22 November 2024
John Eyre

Corroboration and uncertainty

Abstract

In science, uncertainty is always with us, both in observations and in predictions from theory. This paper investigates the important role played by uncertainty in two related problems in philosophy of science: corroboration and the language-dependence of closeness to truth. When predictions from theory are confronted with observations, the theories can be falsified or corroborated. This is an iterative process, since new observations may falsify a previously corroborated theory. Quantification of uncertainty is crucial in determining whether a prediction is consistent with an observation or not. Moreover, quantitative measures of corroboration must be time-dependent, because they rely on estimates of uncertainty, which are always open to reassessment. We also discuss some consequences of these ideas on corroboration for theories of verisimilitude.

In response to Karl Popper's original concept of verisimilitude, Pavel Tichý offered an alternative method for ranking theories in terms of closeness to truth. David Miller raised objections, showing that rankings within Tichý's system did not survive transformation into a different mathematical space. This problem is named here the "Miller-Tichý paradox", and it has implications for the language-dependence of closeness to truth. We show how this paradox can be resolved by taking account of the inevitable uncertainties in observations and in predictions from theory.

1. Introduction

In the history of philosophy, Karl Popper played a key role in prising the concept of knowledge away from the concept of certainty. For Descartes, the only knowledge was certain knowledge, and certain knowledge was the goal of our enquiries. Hume showed that, outside the realms of logic and mathematics, certainty is not possible, and so he concluded that knowledge (understood as certain knowledge) is not possible. Popper (1959, 1963, 1972) showed how knowledge is possible without certainty: that knowledge is embedded in theories, which are conjectural, and that knowledge advances through the falsification of theories.

This is a negative approach; the falsification of a theory tells us what is not the case. This reduces the range of what might be the case, but only a little. A more positive and optimistic approach to the growth of knowledge in general, and scientific knowledge in particular, is as follows: science makes progress – false theories are superseded by better theories and although the latter will be shown in time also to be false, they are better than the theories they replace because they are closer to the truth.

Popper attempted to quantify this progressive process through the concept of verisimilitude. He distinguished sharply between the concepts of corroboration and verisimilitude (although he did regard degree of corroboration as an indicator relevant to verisimilitude). A theory is corroborated when it survives attempts to falsify it – when predictions from the theory are found to be consistent with observations. However, a theory that is corroborated today may be falsified tomorrow, and so corroboration is time-dependent. On the other hand, the verisimilitude (or truthlikeness or closeness to truth) of a theory is conceived as being time-independent, like truth; verisimilitude assesses closeness not to the observations that are available now but to all the consequences (potential observations) derivable from a theory.

Popper's own attempt to develop a logically consistent concept of verisimilitude is widely acknowledged to have been a failure. His theory, as set out in Popper (1963), was shown to be seriously flawed (Miller, 1974; Tichý, 1974). Attempts over many years to rescue the concept have been fraught with difficulty. Niiniluoto (1998) and Miller (2006, chapter 11) summarise 20-30 years of flawed attempts. Niiniluoto surveys several terms that have been used to describe various flavours of verisimilitude, with "expected" or "estimated" verisimilitude being measures of corroboration – measures of closeness to available observations – and with "theoretical" verisimilitude representing Popper's original, time-independent concept. Oddie and Cevolani (2022) and Niiniluoto et al. (2022) summarise more recent developments. The literature on verisimilitude is now rich; it includes a range of concepts, many of which depart substantially from Popper's original idea. It is beyond the scope of this paper to review these.

The concept of verisimilitude, as originally conceived by Popper consists of two components: accuracy and content. Accuracy is important because it represents consistency between predictions from theory and objective facts. Content is important because it guards against a theory being true (or close to true) whilst saying nothing or very little, with the limiting case being tautologies which, though true, have no empirical content.

A problem that has complicated the search for a viable theory of verisimilitude has been the problem of the "language-dependence" of closeness to truth: can a theory that is closer to the truth (relative to another theory) in one language be further from the truth in another? This problem, in the context of verisimilitude, was first identified by Miller (1974) – see below – and then generalised by Miller (1975), in which he suggested that the ranking of the accuracy of predictions from two theories (in terms of closeness to truth or to observation) could always be reversed by a suitable transformation into a different mathematical space. Following his original statement of the problem, Miller concluded that "… no false theory can … be closer to the truth than is another theory". Popper (1979, Appendix 2(5)) gave a simpler mathematical example for the same problem. Miller (2006, chapter 11) reviewed 30 years of work on this problem, and he was of the opinion that no solution was in sight.

If Miller's result is both correct and applicable to science, this is clearly a major problem, because it flies in the face of scientific practice and experience, in which one theory is apparently superseded by a "better" theory and "progress" appears to be made. As most people (scientists and non-scientists) generally accept that this is actually the case, this conflict represents a paradox. In Eyre (2024), we named this the "Miller-Popper paradox" and showed how it can be resolved. The resolution hinges on the importance of uncertainty in science. As we all learn (or should learn) at school, no measurement (observation) is complete without an estimate of its uncertainty. Moreover (and this we probably don't learn at school), predictions from theory are also subject to uncertainties. These arise either from inexactness in the theory (including probabilistic or indeterministic aspects), or from uncertainties and approximations in the predictive models that embody the theory, or from the initial conditions for the predictions, or from all three. Note that we are **not** here trying to account for errors (falsities) in the theory itself, but for errors that arise **even if** the theory is true. Uncertainties in the initial conditions are inevitable because they are based (ultimately) on observations, which are uncertain. In Eyre (2024), it was shown that the inclusion of finite uncertainty, in observations or predictions or both, resolves the paradox – the transformation into a different mathematical space preserves the ranking of the accuracy of predictions (at least for some important cases). Moreover, it was shown that Miller's result concerning the impossibility of ranking predictions is valid only in the limit of zero uncertainty, and that this limit, although an interesting mathematical puzzle, is not relevant for real scientific problems. It was also shown how some concepts important to science – concepts of “closeness”, “consistency”, “agreement” and related concepts – all rely on the concept of uncertainty.

With this in mind, in this paper we investigate the importance of uncertainty for the related problem of corroboration; we show how quantification of uncertainty, in observations and predictions, is central to the process of comparing predictions from theory with observations, and to judgements on whether theories are considered corroborated or falsified. We stress the empirical nature of estimates of uncertainty, and hence the time-dependence of judgements on corroboration/falsification.

We also offer some comments on concepts of verisimilitude. We note that verisimilitude is often approached as a problem in logic – within a framework in which a statement is either true or false. We suggest that, in the face of uncertainty in both observations and predictions, this framework is not the most fruitful. (This is not to say that a solution in logic cannot be found, but only that it is probably not the best place to start.)

