mountain into five equally massive component parts. Then:

(1) the T-vector would be (0.4, 0.4, 0.4, 0.4, 0.4, 1) in a new six-dimensional cause-space,

Footnote 7 continued
But if such comparability were thought important, then the distance could be normalized by dividing it by the magnitude of the T-vector, thus yielding a scalar score.

[8] First, the Manhattan distance is a well-defined Minkowski metric. Second, a false model's [VE] score gives the upper bound on the error resulting from an intervention based on that model (Sect. 7). Third, [VE] as currently formulated has the attractive technical features about to be outlined in the text. In particular, it is clearly preferable to a Euclidean alternative.

(2) the superior big-mountain model's M-vector would be (0.4, 0.4, 0.4, 0.4, 0.4, 0) and
(3) the inferior small-mountain one's would be (0, 0, 0, 0, 0, 1).

Which model would be considered closer to the truth now? For the superior big-mountain model, the Manhattan and Euclidean distances both remain 1. For the inferior small-mountain model, the Manhattan distance remains 5(0.4) = 2, but now the Euclidean distance = $[5(0.4)^2]^{1/2} = 0.89$ approximately. That is, whereas both the Manhattan distances are unchanged, the Euclidean distance has now *altered* for the small-mountain model. Worse, it has altered to such a degree that the overall ranking of the two models has reversed. Yet the only change was to describe the big mountain in terms of five component parts rather than as one whole. In other words, we have exactly the same physical situation and simply by changing our arbitrary partition of the *agreed* causal elements the ranking of the two different models has reversed, which is surely unacceptable.

So according to this approach, verisimilitude is a matter of how closely a model's postulated causal weightings replicate the true causal weightings. Or, put more loosely, has a model captured the causes that really matter? A general advantage of [VE] is that, compared to alternatives in the literature, it is conveniently simple. Another is its naturalness—all the causes and strengths are expressed in the language and units in which the scientist is naturally working. And which causes of some effect might have been the most important is arguably precisely the typical stuff of scientific (and everyday) debate.

One drawback might seem to be that the definition is not applicable everywhere because not all theories are theories about causes. However, as will be argued in the next section, verisimilitude—and our definition—applies to context-specific models rather than to theories as a whole, and such models are in general able to be cast into causal form (Pearl 2000). For example, Newton's law of gravity was from the start famous for being merely associational, yet in token cases, as we have already seen, it is readily taken to imply particular causal strengths. Therefore once we accept the restriction to token applications, the subsequent restriction also to causal models is not in fact any further restriction.[9]

A further category of accuracy claims apparently concerns neither general theories nor causes. For example, consider two estimates of the age of the Earth: 6,000 years; and 11 billion years. Which is nearer the truth? It is true that [VE] does not answer that directly. On the other hand, the two different views will certainly have different causal consequences.[10] Consider their implications, for instance, for the chance that fossil-formation processes will have produced some fossil currently discoverable near the Earth's surface. This is a causal strength, with respect to the effect term of the probability of there being such a fossil. And the 11-billion-year estimate would

[9] One recent strand in philosophy of science and metaphysics in any case takes causal capacities, rather than associational laws, to be ontologically basic, e.g. Cartwright (1989), Bird (2007b). Adopting that view would make our definition applicable all the more easily.

[10] Similar remarks apply to *structural* claims. Take different models of the structure of DNA, say. Although concerning constitution, these models will have different causal consequences, and hence predict different causal strengths, for all manner of token effects.

presumably score *better* here than the 6,000-year one, even though in strict numerical terms it is of course further away from the true figure of 4.6 billion years. Indeed, we also thereby avoid controversies concerning how best to define this 'numerical' distance—arithmetically 6,000 years may be the closer to 4.6 billion, but multiplicatively? Logarithmically? Rather than try to justify some choice as canonical, it seems to me better to accept that with respect to some causal strengths the 11-billion-year estimate will yield values closer to the true ones, while with respect to others the 6,000-year estimate will do better. In causalist eyes there is no univocal answer, much as there was none for which of two theories is closer to the truth. In both cases, this merely reflects that generally there is no univocal answer as to the seriousness of a particular error.

