4 Nov 2023
John Eyre

Resolution of the Miller-Popper paradox

Abstract

A longstanding paradox was first reported by David Miller in 1975 and highlighted by Karl Popper in 1979. Miller showed that the ranking of predictions from two theories, in terms of closeness to observation, appears to be reversed when the problem is transformed into a different mathematical space. He concluded that "… no false theory can … be closer to the truth than is another theory". This flies in the face of normal scientific practice and is thus paradoxical; it is named here the "Miller-Popper paradox".

This paper proposes a resolution of the paradox, through consideration of the inevitable errors and uncertainties in both observations and predictions. It is proved that, for linear transformations and Gaussian error distributions, the transformation between spaces creates no change in quantitative measures of "closeness-to-observation" when these measures are based in probability theory. The extension of this result to nonlinear transformations and to non-Gaussian error distributions is also discussed.

These results demonstrate that concepts used in comparison of predictions with observations – concepts of "closeness", "consistency", "agreement", "falsification", etc. – all imply some knowledge of the uncertainty characteristics of both predictions and observations.

1. Introduction

Under the heading "Crucial experiments in physics", Karl Popper (1979, Appendix 2, Section 5, pp.372-374) presented a result, which he credited to David Miller, addressing the question: how to make a rational choice between competing theories through comparison of their predictions with observations. Miller's result suggested the futility of any ranking of two false theories based on comparison of their predictions with observations, because a transformation of the problem into a different mathematical space could lead to the opposite ranking. Following his original statement of the problem, Miller (1975a) concluded that "… no false theory can … be closer to the truth than is another theory".

Concerning this result, Popper wrote "Frege, when faced with Russell's paradox, said 'Arithmetic is tottering'. One is tempted to say that Miller's result shows that physics is tottering." However, he added that we had very solid reasons to think that physics was not tottering and that Miller's result could not be accepted, but he did not know of a general method by which the problem raised by Miller could be solved or avoided. In this sense, the result is paradoxical: here it is called "the Miller-Popper paradox". It should not be confused with the Popper-Miller Theorem (Popper and Miller, 1983) or Miller's Paradox of Information (Miller, 1966).

Miller (2006, chapter 11) summarised discussion of and further work on the problem in the 30 years following its first statement. He included examples of linear transformations (as in the original work) and also some nonlinear transformations. Discussion of these developments is included in sections 5 and 6 of this paper. More recently, David Miller has advised that he is not aware that the paradox has been successfully resolved (personal communication, 2022).

The purpose of this paper is to propose a resolution of this paradox, by considering the role of inevitable uncertainties in both observations and predictions. Section 2 restates the paradox in the form presented by Popper (and using the same example). Section 3 offers a resolution of the paradox, for the case of a linear transformation in which error distributions take a Gaussian form. It is proved that, in this case, the consistency of prediction with observation, as quantified using standard probability theory, is not changed by the transformation between mathematical spaces. Section 4 presents some illustrative examples, starting with the one originally presented by Popper. Section 5 sketches the extension of these results to nonlinear transformations and to error distributions with more general probability density functions (PDFs). Section 6 discusses these results and their implications, and Section 7 presents some conclusions.

2. The paradox

Popper (1979) presented the problem as follows:

"… if a false theory T_1 yields better metrical approximations than theory T_2 to the true values (or else to the measured values) of at least two parameters, one can always transform these theories into logically equivalent theories which must be given the opposite ranking with respect to another set of parameters, definable in terms of the first set (and the definability is mutual)."

He illustrated it with the following example. Two sets (i) and (ii) of two equations are mutually deducible and therefore equivalent:

$$\begin{aligned} &\text{(i)} \quad x = q - 2p \quad \text{and} \quad y = 2q - 3p \ , \\ &\text{(ii)} \quad p = y - 2x \quad \text{and} \quad q = 2y - 3x \ . \end{aligned} \tag{2.1}$$

Let us assume that the true or measured values are $\{x = 0,\ y = 1\}$ and hence $\{p = 1, q = 2\}$. Let theory T_1 entail prediction P_1: $\{x = 0.100,\ y = 1.000\}$ and hence $\{p = 0.800, q = 1.700\}$, and let theory T_2 entail prediction P_2: $\{x = 0.150,\ y = 1.225\}$ and hence $\{p = 0.925, q = 2.000\}$. These results are presented in Table 1, where O represents the "observed" (true or measured) values.

	x	y	p	q
P_1	**0.100**	**1.000**	0.800	1.700
P_2	0.150	1.225	**0.925**	**2.000**
O	**0.000**	**1.000**	**1.000**	**2.000**

Table 1. Comparing predictions P_1 and P_2 with O. Figures in bold indicate the closer agreement between prediction and observation.