Following his demonstration of the flaws in Popper original concept of verisimilitude, Tichý (1974) outlined an alternative approach for a quantitative theory of verisimilitude and gave a simple example. Miller (1974) criticised Tichý’s proposal and gave a counter-example. This generated a paradox which is named here the “Miller-Tichý paradox”. In the Appendix of this paper, we show in detail how this paradox can be resolved in a manner similar to the resolution of the Miller-Popper paradox.

The paper is structured as follows. Section 2 presents the process of corroboration in science with emphasis on the role of uncertainty. Section 3 examines the link between corroboration and verisimilitude, and it offers some comments on the consequences for some concepts of verisimilitude of the ideas on corroboration presented in section 2. It also discusses the consequences of the proposed resolution of the Miller-Tichý paradox. Section 4 presents further discussion of these ideas, and section 5 offers some conclusions.

2. Corroboration

For the purposes of this analysis, we accept the general Popperian framework; we assume that science progresses through an iterative process of corroboration and falsification, by means of “crucial experiments”. Some may consider this a rather naïve view of scientific progress, but we will adopt it for now, add the concept of uncertainty and see where it leads us. It is therefore instructive to look closely at the process of corroboration in this context. Some of the points in this section may appear rather pedestrian, but they are important for the later sections of the paper.

2.1 Corroboration of comparable theories

By comparable theories we mean theories with comparable consequences – theories that lead to the predictions of the same quantities. Let the n^{th} theory (TH^n) lead to predictions/forecasts, F^n, (denoted F for forecasts, to avoid confusion with P for probability). Let O be the observations used to falsify or corroborate these predictions. Let y_1^o be the value of the first observation and let y_1^n be the value of the prediction of this observation from the n^{th} theory. Note that the symbols F^n and O represent both the values of these measures **and** their uncertainties. This is illustrated in Fig.1 for predictions from three comparable theories.

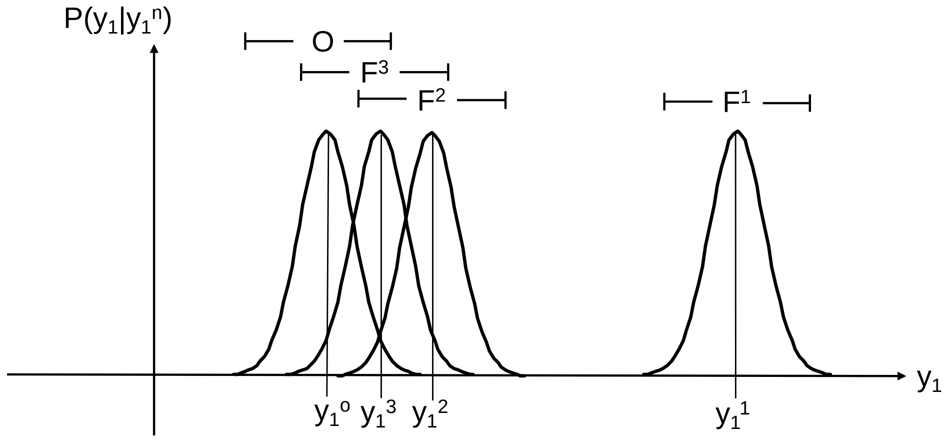

Figure 1. Illustrating the comparison of 3 predictions, F^1, F^2 and F^3, from theories TH^1, TH^2 and TH^3, with observation O, for a single observable quantity, y_1. The ordinate represents the uncertainty expressed as the conditional probability of the true value of y_1 given the observed or predicted value.

A zone of uncertainty is indicated around each predicted value and around the observed value. Initially, we consider all these uncertainties to be of equal magnitude and the associated errors to have a Gaussian distribution or similar shape. (Departures from these error distributions, including those of chaotic systems for which the error distributions of predictions depart substantially from these shapes, will be considered in section 2.5.) The true value of y_1 is unknown but (usually) lies close to y_1^o with a probability distribution as indicated by the ordinate in Fig.1, i.e. this curve should be interpreted as $P(y_1|y_1^o)$, the conditional probability that the true value is y_1 given the observed value y_1^o. Similarly for the predictions: assuming the theory is true, the conditional probability of the true value y_1 given the predicted value y_1^n is $P(y_1|y_1^n)$. For simplicity, we use y_1^o to represent a single observation, which it could indeed be. However, it could also represent the mean of a set of observations of the same type.

Normally, probability density functions (PDFs) of uncertainty express the conditional probability of the observation (or prediction) given the true value, e.g. $P(y_1^o|y_1)$, and Bayes theorem is invoked in order to calculate the posterior probability $P(y_1|y_1^o)$. However, if we assume that the prior probability of the true state is only weakly informative – in practice, that it is flat over the range in which $P(y_1^o|y_1)$ has significant value – then $P(y_1^o|y_1)$ and $P(y_1|y_1^o)$ have the same shape. (This assumption is not always valid – for example, when some physical limit creates a step in the prior probability – but we will examine first cases for which the assumption is valid.)

The usual interpretation of Fig.1 is as follows: the observation O corroborates both prediction F^2 and prediction F^3, because their zones of uncertainty overlap – they are “close” or “consistent”. (Strictly, O corroborates the theories leading to these predictions, but we will use “corroborates the prediction …” as shorthand for this.) On the other hand, the observation O falsifies F^1 – their values are outside each other’s zones of uncertainty and hence they are inconsistent. Note that this “falsification” is provisional, as the uncertainties in O or F^n (their magnitudes or their shapes) may subsequently be found to be substantially in error. Also, for the error distributions shown (Gaussian or similar), their overlap does not fall to zero and so, although the consistency of observation and prediction is very improbable, they are never absolutely inconsistent.

It is instructive to consider Fig.1 without estimated uncertainties. In this case it is only possible to say that y_1^2 is closer to y_1^o than is y_1^1, and that y_1^3 is closer to y_1^o than is y_1^2. Therefore, in this one-dimensional case, the theories can be ranked. However, **nothing can be said about the quantitative degree of corroboration or falsification of the theories.** Indeed, if all the uncertainties were substantially reduced, then all the theories would be considered falsified.

From the evidence presented in Fig.1, TH^2 and TH^3 survive falsification – they are corroborated – and we look for a "crucial experiment" through a new observation y_2^o to compare with y_2^n. This is illustrated in Fig.2, where we now show zones of uncertainty by circles. These circles are to be interpreted as representing equi-probability lines of 2D probability functions equivalent to those shown in Fig.1, with increasing radius indicating decreasing conditional probability. By drawing them as circles we imply that the errors in y_1 and y_2 are equal in magnitude and that they are uncorrelated between y_1 and y_2 (and we discuss other possibilities in section 2.4).