There is one branch of the literature, tracing from Giere (1988), that anticipates this paper's general approach by introducing specifically interpreted extra-logical factors. Aronson et al. (1994), for instance, analyze verisimilitude in terms of type-hierarchies intended to capture actual structural relationships between natural kinds. Smith (1998), influenced by dynamical systems theories, analyzes in terms of theories' postulated geometrical structures. But causal strengths seem a much less unwieldy tool than type-hierarchies and a more widely applicable one than geometrical structures. Barnes (1995) does propose an analysis of approximate causal explanation, but with no allowance for weighting across causes asymmetrically. Weston (1992, p. 68) also supports the adoption of a causal ontology for tackling verisimilitude although his approach is otherwise rather different from ours, concentrating as it does (like Barnes's) on the evaluation of general theories rather than context-specific models. Humphreys (1990, p. 115 note 28) mentions exactly our approach of defining approximate truth in terms of causal strengths, but unfortunately to my knowledge nowhere developed this idea.

Bird (2007a) argues for an interesting alternative approach, namely to define scientific progress as the accumulation of knowledge rather than semantically as increasing closeness to truth. So far as I can tell, this paper's approach fills a previously empty slot in Bird's taxonomy, namely the "accumulation of true scientific belief" (2007a, p. 100), where the beliefs in question are those concerning the values of relevant causal strengths. In any case, [VE] avoids Bird's main critique of semantic views, namely their inability to discount fluke empirical successes (footnote 16). Bird's own view, meanwhile, does not offer any analysis of the language-dependence and false-ontology problems (Sects. 8 and 9). Rather, those classic difficulties in effect get swept under the carpet by his taking 'knowledge' to be a primitive.

## 5 Context-specific versus general

In our scheme, similarity between truth and model is mediated via weights derived from causal strengths. Yet in turn any causal strength will, according to [CS], depend on context and so therefore will similarity and hence verisimilitude. This reflects how the seriousness of any given error itself depends on context. For instance, getting wrong the amount of oxygen in the atmosphere may be unimportant if a match is anyway too wet to be struck, but critically serious if the match is dry. Causal strengths are tokens, not types. The moral is that the seriousness of errors problem can only be solved context by context, application by application. We cannot justify declaring some particular error *generally* serious or not.[11]

One consequence is that it is problematic to compare the verisimilitude of different functional forms. Kuipers has developed a measure of this for the special case where one theory is uniformly better than another, thereby circumventing the seriousness of errors problem. Such a happy situation will likely occur only rarely, alas (as Kuipers acknowledges). And Niiniluoto has considered well-behaved metrics for the general case although again, except for special cases, they are still subject to the seriousness of errors problem. In general, any univocal value that comparisons of functional form yield for the distance between two functions must be implicitly committed in turn to some univocal weighting of errors, yet the latter is precisely what is in question. For example, suppose one theory of Newtonian gravity posits an inverse-cube law, while a second posits instead inverse-square but with the gravitational constant G half its actual value. Which is the better approximation? It seems difficult to justify any univocal answer because the import of the different errors will vary application by application. Sometimes the first theory will yield more accurate predictions and causal strengths, while other times the second will.[12]

Many causal models are non-linear and non-additive. For example, when air resistance is added as a new factor to the ballistic equation, the new equation is a complicated exponential function. Therefore its distance from a postulated linear function may vary greatly, depending on a projectile's speed, wind conditions, and so forth. But for similar reasons to above, such cases are not a problem for [VE] because it does not seek to ascribe verisimilitude to models or functions as a whole, but rather only to particular sets of causal strengths generated by those functions. So for one particular speed and wind conditions the erroneously linear function might be mistaken by a certain amount, for different conditions by a different amount. Ultimately, the non-additivity is just another instance of the seriousness of errors problem implying that we cannot satisfactorily judge the verisimilitudes of general theories or functional forms.

The modern verisimilitude literature has usually concentrated on general theories nevertheless.[13] What could be the root of this generalist emphasis? The answer seems to be that ever since its origin with Popper, the literature has been motivated primarily by the debate between scientific realism and anti-realism. In particular, since it is accepted that few if any of our best theories are literally true, a satisfactory account of verisimilitude has been seen as important, perhaps even essential, to the realist position, or at least to realism about theories (Putnam 1975; Newton-Smith 1981; Miller 1987; Boyd 1990; Psillos 1999). It is just this purpose of buttressing realism that has been

[11] Goodman, in his classic analysis (1972), likewise argues that similarity in general is ill defined without some notion of weighting, and that choice of weighting scheme is in turn highly context-specific and so needs to be adjusted with every application. Niiniluoto (1987, p. 38) and Oddie (1986, p. 56) each quotes Goodman approvingly in this regard. Lewis (1986, pp. 24–25) also argues for a closely related point.