It can be seen that, considering x and y, P_1 appears more accurate (closer to O) than P_2 in both x and y. However, considering p and q, P_2 appears more accurate than P_1 in both p and q.

As discussed in Section 1, this result seems to suggest the futility of any ranking of theories in terms of the closeness of their predictions to truth or measurement, for any theory with at least two variables, because we can always find an equivalent set of variables that will lead to the opposite ranking.

It should be noted that "appearance of accuracy" is measured here by the simple arithmetic difference between predicted and observed values. This notion of "accuracy" is challenged in this paper. Also, in this example, predictions from one theory are closer to observations than predictions from another in all respects. It is in this case that the paradox arises. The case of predictions from one theory being closer to observations than predictions from another in some respects but not in others is discussed in section 6.

3. The proposed resolution

Let us start by considering the assumption: "If a theory T_1 yields better metrical approximations than T_2 to the **true values** (or else to the **measured values**) …". We assume that there **are** such **true values**, but that we do not and cannot know what they are exactly; we cannot obtain them by measurement because all measurements contain uncertainty. Turning to **measured values**, a (well trained) scientist knows that no observation (measurement) is complete without an estimate of its uncertainty – uncertainty is always with us. Herein lies the resolution of the paradox.

Predictions from a theory also contain uncertainties. These arise either from inexactness in the theory, 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. Uncertainties in the initial conditions are inevitable because they are based (ultimately) on observations, which are uncertain. Note that, at this point, we are **not** trying to account for errors (falsities) in the theory itself; for the purposes of these calculations, we assume that a theory is true and we assess this claim through the subsequent calculations. (The prejudgement of a theory as false is discussed in section 6.)

Let us generalise the notation of section 2: an observation or prediction $\{\mathrm{x}, \mathrm{y}\}$ can be represented by a point $\mathbf{x}$ in a 2-dimensional space. Similarly, an observation or prediction $\{\mathrm{p}, \mathrm{q}\}$ can be represented by a point $\mathbf{p}$ in a different 2-dimensional space. (Note that this will allow us to use the same approach and notation for a general N-dimensional space.)

Let us assume, for the time being, that the two spaces are related linearly:

$$\mathbf{x} = \mathbf{A}\mathbf{p} \quad \text{and} \quad \mathbf{p} = \mathbf{B}\mathbf{x} \ , \tag{3.1}$$

where $\mathbf{A}$ and $\mathbf{B}$ are matrices that transform between the two spaces, and $\mathbf{B} = \mathbf{A}^{-1}$ where $^{-1}$ denotes matrix inverse.

Now consider an observation or prediction $\mathbf{x}$, equivalent to $\mathbf{p}$. Let the errors in $\mathbf{x}$ and $\mathbf{p}$ be $\boldsymbol{\varepsilon}_x$ and $\boldsymbol{\varepsilon}_p$ respectively. Note that this implies (unknown) true values, $\mathbf{x}_t$ and $\mathbf{p}_t$, with $\boldsymbol{\varepsilon}_x = \mathbf{x} - \mathbf{x}_t$ and $\boldsymbol{\varepsilon}_p = \mathbf{p} - \mathbf{p}_t$. Assume that these errors are drawn from a large ensemble with (provisionally, and for simplicity) zero mean and with covariance $\mathbf{C}_x$ in $\mathbf{x}$-space and $\mathbf{C}_p$ in $\mathbf{p}$-space.