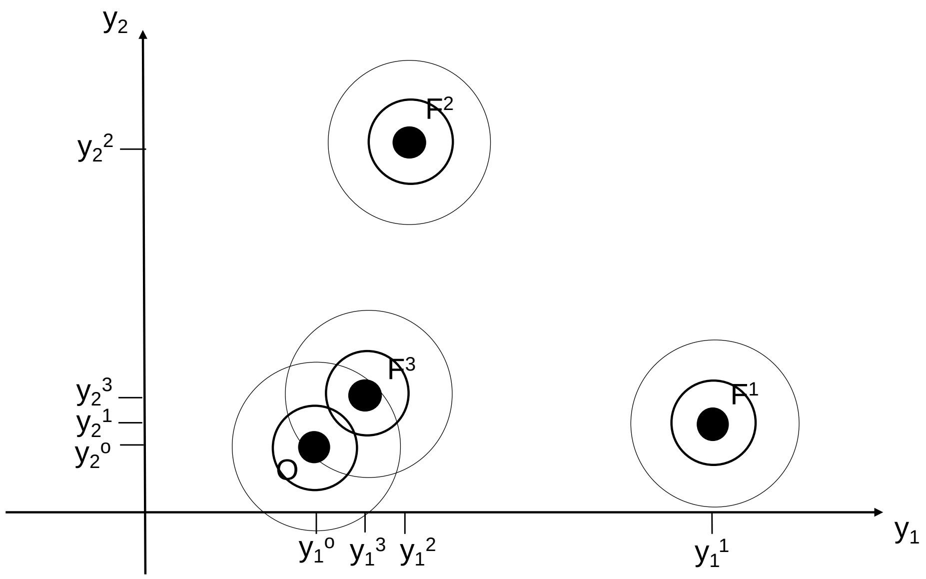

Figure 2. Illustrating the comparison of 3 predictions, F^1, F^2 and F^3, with observations O, for two observable quantities, y_1 and y_2.

The interpretation of Fig.2 is that F^2 is now falsified and F^3 still corroborated. It is important to note that the decision as to what amount of PDF-overlap constitutes corroboration is a procedural one. Fortunately, the decision is always tentative and can be revised in future in the light of new information.

We then seek another crucial experiment through a new observation y_3^o. This is illustrated in Fig.3, which may be interpreted as a 2D projection of predictions F^3 and F^4 and observations O on to the y_2-y_3 plane. Fig.3 shows that F^3 is now falsified whereas F^4 is corroborated. (We assume that F^4 has also survived comparisons with y_1^o.)

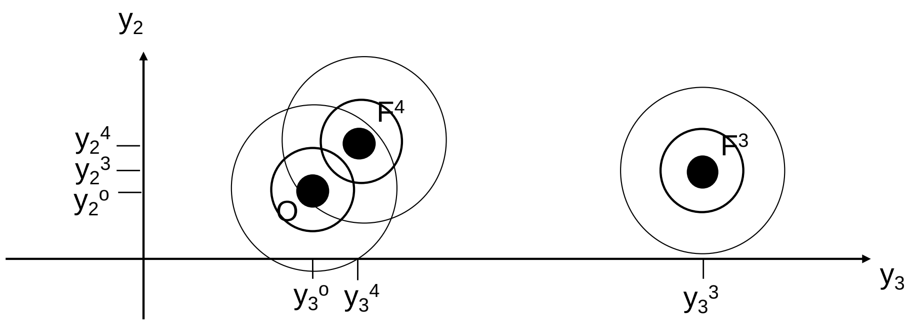

Figure 3. Illustrating the comparison of 2 predictions, F^3 and F^4, with observations O, for two observable quantities, y_3 and y_2.

This process continues iteratively; a theory can remain corroborated by all observations available today but can be falsified by a new observation tomorrow. When this happens, the search starts for a new theory. Note that the order in which theories can be falsified depends on the order in which observations become available; if y_2^o had become available before y_1^o, then F^2 would have been falsified before F^1. In practice the process of falsification is rarely as arbitrary as this suggests, because new experiments yielding new observations are often prompted by conflicts between existing theories.

We have focussed here on simple cases in which one theory gives predictions that are clearly closer to observation than another. In science, life is often more complicated; some observations agree better with predictions from one theory and other observations with those from another. Both theories therefore have shortcomings and judgment as to which is better must await further information. In this case, the choice between theories is a pragmatic one, concerning the applications for which the predictions will be used.

2.2 Corroboration of incomparable theories

By incomparable theories we mean theories with some consequences that cannot be compared – theories for which not all the observables predicted by one theory are predicted by another. It is important that some way can be found for assessing these theories relative to each other.

We suggest that all pairs of incomparable theories can be made comparable as follows: a theory that is incapable of generating a given prediction is equivalent to a theory that can make such a prediction but with a very large (or infinite) uncertainty. This is illustrated in Fig.4 (which is a modified version of Fig.2). The (almost) parallel lines around F^2 can be considered as parts of ellipses of very large length in the y_2 direction in the vicinity of O. In this case, F^2 would also be corroborated by the observations but not as strongly as F^3; the area of the ellipse around F^2 is very much larger than that around F^3. Therefore TH^2 is comparatively a very weak theory and TH^3 is preferred. If Fig.4 had been drawn such that F^3 were falsified, then only F^2 would remain corroborated, but only weakly – the search would be on for a new theory to be tested against y_2^o.

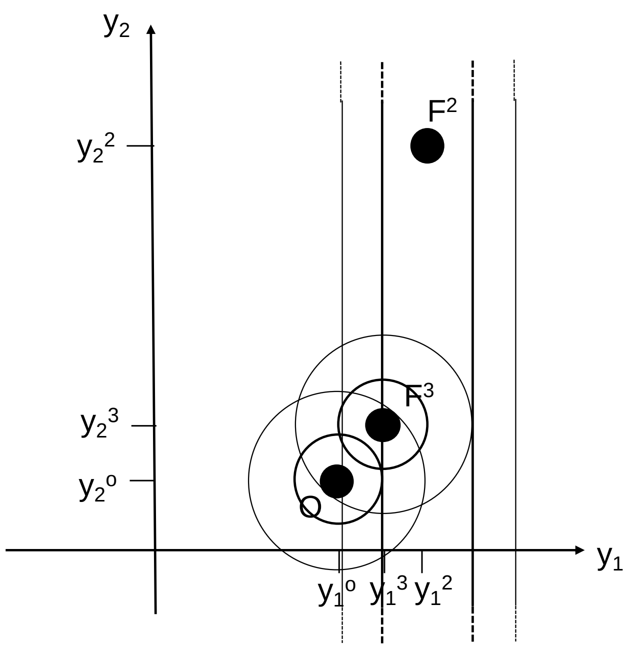

Figure 4. As Fig.2, except with uncertainties of F² increased to very large values in the y_2 direction.