[12] I am therefore pessimistic that a satisfactory definition can be obtained of 'legisimilitude', i.e. of closeness not to empirical facts but rather to laws, notwithstanding several attempts (Cohen 1980; Niiniluoto 1987, Chap. 11; Liu 1999)

[13] Many of the definitions offered are formally applicable to singular statements too (Oddie 1986; Niiniluoto 1987, Chap. 8). The point is that their weaknesses arguably stem from an originally generalist motivation.

the motivation for the most cited works in the literature (Oddie 1986; Niiniluoto 1987; Kuipers 1987; Weston 1992). Related to this debate is the notion of scientific progress that sees successive theories as achieving closer and closer approximations to the truth—'convergent realism'. From Popper onwards, a desire to underpin this notion has also been a frequent spur (Popper 1963, 1972; Miller 1994; Aronson et al. 1994; Barnes 1995).

But the context-specific analysis offered here is likely to disappoint all such motivations because it suggests that the very same theory may score highly in one context but only poorly in another. So any overall verisimilitude score for a theory could only be a crude average of its scores over various particular applications. Accordingly, I side with authors skeptical of global convergent realism (Kuhn 1962; Laudan 1984). A canonical general weighting of the seriousness of errors appears impossible even for one theory as a whole, let alone for science as a whole. Rather, progress should be thought of as relative only to specific explananda: in particular, for a given true set of causal strengths, of course different models may get closer and closer to postulating the correct values. In this contextualized sense we can indeed speak of objective progress, reminiscent perhaps of Kuhnian puzzle-solving.[14] But, notwithstanding our intuitions to the contrary, any grander notion of scientific progress seems impossible to sustain.

Then again, it is not obvious to me how much we really even *need* this grander notion. Perhaps realism can be content with a more qualified kind of convergence, especially as scientific investigation is often aimed at a particular target and so is contextual in the relevant sense. I take it, in line with much recent literature in philosophy of science, that science's applied successes can often be thought of as detailed knowledge of particular mechanisms. This in turn can be cashed out as yielding accurate knowledge of particular causal strengths for any given context. The point is that the verisimilitude of grand background theory is not directly relevant. For instance, the smooth operation of a steam engine or bicycle, triumphs of human ingenuity though they be, need not imply the verisimilitude of Newtonian physics as a whole, only the verisimilitude of particular causal strengths derived from it. Moreover, science can also progress in the sense merely of discovering accurate causal strengths for new explananda. Advances in computer technology, for example, need not imply any *general* theoretical advance closer to the truth.

Moreover, usually philosophical attention in this area has been concentrated on famous paradigm-shifting episodes from the histories of chemistry and fundamental physics. Yet from the perspective instead of the special sciences, a context-specific notion of verisimilitude seems quite sufficient. For example, nobody believes that rational choice models in economics describe the literal truth, or that any particular model is true in any sense universally. Rather, the key desideratum is the applicability of a model to the particular explanandum at hand. Successful applied work depends on a careful selection of the right model for the right problem, often borrowing piecemeal from several different models and then working very hard to see how to fit them to

[14] Indeed the normative authority claimed for verisimilitude scores below (Sect. 7) is one way of fleshing out that claimed by Kuhn for normal science.

a particular real-world situation, doing this in turn often by utilizing much local and informal 'rule of thumb' engineering knowledge, or by running *sui generis* experiments or simulations (Alexandrova 2008; Alexandrova and Northcott 2009). (One excellent example is the design by game theorists of the 1994 US spectrum auction (Plott 1997; Guala 2001). Even if it could be formulated, the *general* verisimilitude of, say, auction theory is an irrelevance.)

## 6 Interest-relativity

Consider two models of the causes of lung cancer. The first cites asbestos and gives a very accurate estimate of its causal strength, but makes no mention of smoking. The second model mentions both asbestos and smoking and gives estimates for their causal strengths, although these estimates are not as accurate as the first model's for asbestos. Which of the two models should be preferred? On one hand, the first one is, as far as it goes, the more accurate of the two. On the other hand, the second one has captured more of the factors at play and so although less accurate is also more complete.