Then, using (3.1), we can derive

$$\boldsymbol{\varepsilon}_p = \mathbf{B}\boldsymbol{\varepsilon}_x \ . \tag{3.2}$$

The uncertainty in $\mathbf{x}$ can be related to the uncertainty in $\mathbf{p}$ by taking the expected values of $\boldsymbol{\varepsilon}_x\boldsymbol{\varepsilon}_x{}^T$ and $\boldsymbol{\varepsilon}_p\boldsymbol{\varepsilon}_p{}^T$, which leads to

$$\mathbf{C}_p = E\{\boldsymbol{\varepsilon}_p \boldsymbol{\varepsilon}_p{}^T\} = E\{\mathbf{B}\boldsymbol{\varepsilon}_x \boldsymbol{\varepsilon}_x{}^T \mathbf{B}^T\} = \mathbf{B}\mathbf{C}_x\mathbf{B}^T \ , \tag{3.3}$$

where $E\{\ldots\}$ is the expected value operator and ^{T} denotes matrix transpose.

To measure the consistency between observation O with value $\mathbf{x}_o$ and prediction P_n with value $\mathbf{x}_n$, we evaluate, for all values of $\mathbf{x}$, the joint likelihood of $\mathbf{x}$ given $\mathbf{x}_o$ and $\mathbf{x}$ given $\mathbf{x}_n$, i.e. $P(\mathbf{x}|\mathbf{x}_o, \mathbf{x}_n)$. Assuming that the errors in $\mathbf{x}_o$ and $\mathbf{x}_n$ are independent, this gives:

$$P(\mathbf{x}|\mathbf{x}_o, \mathbf{x}_n) = P(\mathbf{x}|\mathbf{x}_o)P(\mathbf{x}|\mathbf{x}_n) \tag{3.4}$$

or

$$\ln\{P(\mathbf{x}|\mathbf{x}_o, \mathbf{x}_n)\} = \ln\{P(\mathbf{x}|\mathbf{x}_o)\} + \ln\{P(\mathbf{x}|\mathbf{x}_n)\} \ . \tag{3.5}$$

The observation and prediction will be considered "consistent" if there exists a range of $\mathbf{x}$ for which the joint likelihood (or its logarithm) is sufficiently high. Note that this implies that the concept of "consistency" involves not only the values $\mathbf{x}_o$ and $\mathbf{x}_n$ but also the PDFs of their uncertainties. This is what we would normally mean by "consistency"; broadly speaking, two estimates of the same quantity are considered consistent if they agree to within their respective uncertainties.

Assuming (again, provisionally and for simplicity) that the errors in the observation and the prediction have Gaussian PDFs, it can easily be shown that:

$$P(\mathbf{x}|\mathbf{x}_o) \propto \exp\{-\tfrac{1}{2}(\mathbf{x}-\mathbf{x}_o)^T \mathbf{C}_x^{o\,-1}(\mathbf{x}-\mathbf{x}_o)\} \tag{3.6}$$

$$P(\mathbf{x}|\mathbf{x}_n) \propto \exp\{-\tfrac{1}{2}(\mathbf{x}-\mathbf{x}_n)^T \mathbf{C}_x^{n\,-1}(\mathbf{x}-\mathbf{x}_n)\} \tag{3.7}$$

where $\mathbf{C}_x^o$ and $\mathbf{C}_x^n$ are the error covariances of $\mathbf{x}_o$ and $\mathbf{x}_n$ respectively.

Substituting (3.6) and (3.7) into (3.5), we obtain

$$\begin{aligned} &-2\ln\{P(\mathbf{x}|\mathbf{x}_o, \mathbf{x}_n)\} + k = \\ &(\mathbf{x}-\mathbf{x}_o)^T \mathbf{C}_x^{o\,-1}(\mathbf{x}-\mathbf{x}_o) + (\mathbf{x}-\mathbf{x}_n)^T \mathbf{C}_x^{n\,-1}(\mathbf{x}-\mathbf{x}_n) = J_x^n(\mathbf{x}) \end{aligned} \tag{3.8}$$

where k is a constant. This equation defines $J_x^n(\mathbf{x})$, which is known as a "cost" or "penalty" function, and it is directly related to the likelihood that (for this problem) the prediction is consistent with the observation. It quantifies the fit of any value of $\mathbf{x}$ to the observation and the prediction. The first term on the right-hand side represents the fit to the observation and the second the fit to the prediction. Note that, in the case of Gaussian PDFs, the logarithmic terms in (3.5) become quadratic terms in (3.8), and this is why this form of the cost function is usually convenient.