2.3 Corroboration by incomplete observations

An observation that is not available is equivalent to an observation that is available but with a very large (or infinite) uncertainty. This is illustrated in Fig.5. F^2 and F^3 are both corroborated by y_1^o. They differ significantly in their prediction of y_2^o, but this is not yet available and so both remain corroborated for the time being. Note that, as observations of lower and lower uncertainty become available, there will come a point at which y_2^o corroborates F^2 or F^3 or neither, but not both.

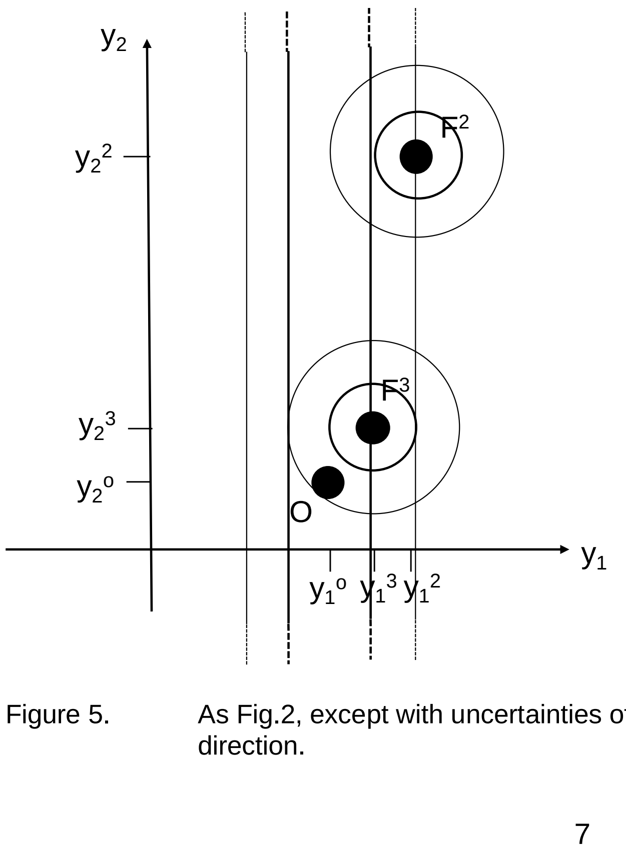

Figure 5. As Fig.2, except with uncertainties of O increased to very large values in the y_2 direction.

2.4 Corroboration with varying uncertainties

The degree to which predictions are corroborated or falsified depends on the uncertainties in observations and predictions – both their magnitudes and their correlations. This is illustrated in Fig.6. In Fig.6a, both predictions F^5 and F^6 are falsified to a similar degree. However, they could be falsified more strongly if observation uncertainties were reduced, or they could corroborated if these uncertainties were increased. The correlations of uncertainty between different observations or predictions are also important, as illustrated in Fig.6b. Positively correlated uncertainties, as drawn, will now lead to F^5 being corroborated but not F^6. Note that there is no change here in the predicted values, only in their estimated uncertainties.

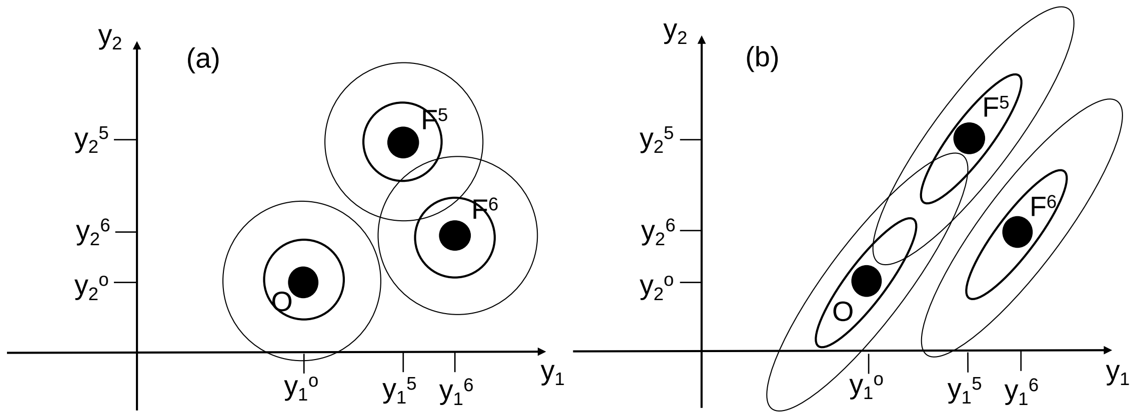

Figure 6. Illustrating the comparison of 2 predictions, F^5 and F^6, with observations O, for two observed quantities, y_1 and y_2, (a) for equal and uncorrelated errors in observations and predictions and (b) for equal and positively correlated errors.

2.5 Corroboration for predictions with different uncertainty distributions

In the sections above, we have assumed that observations and predictions have error distributions that are either Gaussian or similar to Gaussian, i.e. that they are unimodal, that their error PDFs decrease monotonically away from a maximum value, and that they never fall to zero. Firstly, consider the third of these: if the error PDFs of both observation and prediction fall to zero beyond a certain distance, then beyond a certain separation of observed and predicted value the probability of their being consistent falls to zero. This is the region in which the condition identified by Miller (1975) is valid, i.e. no prediction is better than any other prediction, whatever their separation from observations. However, this condition is non-physical – it indicates that at least one of the PDFs has been mis-specified. In science the opposite case is more common, i.e. the tails of the distributions are often super-Gaussian. For example, observations may consist of “good” observations (having small errors with Gaussian or other “well-behaved” PDFs) and “bad” observations (having much larger errors than normal). Such situations are common and call for methods of quality control, to separate “good” from “bad”. This can be achieved by specifying an appropriate total PDF, such as a Gaussian plus a constant (for example, see Lorenc and Hammon, 1988). Non-Gaussian PDFs give rise to non-quadratic penalty functions; the PDF of uncertainty determines the appropriate objective penalty function (see Eyre, 2024).

Another common problem arises in nonlinear systems. Here prediction errors may initially be unimodal and close to Gaussian but, as the state of the nonlinear system evolves, the error distributions become highly non-Gaussian and multi-modal. In chaotic systems the prediction errors will grow to populate an attractor (e.g. see Palmer 2022).

In these cases we must abandon the comparison between individual predictions and "verifying" observations and instead consider the statistics of an ensemble of predictions. Now a theory can be considered corroborated if it leads to an ensemble of predictions with statistical characteristics close to those of "verifying" observations – for example, that the spread of the ensemble of predictions is close to the spread of the "errors" (difference from observation) of the verified predictions. In this case more complex metrics are needed to measure "closeness" (but this will not be discussed further in this paper).