This example illustrates how verisimilitude cannot be calculated without an exact specification of just what we are interested in. Here, we could be interested in one of two different things—either asbestos alone or else all the other causes of lung cancer too. If the former, then all that would matter would be the causal strength of asbestos, so our narrow but accurate first model would score the better and the second model's additional coverage of smoking would garner it no extra credit. But if the latter, then now the less accurate but more complete second model will likely be preferred.[15]

In our scheme, interest-relativity is readily represented as the initial *choice* of causal strengths to be captured, i.e. the choice of T-vector. Which of two models scores best will turn on this choice, i.e. on what we are trying to explain. Mere accuracy is not itself enough; we need also *relevant* accuracy. A corollary of this is that it is possible for many different models of the same event each to score highly for verisimilitude—with respect to different focuses of interest (e.g. are we trying to capture the event's proximate or underlying causes?).

Although in the literature it is rarely discussed at length, of course some allowance for interest-relativity is obvious and uncontroversial (Miller 1994; Niiniluoto 1998, pp. 13–14). Nevertheless I think there is value in highlighting the issue so explicitly. The reason is that the literature's generalist focus has fuelled the attempt to capture in a single measure both a theory's accuracy and its comprehensiveness. But as well as causing its own difficulties this also now seems just ill motivated since our treatment of interest-relativity does, for any one context, just the job that Popper originally earmarked generally for comprehensiveness. In particular, from a context-specific point of view, a theory's generality is useful only if it results in the theory covering what we are interested in here and now; who cares whether it also covers *other* problems?

---

15 In this way our definition satisfies (for any given context) a common desideratum in the literature, namely the ability to rank Newtonian mechanics above a trivial tautology, i.e. to rank a false but useful theory ahead of a true but uselessly narrow one.

That is, its comprehensiveness per se is irrelevant. Thus while our account of verisimilitude indeed concentrates only on accuracy, in a sense this omission is made good because the motivation for including comprehensiveness at all is satisfied instead by our explicit incorporation of interest-relativity. (This also shows that although verisimilitude is indeed inextricably bound up with interest-relativity, there is no necessity to abandon ambitions for a unified analysis of it just on that account—contrary to the pessimism on this point of Smith (1998) and Psillos (1999).)

Another difficulty also now falls into place, namely the troublesome conundrum that sometimes the consequences of an approximately true theory are not themselves approximately true. For example, the butterfly effect in dynamical systems means that a very good approximation of the underlying theoretical equations may yield predictions that quickly become very bad approximations of the actual outcomes, even assuming perfect knowledge of initial conditions. That is, there seems to be a divergence between achieving approximate truth with respect to one part of the problem and achieving it with respect to another part (Miller 1994,pp. 198–202). Should we put more weight on the underlying equations or on the final predictions? No approach emphasizing the verisimilitude of general theories would seem to have any solution to this dilemma. Our account, by contrast, only even defines verisimilitude with respect to particular causal strengths. Thus in this example we need feel no obligation (contrary to Smith 1998) to pronounce on the 'overall' success of some theory that got the equations almost right but the outcomes wildly wrong. Rather, we would first specify whether we were *interested* primarily in the equations or in the outcomes, or—more precisely—in which particular outcomes, initial or final. Thus there is no paradox whereby it seems that the same theory must be adjudged simultaneously both approximately true and not approximately true. Rather, it may be adjudged approximately true with respect to one focus of interest but not with respect to another. On our account, there is no additional general calculation to be made.

This example also illustrates the role of subjectivity. Always there is an unavoidable subjective element—in this case, are we interested in the initial conditions or the final outcome? *Once* we have characterised our subjective interest, *then* we can formulate an objective measure. Niiniluoto explicitly defends this conception of objectivity as well. Perhaps surprisingly we can also enlist in support Popper himself since Miller (1975, p. 186) reports Popper's opinion that: "We can relativize any theory, in a perfectly objective manner, to its historical setting, and thus we may be able to construct an objective … mode of comparison of two theories." As just argued, interest-relativity, and hence a certain kind of subjectivity, cannot be avoided anyway. So no extra subjectivity is being unduly imported—and none unduly denied either.

Notice also that interest-relativity alone is enough to imply the context-specificity of verisimilitude just as much as did the seriousness of errors problem. In a sense, indeed, interest-relativity might be seen merely as an instance of the seriousness of errors problem, broadly construed. In any case, both imply that once a theory has fallen short of full truth, thereafter there just is no univocal answer as to how *much* it has fallen short.