In $\mathbf{p}$-space, the equivalent relation is:

$$J_p^n(\mathbf{p}) = (\mathbf{p}-\mathbf{p}_o)^T \mathbf{C}_p^{o\,-1}(\mathbf{p}-\mathbf{p}_o) + (\mathbf{p}-\mathbf{p}_n)^T \mathbf{C}_p^{n\,-1}(\mathbf{p}-\mathbf{p}_n) \tag{3.9}$$

Substituting (3.3) and (3.1) into (3.9) gives:

$$\begin{aligned} J_p^n(\mathbf{x}) &= (\mathbf{x}-\mathbf{x}_o)^T \mathbf{B}^T(\mathbf{B}\mathbf{C}_x^o\mathbf{B}^T)^{-1}\mathbf{B}(\mathbf{x}-\mathbf{x}_o) + (\mathbf{x}-\mathbf{x}_n)^T \mathbf{B}^T(\mathbf{B}\mathbf{C}_x^n\mathbf{B}^T)^{-1}\mathbf{B}(\mathbf{x}-\mathbf{x}_n) \\ &= (\mathbf{x}-\mathbf{x}_o)^T \mathbf{C}_x^{o\,-1}(\mathbf{x}-\mathbf{x}_o) + (\mathbf{x}-\mathbf{x}_n)^T \mathbf{C}_x^{n\,-1}(\mathbf{x}-\mathbf{x}_n) = J_x^n(\mathbf{x}) \end{aligned} \tag{3.10}$$

This shows that the transformation from $\mathbf{x}$ to $\mathbf{p}$ makes no difference to the likelihood that the observation is consistent with the prediction, P_n . It thus resolves the paradox, because it shows that there are at least some important cases (i.e. those involving linear transformations and Gaussian PDFs) where the quantitative fit of observation to prediction, and hence the ranking of theories, is unaffected by the transformation between spaces.

4. Some examples

In the example given by Popper (1979), $\mathbf{A}$ and $\mathbf{B}$ are 2x2 matrices:

$$\mathbf{A} = \begin{bmatrix} -2 & 1 \\ -3 & 2 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} -2 & 1 \\ -3 & 2 \end{bmatrix} . \tag{4.1}$$

(The fact that $\mathbf{A} = \mathbf{B}$ here appears to be purely coincidental; nothing relies on this, as the proof in section 3 shows. The example given by Miller (1975a) does not have this property.)

Let the two theories, T_1 and T_2, lead to predictions P_1 and P_2 respectively. In the example given, these predictions have values:

$$\mathbf{x}_1 = \begin{bmatrix} 0.1 \\ 1 \end{bmatrix} \text{ and so } \mathbf{p}_1 = \begin{bmatrix} 0.8 \\ 1.7 \end{bmatrix} \text{, and } \mathbf{x}_2 = \begin{bmatrix} 0.15 \\ 1.225 \end{bmatrix} \text{ and so } \mathbf{p}_2 = \begin{bmatrix} 0.925 \\ 2 \end{bmatrix} .$$

We are also given

$$\mathbf{x}_o = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \text{ and so } \mathbf{p}_o = \begin{bmatrix} 1 \\ 2 \end{bmatrix} .$$

Let us first consider the (limiting and unrealistic) case where the predictions are exact. Then $P(\mathbf{x}|\mathbf{x}_n)$ is non-zero only at $\mathbf{x} = \mathbf{x}_n$. The last term in (3.8), the fit to prediction, becomes infinite for $\mathbf{x} \neq \mathbf{x}_n$. Therefore we need only consider the fit to observation for $\mathbf{x} = \mathbf{x}_n$.