3. Corroboration, verisimilitude and the language-dependence of closeness to truth

As we have seen in section 2, quantitative degrees of corroboration depend on the values and the estimates of uncertainty of both predictions and observations. (For the mathematics of the quantification, see Eyre (2024).) Quantitative estimates of uncertainty have the same empirical status as the values to which they apply – both are inter-subjectively shareable and criticisable, and thus they have the status of objective information. Estimates of uncertainty are, however, time-dependent – they can be revised as experimental evidence is reassessed and they can be changed (usually reduced) as a result of improved observational technology and improved ability to generate predictions from their underlying theories. It could be argued that the values themselves may also be time-dependent; they might occasionally be revised. Or, indeed, it could be argued that the observation/prediction should always be considered as a value **and** its uncertainty, as stressed in section 2.1. In either case, we suggest that the process of corroboration should start again with these new values. However, the key point here concerns the importance of uncertainty to the measure of the "closeness" of prediction to observation, and hence the degree of corroboration, even if the values are unchanged.

Popper's original idea was that corroboration is time-dependent, as results from new crucial experiments become available, but that verisimilitude is time-independent. This follows from the notion that verisimilitude can be equated with "ultimate" or "theoretical" corroboration – with testing all the consequences of a theory. Although verisimilitude in this sense has always been impractical – a goal to which one might aspire, or a normative concept – the analysis of corroboration presented above suggests that it is a flawed concept in principle, because any measure of "closeness" relevant to science must rely on measures of uncertainty. There is no absolute standard of uncertainty – it is intrinsically a time-dependent concept – and so this criticism will apply to any time-independent concept of verisimilitude. It is also suggested that time-dependent measures of verisimilitude will not be relevant to science unless they account for uncertainty. This is a tentative conclusion; a stronger statement would require a more thorough review of the literature on different flavours of verisimilitude, which is beyond the scope of this paper.

The process developed in section 2 could be described as one of "iterative corroboration" and would belong to the family of measures of "estimated verisimilitude" as described by Niiniluoto (1998). Through this process, we continually find "better" theories, in that they are better corroborated, i.e. consistent with observations to within estimated uncertainties. As more crucial experiments are performed, the dimensions of the space in which a theory is corroborated increases and/or the hypervolume of the uncertainty within this space decreases.

The last point would appear to address Popper's concern about theory content, i.e. that accuracy alone is not an adequate measure, because it may be possible to increase accuracy at the expense of content (e.g. through theories which, in the limit, express a set of tautologies). This can be avoided through the above concept of a hypervolume of uncertainty; as theories improve (become stronger or "bolder") and yield more and more accurate predictions, the hypervolume of uncertainty is reduced. The magnitude of the hypervolume of uncertainty is related to the information content of a prediction (or of an observation). According to Shannon's information theory (Shannon and Weaver, 1963), the information content of an observation is equal to the reduction in the entropy of the associated PDF and, if the PDF is Gaussian, this is equal to the logarithm of the ratios of the hypervolumes of uncertainty before and after the observation. These ideas are discussed in more detailed in the literature of the geosciences; see, for example, Rodgers (1976) or Rodgers (2000).

We accept that the history of science suggests that most theories are probably false and also Popper's assertion that all universal theories are, *a priori*, probably false. However, the claim that no false theory can yield predictions that are closer to the truth that any other false theory can be contested. In Eyre (2024), it was shown how this claim, termed the "Miller-Popper paradox", can be resolved; the transformation of predictions and observations into a different mathematical space preserves the ranking of theories in terms of closeness to observation, because the associated error distributions are also transformed by the projection into the new space and in such a way as to preserve objective measures of closeness. As introduced in section 1, it can also be shown (see Appendix) that the paradox arising from Miller's objection (1974) to Tichý's attempt (1974) to rescue Popper's theory of verisimilitude is an example of the same problem of projection between spaces and can be resolved in the same way.

So, if Miller's objection can be rejected in this way, what are we to make of Tichý's proposal for a new theory of verisimilitude? We show in the Appendix that Tichý's proposal represents a strategy in logic that does not map on to a real problem in science, because it does not consider uncertainties. When uncertainties are taken into account, we find the same problem with a concept of (time-independent) verisimilitude as discussed above. So, can one false theory yield better predictions than another, as Tichý's example attempted to illustrate? We suggest that it can, but using the process of iterative corroboration described in section 2. In this way, the proposed solution to the Miller-Tichý paradox is relevant not only to Popperian-inspired measures of verisimilitude, but also to all other kinds of comparisons of uncertain measurements and predictions. It therefore supports claims for the language-independence of acceptable measures of closeness to truth.

4. **Discussion**

The analysis in sections 2 and 3 has stayed firmly in multi-dimensional observation space; the concepts of corroboration and verisimilitude have been discussed in this space, as have the measures through which these concepts may be quantified. All this relies on the concept of an "observation operator", i.e. a function or procedure for mapping quantities from variables represented by theories (from theory space) to quantities that can be compared with observations (to observation space). Predictions are compared with theories in observation space, not in theory space, and likewise "comparability" is described in observation space. The process of iterative corroboration allows us to get "closer to truth" in observation space, i.e. in terms of the consequences of theories, rather than of the theories themselves.

All observations and their operators are theory-laden. An observation operator may be considered as a set of theories auxiliary to the theory under test. However, the process of corroboration relies on a large degree of independence between the theories underlying the observations and those underlying the predictions. Whilst science strives to achieve this independence, it is not always fully possible. The practical consequence is some degree of correlation between uncertainties in predictions and in the corroborating observations. Scientists need to be aware of this (and they generally are).

Another consequence of the theory-laden nature of observations is that theory-change can lead to a change in the interpretation of observations. We assumed in section 2 that, when a new theory comes along, its predictions can be compared with old observations, as well as with new ones. However, old observations may need to be reinterpreted in the light of new theory. This complicates the process of iterative corroboration outlined in section 2 but, we suggest, does not invalidate it.

Central to the concept of uncertainty is the concept of truth, i.e. the idea that there is a "true value" of the observed quantity around which the observations (with their uncertainties) are clustered (but see below). This amounts to "realism" in observation space, which we accept. (We also postulate that these true values in observation space are the result of a true state in a theory space corresponding to the real world, but this assumption is not necessary for the arguments in this paper.)

Whilst observations are normally clustered around the truth, this is not the case when systematic errors are greater than random errors. In this case, the observation cluster will be offset from the truth, and this is well understood in science. We note the comment by Hacking (1982) on this: "Although the idea of systematic error presents interesting conceptual problems, it seems to be unknown to philosophers".