## 7 Interventions and quantitative authority

What status should we attach to scores for verisimilitude based on causal strengths? First, they are relativized to choice of causal *ontology*. Second, even when two models agree ontologically still they may disagree on choice of *vocabulary* within that ontology, so scores are relativized to that too. I discuss these two issues in the next two sections. Third, scores are also relativized to the choice of *measure* of similarity. Fourth, as we have just seen, they are also relativized to choice of *interest*.

Given these multiple relativizations, whence verisimilitude scores' objective authority? The answer is that they carry it with respect to *interventions*. Our definition of causal strength is such that if a cause has a strength of 2 units, say, then, given the context, implementing that cause—starting from the neutral level—would lead to an extra 2 units of the specified effect. Thus our absolute scores for verisimilitude, by reflecting the relevant causal strengths, in turn reflect the absolute impact of particular causal interventions. Assuming that the evaluations of the relevant counterfactuals carry objective weight, so therefore do our approach's verisimilitude scores.

Moreover, arguably it is only with respect to interventions that we really care about quantitative results anyway. Regarding verisimilitude itself, in contrast, typically our concern is only qualitative. That is, typically we are concerned to know only whether one model is nearer the truth *than another*. The classic concentration on scientific progress, for example, is over whether successive theories or paradigms are nearer the truth than their predecessors. It is rare for philosophers to ask, 'how approximately true is Newtonian mechanics?' More common rather, 'is Newtonian mechanics *more* approximately true than Aristotelian mechanics?' So although our definition's quantitative advice may carry authority only with respect to context-specific interventions, that is also the only time that such quantitative succor is especially desired.

The authority of interventions also establishes the link between our scores for a model's verisimilitude and that model's empirical success. In the mainstream logical-similarity approach such a link is difficult to establish, or at least can be established only with respect to a rather abstract understanding of empirical success in terms of true and false deductive consequences (for references see Niiniluoto 1998, p. 23). But on our scheme a high score for verisimilitude directly underwrites a high accuracy of intervention, i.e. underwrites that the impact on the effect variable of an intervention will be near to that anticipated. There is thus direct empirical endorsement of those models scoring well for verisimilitude—the truer our model, the better we will be able to manipulate the world. In turn, good empirical performance implies high verisimilitude.[16]

[16] An exception to this latter claim are 'fluke' empirical successes. For example, suppose that every time I perform a rain-dance, the next day it indeed rains. On the face of it, a model postulating a causal link between the two events therefore boasts excellent empirical success. But it would not thereby score well for verisimilitude on our approach, since formula [CS] stipulates that the causal strength of the rain-dance is defined relative to the *counterfactual* of not performing that dance. And assuming that even without the dance it would have rained each time anyway, it follows that the dance has zero causal strength here and so any model that postulates otherwise will score badly. Thus, happily the empirical successes that our definition fails to reward are just those fluke ones that we would not *want* it to reward.

## 8 Miller's problem

Turn now to the relativization to vocabulary. David Miller famously demonstrated that rankings of approximate truth can in principle always be reversed simply by arbitrary switches between logically equivalent languages (Miller 1974, 1975). The shadow of this result has loomed large. Of the other relativizations mentioned earlier, that to focus of interest might still allow an objective definition once such a focus has been specified; that to choice of measure might be resolved by the literature settling on a preferred metric; and that to ontology might be overcome by privileging whatever is the true ontology of the world. But no similar resolution seems possible of the fourth relativization, highlighted by Miller, namely that to vocabulary. For, the thought runs, even if science ever were to establish conclusively the true ontology of the world, still that would underdetermine the predicates we might use consistent with that ontology.

It is certainly true that our own definition is relativized to language/conceptual scheme. The question is: what follows from that? I endorse the answer of Mormann (1988). He points out (as have others, for instance Niiniluoto) that Miller's ingenious examples are all instances of a single underlying mathematical fact—namely that homeomorphisms or continuous bijective mappings between topological spaces need not preserve metric structure. Mormann shows that our choice of predicates, i.e. of physical variables, in effect determines the metrical structure of the hypothetical magnitude-space on which verisimilitude is defined. Therefore a different choice of variables may lead to a different metrical structure and thus to different values for verisimilitude, as per Miller's demonstrations. The key question is: are such transformations, just because mathematically possible, therefore also philosophically significant?