If we choose, as an example, $\mathbf{C}_x^o = a^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, where a is a constant defining the magnitude of the uncertainty, then, through (3.3) and (4.1), $\mathbf{C}_p^o = a^2 \begin{bmatrix} 5 & 8 \\ 8 & 13 \end{bmatrix}$. Substituting these values into (3.10) gives

$$J_p^1 = J_x^1 = 0.01/a^2$$

and $$J_p^2 = J_x^2 = 0.073125/a^2.$$

Therefore, whether measured in the space of $\mathbf{x}$ or $\mathbf{p}$, the “cost” of the observation O with prediction P_2 is 7.3125 times higher than the cost with prediction P_1.

These results are plotted (to scale) in Figure 1, with the data points offset such that the observation is placed at the origin in each diagram. All points on the circle (left) or ellipse (right) are points of equal probability that the true value (of $\mathbf{x}$ or $\mathbf{p}$) will lie at these points in the plane. Points inside the circle/ellipse are more likely and those outside less likely. Figure 1 illustrates the resolution of the paradox: the transformation from $\mathbf{x}$- to $\mathbf{p}$-space does not change the ranking of P_1 and P_2 in terms of their closeness to O – P_1 remains more likely than P_2. Also, Figure 1 shows that the shorter Euclidean distance does not necessarily represent the “closer” fit of prediction to observation; the uncertainties in the observation must be taken into account.

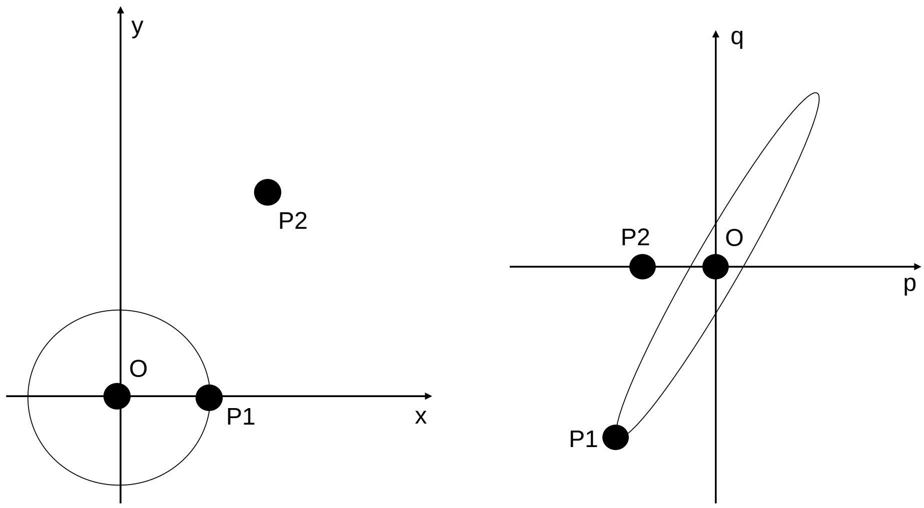

Fig 1. Illustrating the example in the **x**-space (left) and the **p**-space (right). “O”, “P1” and “P2” are the locations of the observation, prediction P_1 and prediction P_2 respectively. The circle (left) and ellipse (right) are surfaces of equal probability passing through P_1.

Let us now consider the case in which both observations and predictions are uncertain. This is illustrated in Figure 2, where we have chosen an error covariance for each prediction equal to that for the observation. The regions of overlap between the circles/ellipses in Fig.2 can be roughly interpreted as the regions of consistency between observation and prediction. Again, the closeness-of-fit of prediction to observation is not affected by the transformation from **x**-space to **p**-space.