The figures in section 2 have been drawn on the assumption that truth is point-like – that the true value is a point in a multi-dimensional observation space. It is plausible that the analysis presented here could be extended to cover either an indeterministic or probabilistic notion of truth – that, within the framework proposed here, this possibility could be included in the description of uncertainty – but we do not pursue it further here.

In this paper, we stress that uncertainty in observations and predictions is fundamental and unavoidable, both in science itself and in any coherent theory of corroboration. Why, then, has uncertainty been largely ignored in most previous work in this field? We suggest that it is because uncertainty has been, or has been perceived to be, very low or unimportant in most of the key experiments discussed in the philosophy of science - that the agreement or otherwise between observation and prediction has been obvious without a detailed analysis of their uncertainties. However, we hope it is clear from the analysis presented in this paper that the neglect of uncertainty is a mistake, because corroboration relies on "consistency" – a prediction cannot be declared consistent with an observation without some implicit assumption concerning their uncertainties.

It could be objected that the original concept of verisimilitude was a postulated measure of closeness to truth, whereas we have focussed here on closeness to observations. However, as we have demonstrated, measures of "closeness" relevant to science involve estimates of uncertainty; in the limit of zero uncertainty, these measures lose their scientific relevance. In other words, if we wish to say that the value of a predicted or observed quantity is **close** to the true value, then we imply values of uncertainty in the predicted or observed value. This is not to deny that true values exist, but only to stress that measures of closeness to them involve estimates of uncertainty.

As noted above, we have stayed firmly in observation space – we have identified corroboration with the claim that predictions from a theory are consistent with observations. In this way we have stayed close to the methods of science as described numerous times by Popper (e.g. see Popper, 1979, section 23 and 24) and with his view that rational choice between competing theories can be made using this approach. We have stayed away from ideas of “closeness to truth in theory-space” and whatever this might mean. We note that it is possible to make radical changes to theories, even at a fundamental level, whilst making relatively minor changes to some of the observable consequences (but perhaps radical changes to others). In this context and as an example, we note the radical theory change from classical to quantum physics. We also note the quotation in Palmer (2022) from Penrose (1997): “My own view is that to understand quantum nonlocality we shall require a radical new theory. This theory will not just be a slight modification of quantum mechanics but something as different from standard quantum mechanics as general relativity is from Newtonian gravity. It would have to be something which has a completely different conceptual framework.”

We have focussed in this paper on quantitative theories, and it might be asked to what extent these arguments apply also to qualitative theories. We have not considered these in detail. However, it seems plausible that the inevitable vagueness inherent in a qualitative prediction must have an effect similar to an uncertainty in a quantitative prediction. The discussion of the Miller-Tichý paradox (Appendix) illustrates this.

5. **Conclusions**

Popper stressed the distinction between knowledge and certainty and thus highlighted the importance of uncertainty, in knowledge in general and in scientific knowledge in particular. Scientists also follow this maxim by acknowledging that all observations come with uncertainty, as do all predictions. In this paper, we have attempted to follow through on both these central ideas in a consistent way; we have illustrated the pervasive role of uncertainty in the corroboration of theories, by comparison of their predictions with observations. We have noted that decisions concerning corroboration or falsification are always provisional, and that they require some convention concerning the magnitude of the overlap of relevant uncertainties.

We have also examined the concept of “theoretical” or time-independent verisimilitude from this perspective and found it wanting in terms of its relevance to science, because estimates of uncertainty are inherently time-dependent. This criticism is likely to apply also to concepts of “expected” or ”estimated” verisimilitude, if they too fail to account for uncertainty in observations and predictions.

In Eyre (2024), we showed that introduction of uncertainty resolves the problem of language-dependence in the Miller-Popper paradox. Here we have shown that the same approach can be used to resolve the Miller-Tichý paradox (the “minnesotan-arizonan paradox”), which arose out of early work on verisimilitude.

We have conducted this analysis in observation space – in the space of the consequences of theories rather than the space of theory variables themselves. We contend that this is how science works and how it accommodates radical theory change. In summary, we hope we have put Popperian theory of falsification and corroboration on a firmer footing. Key to this analysis is the insistence that an observation or a prediction is not simply a value – it is value together with an estimate of its uncertainty .

Appendix. **Resolution of the “Miller-Tichý paradox**”

A.1. **The paradox**

In a paper discussing flaws in Popper's concept of verisimilitude, Tichý (1974) introduced a "rudimentary weather-related language" with only 3 primitive sentences equivalent to: "it is hot" $(= h)$, "it is raining" $(= r)$ and "it is windy" $(= w)$. Within this language it is possible to construct sentences such as it is hot, raining and windy $(h\&r\&w)$, or it is hot, not raining and not windy $(h\&{\sim}r\&{\sim}w)$. If the true state is $h\&r\&w$, Tichý contended that a prediction, $h\&r\&{\sim}w$, was clearly closer to the truth than a prediction, ${\sim}h\&{\sim}r\&{\sim}w$. Tichý proposed that a different theory of verisimilitude could be built on closeness measures such as this.

In a paper also showing flaws in Popper's concept, Miller (1974) included a response to Tichý's proposal in which he contested Tichý's statement about the ranking of theories in this language. He proposed a translation of Tichý's hot-rainy-windy language into a logically equivalent language using 3 different primitive sentences: "it is hot" $(= h)$, "it is minnesotan" $(= m)$, and "it is arizonan" $(= a)$, in which "minnesotan" means $(h\&r.\mathrm{OR}.{\sim}h\&{\sim}r)$ and "arizonan" means $(h\&w.\mathrm{OR}.{\sim}h\&{\sim}w)$. Within this language the ranking of closeness to truth in Tichý's language does not hold, as demonstrated in Table 1, where the two sentences in each row are logically equivalent, and where "distance" is the distance from truth, i.e. the number of primitive sentences that are false in each compound sentence.

Tichý's language		Miller's language	
	distance		distance
$h\&r\&w$	0	$h\&m\&a$	0
$h\&r\&{\sim}w$	1	$h\&m\&{\sim}a$	1
$h\&{\sim}r\&w$	1	$h\&{\sim}m\&a$	1
$h\&{\sim}r\&{\sim}w$	2	$h\&{\sim}m\&{\sim}a$	2
${\sim}h\&r\&w$	**1**	${\sim}h\&{\sim}m\&{\sim}a$	**3**
${\sim}h\&r\&{\sim}w$	2	${\sim}h\&{\sim}m\&a$	2
${\sim}h\&{\sim}r\&w$	2	${\sim}h\&m\&{\sim}a$	2
${\sim}h\&{\sim}r\&{\sim}w$	**3**	${\sim}h\&m\&a$	**1**

Table 1. Tichy's and Miller's weather-related languages.