Mormann argues that to accept that they are is to adopt a species of geometric conventionalism. In particular, it is to claim that it is entirely a matter of convention which variables we should adopt when measuring two theories' experimental outcomes—just as much as would be the choice of measuring length using metres or yards. But, following Putnam and others, Mormann rejects such a conventionalism, concluding:

> ... many properties are part of a physical magnitude and ... the meaning of such a term is not exhausted by a simple formal definition. The relevant structure of physical magnitude spaces is much richer and it depends in such complicated ways on other empirical theories and conceptions that it cannot bear just any *prima facie* possible formal manipulation as Miller asserts. (Mormann 1988, p. 517)

In other words, not just any variables will do; rather, they must also be physically meaningful. This puts rigorous flesh on what has anyway been the most common response to Miller's problem, namely the obvious recommendation to privilege those predicates preferred by working scientists themselves (Oddie 1986, p. 159; Weston 1992, p. 68; Niiniluoto 1998, p. 17).

Footnote 16 continued

This successful treatment of fluke empirical successes also undermines the main advantage that Bird (2007a) claims for his knowledge account of scientific progress over semantic accounts such as this paper's.

Tellingly, Mormann points out that Miller himself is forced to reject his own conventionalism. Consider the one-dimensional case regarding accuracy of predictions. Can we even say that, for example, 5 (units of some physical quantity) is closer to 6 than to 7? If he is not anti-conventionalist here, then Miller is hoist on his own petard. His own demonstrations of Miller-reversals themselves rely on us being able to make assessments of approximate truth unambiguously for such one-dimensional cases, since the very notion of a reversal implies the existence of some well-established ordering that can *be* reversed. Miller responds reasonably that: “reversals of ordering by accuracy can indeed be obtained even in the one-dimensional case if we are prepared to allow discontinuous transformations ... But this cannot be thought to be anything like as interesting, since some topological restraints must be insisted on if our reformulated hypotheses are to be reformulations at all” (Miller 1994, p. 226). But as Mormann asks (1988, p. 518): “why is the metrical structure [of a magnitude space] conventional while its topological structure is not?” And as he comments, there can be no general answer, only consideration in each individual case of what the scientific context tells us should and should not be taken as conventional. Moreover, Miller himself appears to accept Mormann’s analysis (1994, p. 231).

Note that Mormann’s critique of Miller does no damage to the argument (Sect. 4) favoring the Manhattan over Euclidean distance for our own measure. This is because there are often many different natural ways of stating a given causal ontology. For instance, often the conjunction of one cause and another cause will itself be a true cause too—such as, in our earlier example, ‘mountain’ rather than the different component parts of that mountain. Either choice of predicates is endorsed by the wider web of scientific theory, so this time it *is* imperative that a measure be constant with respect to a switch between them.

## 9 Models with false ontologies

Perhaps partly as a consequence of the attention paid to Miller’s problem, unfortunately rather less has been paid to a different one, this time one not of vocabulary but rather of ontology (although see Niiniluoto 1987, p. 394ff). In the literature, verisimilitude is typically relativized to a choice of target in a particular language and then defined as a distance between propositions. Yet at the core of the pre-philosophical conception of verisimilitude is arguably distance between worlds or states of the world. The issue here is not how to measure distance. Rather, it is the different one of whether the syntactic structure of a theory in some way reflects the common structure of the worlds in which it is true (Brink 1989, p. 188). That is, can we use distance between propositions as a proxy for distance between worlds?

The heart of the matter is this: can we say anything about when a false ontology is acceptable and when not? Suppose one Newtonian model of a spacecraft’s trajectory leads it to land safely whereas a less accurate second one leads it to crash. This would seem to be paradigmatic of a case suitable for a judgment of approximate truth. By contrast, a diagnosis of my fever in terms of four hobgoblins would seem no more or less true than one in terms of two hobgoblins and a dragon. Yet why is the Newtonian false ontology less troublesome than the one of hobgoblins and dragons? That is, there is a philosophical difficulty here not only when two competing models are ontologically incompatible but also, more simply, whenever any model is ontologically mistaken. It is only the first of these two difficulties that has preoccupied the scientific realism debate, yet a solution to the first would seem to depend on a solution to the second.