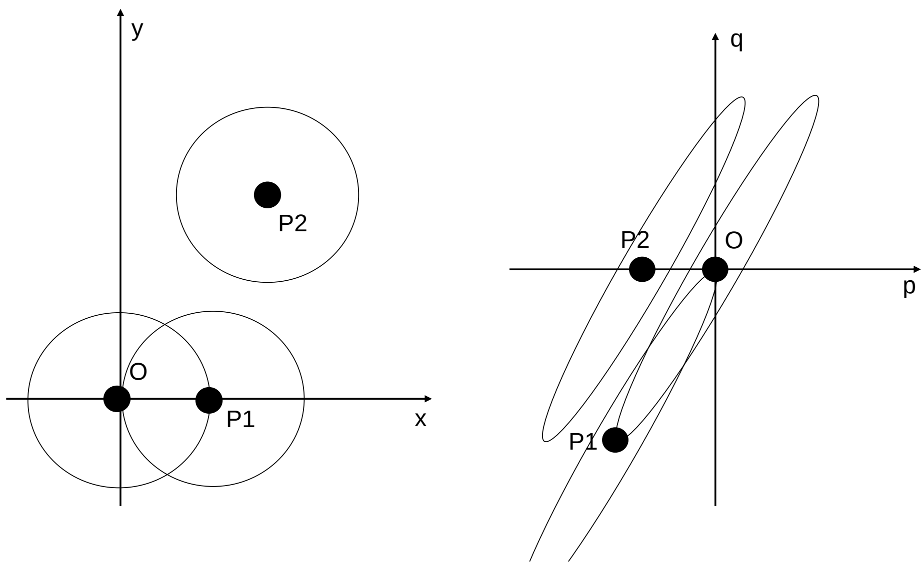

Fig 2. As Fig.1, including the effects of uncertainties in the predictions. The circles (left) and ellipses (right) are surfaces of equal probability centred on O, P_1 and P_2. The magnitudes of the uncertainties in observation and prediction have been set equal to each other.

The value of $\mathbf{C}_x^n$ is crucial in determining the utility of the prediction. As $\mathbf{C}_x^n$ becomes very large (compared to $\mathbf{C}_x^o$) the second term in (3.10) becomes negligible; the most probable value of $\mathbf{x}$ (i.e. the minimum of J) is determined only by the observation. This means that any observation is consistent with the theory and (in Popperian terms) that the theory is very weak.

Conversely, as $\mathbf{C}_x^n$ becomes very small, the region of $\mathbf{x}$-space (or $\mathbf{p}$-space) for which the probability of the prediction (given the theory) is significant becomes very small. Therefore the theory is very strong; it is, *a priori*, improbable and easy to falsify given sufficiently precise observations. If the PDFs of $\mathbf{x}_o$ and $\mathbf{x}_n$ do not significantly overlap, then the likelihood of any value of $\mathbf{x}$ being the true value is very low (given theory T_n); high cost (low likelihood) shows that the theory is inconsistent with the observation. This is what is meant by "falsification" in probablistic terms.

In the limit that both observation and prediction are exact and have different values, the cost is infinite (and the joint likelihood zero) for all values of $\mathbf{x}$. The theory is therefore false, and the quantitative cost or distance measure loses its meaning. This is effectively the limiting case considered by Miller and Popper. In this limit, the ranking of predictions based on "closeness" measures breaks down, consistent with Miller's original result. However, this result is not relevant to science, because in science uncertainty is always with us. The introduction of finite uncertainty resolves the paradox, and the parameters describing uncertainty lead to natural metrics for evaluating "agreement" or "consistency" between prediction and observation.

When we consider predictions P_1 and P_2, we can see that:
(i) If the values of $\mathbf{C}_x^1$ and $\mathbf{C}_x^2$ are sufficiently large that the PDFs of $\mathbf{x}_1$ and $\mathbf{x}_2$ are substantially overlapping, then an observation can falsify both or neither.
(ii) If the values of $\mathbf{C}_x^1$ and $\mathbf{C}_x^2$ are sufficiently small that the PDFs of $\mathbf{x}_1$ and $\mathbf{x}_2$ do not overlap, then an observation can falsify one theory or the other or both.
(iii) The value of $\mathbf{C}_x^o$ is crucial in determining the potential of the observation to falsify either theory.

These points are illustrated in Figure 3.

The above conclusions are only valid if the estimates $\mathbf{C}_x^o$, $\mathbf{C}_x^1$ and $\mathbf{C}_x^2$ are sufficiently accurate. Non-overlap of PDFs can arise because the uncertainties have been underestimated. It is prudent, therefore, to re-examine the uncertainties in the observations and/or the predictions, both their magnitudes and the forms of their PDFs, before rejecting the theories.