It can be seen that the ranking of the distances are reversed for two of the compound sentences in the table. On this basis, Miller concluded that Tichý's proposal was also flawed. Miller's result is paradoxical, as it conflicts with the ranking that Tichý (and others) claimed to be intuitively obvious.

A.2. **Proposed resolution of the paradox**

Although Tichý's language is very simple and uses weather-related terms, it differs substantially from the language used in the science of meteorology; when told it is "hot", "rainy" or "windy", a meteorologist would ask: how hot, how rainy, how windy? Quantitative measures, in this case temperature and rainfall rate (or rainfall amount) and wind speed, would be used. Let us use the same symbols – h, r and w – but let them now stand for the quantitative departure of each of the continuous meteorological variables from its mean value. Also let us use scaled variables such that for a typical hot day, $h = 1$, and for a typical cold (not-hot) day, $h = -1$, and similarly for r and w.

Next we need to create continuous variables to represent the concepts "minnesotan" and "arizonan". There are many ways in which this might be done, but let us choose a very simple one:

$$m = hr \text{ and } a = hw \,. \quad (1)$$

Then the corresponding table is:

Variables h, r, w		Variables h, m, a	
	distance		distance
$h = 1, r = 1, w = 1$	0	$h = 1, m = 1, a = 1$	0
$h = 1, r = 1, w = -1$	1	$h = 1, m = 1, a = -1$	1
$h = 1, r = -1, w = 1$	1	$h = 1, m = -1, a = 1$	1
$h = 1, r = -1, w = -1$	2	$h = 1, m = -1, a = -1$	2
$h = -1, r = 1, w = 1$	**1**	$h = -1, m = -1, a = -1$	**3**
$h = -1, r = 1, w = -1$	2	$h = -1, m = -1, a = 1$	2
$h = -1, r = -1, w = 1$	2	$h = -1, m = 1, a = -1$	2
$h = -1, r = -1, w = -1$	**3**	$h = -1, m = 1, a = 1$	**1**

Table 2. Conversion of Tichý's and Miller's weather-related languages into values for continuous variables.

Note that the relationship of (h, m, a) to (h, r, w) in Table 2 mirrors exactly their relationship in Table 1.

We now recognise that observations or predictions of these variables are uncertain, and we need to understand how an uncertainty in (h, r, w) transforms into an uncertainty in (h, m, a). Let us introduce a vector-matrix notation in which

$$\mathbf{x} = (h, r, w) \text{ and } \mathbf{p} = (h, m, a) \,. \quad (2)$$

$\mathbf{x}$ and $\mathbf{p}$ are related through

$$\mathbf{p} = B(\mathbf{x}) \quad (3)$$

where $B(\ldots)$ is a nonlinear matrix function representing the relations in (1).

Let the errors in observations or predictions of these quantities be $\boldsymbol{\varepsilon}_x$ and $\boldsymbol{\varepsilon}_p$ respectively. This is the same notation as used in Eyre (2024), in which the Appendix discusses nonlinear problems such as this one.

Errors in $\mathbf{x}$ and $\mathbf{p}$ are related by

$$\boldsymbol{\varepsilon}_p \approx \mathbf{B}(\mathbf{x})\boldsymbol{\varepsilon}_x \,, \quad (4)$$

where $\mathbf{B}(\mathbf{x}) = \nabla_{\mathbf{x}} B(\mathbf{x})$. The relation in (4) becomes exact as $\boldsymbol{\varepsilon}_x \to \mathbf{0}$ and $\boldsymbol{\varepsilon}_p \to \mathbf{0}$.

As shown in Eyre (2024), the corresponding transformation of the associated error covariances, $\mathbf{C}_x(\mathbf{x})$ and $\mathbf{C}_p(\mathbf{p})$, is given by

$$\mathbf{C}_p(\mathbf{p}) \approx \mathbf{B}(\mathbf{x})\mathbf{C}_x(\mathbf{x})\mathbf{B}(\mathbf{x})^T \,, \quad (5)$$

where ^{T} denotes matrix transpose. $\mathbf{B}(\mathbf{x})$ is obtained by differentiation of equations (1) with respect to the elements of $\mathbf{x}$:

$$\mathbf{B}(\mathbf{x}) = \nabla_{\mathbf{x}} B(\mathbf{x}) = \begin{bmatrix} 1 & 0 & 0 \\ r & h & 0 \\ w & 0 & h \end{bmatrix}. \quad (6)$$

If the errors in $\mathbf{x}$-space are equal and uncorrelated, then

$$\mathbf{C}_x = k^2 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{7}$$

where k is a constant giving the magnitude of the expected error. Then, from (5),

$$\mathbf{C}_p \approx k^2 \begin{bmatrix} 1 & r & w \\ r & r^2 + h^2 & rw \\ w & rw & w^2 + h^2 \end{bmatrix}. \tag{8}$$

This shows that the errors in $\mathbf{p}$-space are correlated between variables, except for the points at which $r = w = 0$. For the value of $\mathbf{x}$ at the first row on Table 2 ($h = 1, r = 1, w = 1$), the values of $\mathbf{C}_p$ and $\mathbf{C}_p{}^{-1}$ are:

$$\mathbf{C}_p \approx k^2 \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}, \ \mathbf{C}_p{}^{-1} \approx \frac{1}{k^2} \begin{bmatrix} 3 & -1 & -1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}, \tag{9}$$

where $^{-1}$ denotes matrix inverse.

For the values at the other rows in Table 2, the diagonal elements of $\mathbf{C}_p$ are the same and the off-diagonals take values of either +1 or -1. For $\mathbf{C}_p{}^{-1}$, the values that are -1 for the first row take values of either +1 or -1 for other rows.

When visualised geometrically, an error covariance that is spherical in the $\mathbf{x}$-space becomes ellipsoidal when transformed into $\mathbf{p}$-space.

Let us now return to Tichý's problem (Table 1 or 2) in which the first row represents the true values and the other rows represent predictions with different levels of accuracy. If we now regard the first row as representing an observation (which, in Tichý's example, contains zero error) and the other rows as predictions (all containing some degree of error), then we can use the result obtained in Eyre (2024), i.e. that the transformation from $\mathbf{x}$-space to $\mathbf{p}$-space (i.e. to the minnesotan-arizonan weather representation) makes no difference to the probability that the observation is consistent with any prediction, and that the ranking of predictions according to accuracy (i.e. misfit to truth) is preserved by the transformation. This is confirmed in Table 3, which shows the "penalty" (i.e. quantified misfit of prediction to truth) for each row in Table 2, in both $\mathbf{x}$-space and $\mathbf{p}$-space. Here we have used the term representing the misfit of prediction to truth from Eyre (2024, eq.3.10), noting that the transformation preserves not only the total penalty of misfit to observation and prediction but also each of its component parts, i.e. misfit to observation and misfit to prediction (and also noting that, in Tichý's example, the misfit of observation to truth is zero.)