For a false-ontology model to be verisimilitude-apt, according to our scheme there must be some implicit reference by the false ontology to entities in a true ontology, in particular enough to enable us to assign—sufficiently non-arbitrarily—objective causal strengths to the entities of the false ontology. I do not have any more precise criterion to offer. In a similar vein, Aronson et al. (1994) state that a theory's ontological adequacy can be determined by the ability it affords us to manipulate individuals of the kinds that it treats. I take this position to be similar to ours since according to [VE], a causal strength is well established if and only if it licenses a (sufficiently) successful associated intervention (or 'manipulation').

In some cases, such as Newtonian models of spaceflight, passing our criterion seems to be widely perceived as unproblematic. In others, such as medicinal hobgoblins, quite the opposite. In still others, matters appear to be more borderline. Might we legitimately interpret 'dephlogisticated air' to be oxygen, for example? Sometimes, doing so licenses correct causal interventions, and in such cases does that not indeed reflect some genuine knowledge of the world? When the criterion does deliver a clearly negative verdict, no values for true causal strengths can be inferred and hence according to our scheme no judgment of verisimilitude can be delivered—which, as the hobgoblins illustrate, is surely just as we would want it in those cases.

There is no denying that the whole issue is unattractively messy, touching as it does on deep and controversial questions of epistemological warrant. Unfortunately, the issue is also unavoidable. In particular, notwithstanding the unattractive messiness, some resolution of it is necessary for underpinning clearly useful notions such as deeming one Newtonian model more approximately true than another. The alternative is just to wash our hands of actual science, ruling instead simply that all (or most of) its models are as bad as each other. My own view is that the factors guiding our judgments here are not so arbitrary as to support such a skeptical conclusion. But ultimately, the matter turns on how impressed one is by successful interventions.[17]

## 10 Conclusion

Verisimilitude has long acquired a certain notoriety, many questioning whether any rigorous sense can be made of the notion at all. This may be because philosophical attention has focused unduly on the case of famous general theories, for instance on the task of establishing objectively that Newtonian mechanics is somehow *as a whole*

[17] The problems surrounding false-ontology cases might seem to argue in favour of a simple empirical conception of verisimilitude, notwithstanding that conception's other difficulties. After all, the thought runs, at least we always have a common currency of accurate predictions. But this strategy is *also* dependent on there being available an acceptable true-ontology interpretation of a false-ontology term—this time of the parameter we wish actually to measure, e.g. 'quantity of dephlogisticated air'. So the same issue reappears. (Miller 1994 also makes this point.)

closer to the truth than is Aristotelian mechanics. Skepticism here, I have argued, is well founded. Any viable definition must inevitably be applicable instead only to context-specific models rather than general theories.

That leaves verisimilitude unable to offer any particular support for the notion of global scientific progress. But in return, we are left free to develop a definition that can solve the seriousness of errors problem. Such a definition is well framed in terms of causal strengths since that in turn ensures: that it is widely applicable; that it is simple; and that it still carries quantitative normative authority with respect to interventions. When a model is framed in a false ontology, our definition fails to offer a verdict only when we would not want one anyway. And given its normative authority, it can underwrite just those technological successes that arguably motivate the more diffuse notion of global progress in the first place. The picture of scientific progress yielded by [VE] is not grand overall convergence on a single truth but instead the worthy mastering of miscellaneous particular problems, one by one.

## References


Adams, E. (1990). Review article: Ilkka Niiniluoto. Truthlikeness. *Synthese, 84*, 139–152.

Alexandrova, A. (2008). Making models count. *Philosophy of Science, 75*, 383–404.

Alexandrova, A., & Northcott, R. (2009). Progress in economics: Lessons from the spectrum auctions. In D. Ross & H. Kincaid (Eds.), *Oxford Handbook of Philosophy of Economics* (pp. 306–337). Oxford.

Aronson, J., Harre, R., & Way, E. (1994). *Realism rescued: How scientific progress is possible*. Chicago: Open Court.

Barnes, E. (1995). Truthlikeness, translation, and approximate causal explanation. *Philosophy of Science, 62*, 215–226.

Bird, A. (2007a). What is scientific progress? *Nous, 41*, 92–117.

Bird, A. (2007b). *Nature's metaphysics*. Oxford: Oxford University Press.