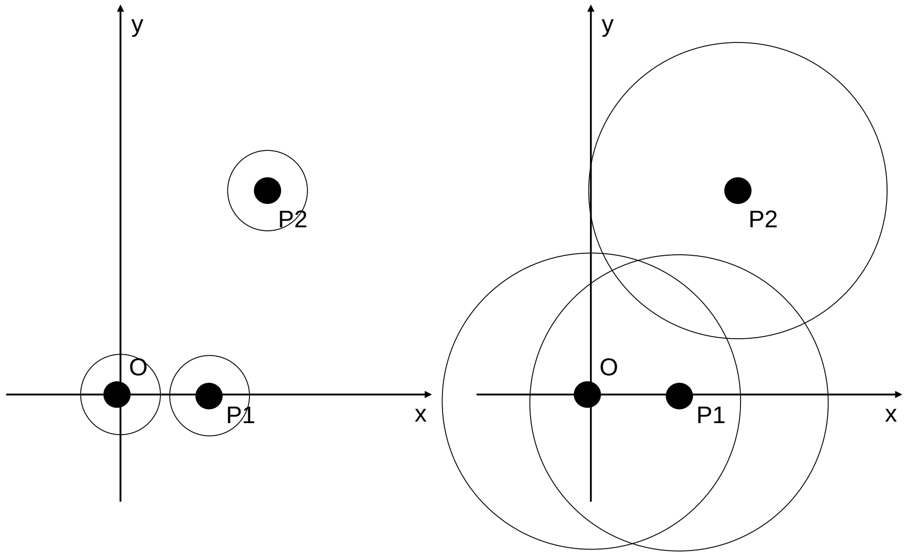

Fig 3. Illustrating the effects of varying the magnitudes of uncertainty in observations and predictions (in the $\mathbf{x}$-space). On the left they have been decreased relative to Fig.2 (left panel) and on the right increased. The circles show the associated surfaces of equal probability.

5. Nonlinear transformations and non-Gaussian error distributions

The extension of the theory presented in section 3 to the general case of nonlinear transformations is given in the Appendix, where the two nonlinear examples studied by Miller (2006) are analysed. Both examples exhibit essentially the same behaviour as found for the linear case; PDFs of uncertainty that are circular in $\mathbf{x}$-space distort to become near-elliptical or oval in $\mathbf{p}$-space, and in such a way that the ranking of "closeness" between prediction and observation is preserved by the transformation. Numerical experiments (not shown) confirm this, generating diagrams similar to Fig.2.

In the example in section 4, in order to keep the algebra simple, we chose a PDF of uncertainty centred on the observed/predicted value with a 2-dimensional Gaussian distribution about this value. We have shown that, in this case, the value of the penalty function is independent of the space in which it is calculated.

However, this result can be generalised; it will arise whatever the PDFs. The PDFs are defined for a specific space, in this case the $\mathbf{x}$-space. The values of $\mathbf{x}_o$ and $\mathbf{x}_n$ lie at positions with specific values of the PDFs in this space. Transformation to another space (e.g. the $\mathbf{p}$-space) will transform both the predicted values and the PDFs in the same way, giving the same values of the PDFs at the equivalent locations in the new space.

The computation of the cost function generalises in the same way: (3.8) remains

$$J_x^n = -2\ln\{P(\mathbf{x}|\mathbf{x}_o, \mathbf{x}_n)\} + k \ . \tag{5.1}$$

This can be used to evaluate the cost of the prediction relative to the observation, provided that mathematical expressions and quantitative values can be given for the conditional probabilities, $P(\mathbf{x}|\mathbf{x}_o)$ and $P(\mathbf{x}|\mathbf{x}_n)$.

Note that the two issues discussed in this section interact: a nonlinear transformation will convert a PDF that is Gaussian in one space into a PDF that is non-Gaussian in another.