Variables h, r, w			Variables h, m, a		
$\mathbf{x}^T=[h, r, w]$	$\boldsymbol{\delta}\mathbf{x}^T=\mathbf{x}^T\text{-}\mathbf{x}_o{}^T$	Penalty = $\boldsymbol{\delta}\mathbf{x}^T\mathbf{C}_x{}^{-1}\boldsymbol{\delta}\mathbf{x} \,/\, k^2$	$\mathbf{p}^T=[h, m, a]$	$\boldsymbol{\delta}\mathbf{p}^T=\mathbf{p}^T\text{-}\mathbf{p}_o{}^T$	Penalty = $\boldsymbol{\delta}\mathbf{p}^T\mathbf{C}_p{}^{-1}\boldsymbol{\delta}\mathbf{p} \,/\, k^2$
1,1,1	0,0,0	0	1,1,1	0,0,0	0
1,1, −1	0,0, −2	4	1, 1, −1	0,0, −2	4
1, −1, 1	0, −2,0	4	1, −1,1	0, −2,0	4
1, −1, −1	0, −2, −2	8	1, −1, −1	0, −2, −2	8
−1,1,1	−2,0,0	4	−1, −1, −1	−2, −2, −2	4

$-1,1,-1$	$-2,0,-2$	8	$-1,-1,1$	$-2,-2,0$	8
$-1,-1,1$	$-2,-2,0$	8	$-1,1,-1$	$-2,0,-2$	8
$-1,-1,-1$	$-2,-2,-2$	12	$-1,1,1$	$-2,0,0$	12

Table 3. Penalty of prediction for Tichý's variables (3 left columns) and Miller's variables (3 right columns). $\mathbf{x}_o$ and $\mathbf{p}_o$ are the true values given in the first row.

This is the proposed resolution of the paradox. This result also demonstrates that "closeness to truth" depends on the magnitudes and correlations of errors in the different variables.

The example above assumes that the uncertainties in h, r and w are equal and uncorrelated. It is possible to reverse the result if $\mathbf{C}_x$ is chosen differently; from (5), we can see that, if $\mathbf{C}_p$ is a unit matrix, then $\mathbf{C}_x = \mathbf{B}(\mathbf{x})^{-1}\mathbf{B}(\mathbf{x})^{T^{-1}}$. In this limit the ranking of predictions using (h, m, a)-language in Table 1 would be correct.

Of course, the conversion of the problem from discrete, logical quantities to continuous variables is arbitrary; other functions could have been chosen to achieve the same result. However, any function with the desired properties (i.e. properties that simulate the original Miller-Tichý problem but including the attributes of uncertainty required to make it a scientific problem) could be used to replace (1), thus yielding a different value of $\mathbf{B}(\mathbf{x})$ in place of (6). But the transformation of error covariances given by (5) is general, and the ranking of accuracies of predictions to which it leads is preserved for differentiable, non-singular transformations.

Even if we were to concede that Tichý's language of discrete in h,r,w variables was acceptable as a scientific language, we would still encounter the problem of uncertainty in both observations and predictions, and these would be particularly large close to the boundary between hot and not-hot, etc. The PDFs of uncertainty would be more complex than those in the problem of continuous variables but, we suggest, they would have similar properties when transformed into a different space.

Acknowledgements

I am very grateful to Dr Antony Galton and Dr Mark Webb for their comments, encouragement and advice. I am also grateful to two anonymous reviewers for the helpful comments.

References

Eyre, John (2024). Resolution of the Miller-Popper paradox. Synthese, **203**, 1 (2024). https://doi.org/10.1007/s11229-023-04425-7

Hacking, Ian (1982). Experimentation and scientific realism. *Philosophical Topics*, **13**, 154-172.

Lorenc, A.C., and Hammon, O. (1988). Objective quality control of observations using Bayesian methods. Theory, and a practical implementation**.** *Quarterly Journal of the Royal Meteorological Society*, **114**, 515-543**.** https://doi.org/10.1002/qj.49711448012

Miller, David (1974). On Popper's quantitative theory of verisimilitude. *British Journal for the Philosophy of Science*, **25**, 166-177. https://doi.org/10.1093/bjps/25.2.166

Miller, David (1975). The accuracy of predictions. *Synthese*, **30**, 159-191. https://doi.org/10.1007/BF00485304

Miller, David (2006). *Out of error: further essays on critical rationalism*. London: Routledge. https://doi.org/10.4324/9781315247465

Niiniluoto, Ilkka (1998). Verisimilitude: the third period. *British Journal for the Philosophy of Science*, **49**, 1-29. https://www.jstor.org/stable/i227852

Niiniluoto, I., Cevolani, G. and Kuipers, T. (2022). Approaching probabilistic truths: introduction to the Topical Collection. *Synthese*, ***200***, 1139. https://doi.org/10.1007/s11229-022-03516-1

Oddie, Graham, and Cevolani, Gustavo (2022). Truthlikeness. *The Stanford Encyclopedia of Philosophy* (Winter 2022 Edition), Edward N. Zalta & Uri Nodelman (eds.). https://plato.stanford.edu/archives/win2022/entries/truthlikeness/

Palmer, Tim (2022). *The primacy of doubt*. Oxford: Oxford University Press.

Penrose, R. (1997). *The large, the small and the human mind.* Cambridge: Cambridge University Press.

Popper, Karl (1959, 1968). *The Logic of Scientific Discovery*. London and New York: Routledge (Second edition 1968).

Popper, Karl (1963, 1965, 1969). *Conjectures and Refutations*. London and New York: Routledge (Third edition 1969)

Popper, Karl R. (1972, 1979). *Objective knowledge: an evolutionary approach*. Oxford: Clarendon Press. (Second edition 1979).

Rodgers, C. D. (1976). Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation. *Reviews of Geophysics and Space Physics*: **14**, 609-624. https://doi.org/10.1029/RG014i004p00609

Rodgers C.D. (2000). Inverse methods for atmospheric sounding: theory and practice. World Scientific. Series on Atmospheric, Oceanic and Planetary Physics: Volume 2. https://doi.org/10.1142/3171

Shannon, Claude E., and Weaver, Warren (1963). *The mathematical theory of communication*. Urbana and Chicago: University of Illinois Press.

Tichý, Pavel (1974). On Popper's definitions of verisimilitude. *British Journal for the Philosophy of Science*, **25**, 155-60. https://doi.org/10.1093/bjps/25.2.155