Boyd, R. (1990). Realism, approximate truth, and philosophical method. In C. Savage (Ed.), *Scientific theories* (pp. 355–391). Minneapolis: University of Minnesota Press.

Brink, C. (1989). Verisimilitude: Views and reviews. *History and Philosophy of Logic, 10*, 181–201.

Cartwright, N. (1989). *Nature's capacities and their measurement*. Oxford: Clarendon.

Cohen, L. J. (1980). What has science to do with truth? *Synthese, 45*, 489–510.

Giere, R. (1988). *Explaining science: A cognitive approach*. Chicago & London: University of Chicago Press.

Goodman, N. (1972). Seven strictures on similarity. In: *Problems and projects* (pp. 437–447). Indianapolis: Bobbs-Merrill.

Guala, F. (2001). Building economic machines: The FCC auctions. *Studies in History and Philosophy of Science, 32*(3), 453–477.

Hart, H. L. A., & Honore, A. (1985). *Causation in the law*. Oxford: Clarendon Press.

Hilpinen, R. (1976). Approximate truth and truthlikeness. In M. Przelecki, K. Szaniawski, & R. Wojcicki (Eds.), *Formal methods in the methodology of empirical sciences* (pp. 19–42). Dordrecht: Reidel.

Hitchcock, C. (1996). The role of contrast in causal and explanatory claims. *Synthese, 107*, 395–419.

Humphreys, P. (1990). *The chances of explanation*. Princeton: Princeton University Press.

Kuhn, T. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Kuipers, T. (Ed.). (1987). *What is closer-to-the-truth?* Amsterdam: Rodopi.

Laudan, L. (1984). *Science and values: The aims of science and their role in scientific debate*. Berkeley: University of California Press.

Lewis, D. (1986). *On the plurality of worlds*. Oxford: Basil Blackwell.

Liu, C. (1999). Approximation, idealization, and laws of nature. *Synthese, 118*, 229–256.

Miller, D. (1974). Popper's qualitative theory of verisimilitude. *British Journal for the Philosophy of Science, 25*, 166–177.

Miller, D. (1975). The accuracy of predictions. *Synthese, 30*, 159–191.

Miller, D. (1994). *Critical rationalism: A restatement and defence*. Chicago: Open Court.

Miller, R. (1987). *Fact and method*. Princeton: Princeton University Press.

Mormann, T. (1988). Are all false theories equally false?. *British Journal for the Philosophy of Science, 39*, 505–519.

Newton-Smith, W. (1981). *The rationality of science*. Boston MA & London: Routledge & Kegan Paul.

Niiniluoto, I. (1978). Truthlikeness in first-order languages. In J. Hintikka, I. Niiniluoto, & E. Saarinen (Eds.), *Essays on mathematical and philosophical logic* (pp. 437–458). Dordrecht: Reidel.

Niiniluoto, I. (1987). *Truthlikeness*. Dordrecht: Reidel.

Niiniluoto, I. (1998). Verisimilitude: The third period. *British Journal for the Philosophy of Science, 49*, 1–29.

Northcott, R. (2008a). Weighted explanations in history. *Philosophy of the Social Sciences, 38*, 76–96.

Northcott, R. (2008b). Can ANOVA measure causal strength?. *Quarterly Review of Biology, 83*, 47–55.

Oddie, G. (1986). *Likeness to truth*. Dordrecht: Reidel.

Pearl, J. (2000). *Causality*. New York: Cambridge University Press.

Plott, C. (1997). Laboratory experimental testbeds: Application to the PCS auction. *Journal of Economics and Management Strategy, 6*(3), 605–638.

Popper, K. (1963). *Conjectures and refutations: The growth of scientific knowledge*. London: Routledge & Kegan Paul.

Popper, K. (1972). *Objective knowledge: An evolutionary approach*. Oxford: Clarendon Press.

Psillos, S. (1999). *Scientific realism: How science tracks truth*. London: Routledge.

Putnam, H. (1975). How not to talk about meaning. In *Philosophical papers, Vol. 1, mathematics, matter and method* (pp. 250–269). Cambridge: Cambridge University Press.

Smith, P. (1998). Approximate truth and dynamical theories. *British Journal for the Philosophy of Science, 49*, 253–277.

Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search* (2nd ed.). Cambridge: MIT Press.

Weston, T. (1992). Approximate truth and scientific realism. *Philosophy of Science, 59*, 53–74.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.