6. **Discussion**

In response to the original statement of the problem by Miller (1975a), Good (1975) sketched briefly a response with similarities to the one given in section 3 of this paper. He warned that "closeness" is not necessarily best measured by Euclidean distance, that the PDFs of errors need to be considered, and that "the ratio of the probability densities … is invariant under all differentiable non-singular transformations of the plane". He also illustrated this with a diagram similar to Fig.1. However, as pointed out by Miller (1975b) in reply, Good's diagram did not show what it purported to show. In summary, however, Good's response was on the right track; the current paper attempts to give a more thorough analysis of the problem, with extensions to consider the effects of uncertainties in both predictions and uncertainties, and to consider in more depth the implications of the resolution of the paradox.

Miller (1975a, 1975b) placed great stress on the symmetry of the problem, i.e. of the symmetric transformation between the $\mathbf{x}$-space and the $\mathbf{p}$-space. We have shown here how the introduction of the relevant error covariances and PDFs breaks this symmetry, as illustrated in Fig.1 and Fig.2. However, a new symmetry is introduced: if a PDF is elliptical in $\mathbf{x}$-space then, with a suitable transformation, it will become circular in $\mathbf{p}$-space.

Popper (1979) speculated that the resolution of this problem, i.e. of how to avoid theory-choice being arbitrary, may lie with "the parameters that enter into the evaluation of competing physical theories". The analysis presented above shows how objective choice is possible; the relevant "parameters" are those describing the PDFs of uncertainty. These will never be known perfectly, but evidence from the real world of science suggests that (a) they are often known accurately enough and (b) their conjectural nature will require their periodic re-evaluation.

Both Popper and Miller presented the problem as applying only to the choice between **false** theories. This is because, if the theories are true and exact predictions from them are compared with perfect observations, then there will be no difference between them and the paradox will not arise. However, this is never the case – observations are always uncertain and predictions from theories, even true theories, are uncertain, if only because the initial conditions are uncertain. Therefore, the analysis in this paper applies to both true and false/approximate theories. In fact, sections 3 and 4 show how the effects of errors in theories themselves and in errors arising from other sources of uncertainty in the predictions can be separated; if observation and prediction differ by more than their expected uncertainties, then this indicates a problem with the theory itself (or else that the uncertainties have been mis-specified, either in their magnitudes or the shapes of their PDFs).

Moreover, in order to establish that a theory is false, a procedure similar to the one outlined in this paper must be followed; a theory will be judged false if its predictions are inconsistent with observations, and the judgement of "inconsistency" relies on quantification of uncertainty. More generally, the concepts of "closeness", "consistency", "agreement", etc., imply some concept of uncertainty; neglect of uncertainty lies at the heart of the paradoxical nature of the original problem.

As emphasised above, the values used for the PDFs of error of both observations and predictions play an important role in quantifying closeness-to-observation and, in practice, an important part of science involves the quantification of parameters describing these PDFs to an adequate approximation. One specific problem is systematic error (bias) in the observations. Biased observations can give erroneous support to predictions that happen to be biased in the same direction.

It could be argued that observational error can be reduced by averaging over many observations. However, it can never be reduced to zero. Also, although random error can be reduced by averaging, systematic error will remain and will eventually become dominant. Observational error is always with us!

In the figures in section 4, no values have been assigned to the equi-probability surfaces. It is not necessary to do so in order to demonstrate that the ranking of predictions is unaffected by the transformation between spaces. Moreover, no specific value of likelihood or cost can be associated with the judgement that a prediction is or is not inconsistent with an observation. Sufficient to say that the judgement on consistency is not a yes/no decision.

Although, it is suggested, the paradox as originally stated has been resolved – and this means that theory choice is not always rationally undecidable – it does not mean that it **is** decidable in all cases. This is for (at least) two reasons. Firstly, we have assessed quantitatively the case of predictions and/or observations with Gaussian errors. PDFs of relevant error distributions can be highly non-Gaussian. In fact, nonlinearity of theories/models is a major cause of non-Gaussianity in the errors of their predictions. As a result, the associated cost functions can become multi-modal and, in such cases, the measure of closeness-to-observation does not give a unique answer. Secondly, one false/approximate theory may give predictions that are closer to observations in some variables whereas another theory may be closer for other variables. Then, the question of theory choice is a pragmatic one, concerning the applications for which the predictions will be used.

We have assumed in this analysis that truth is point-like (as in the original paradox). However, it seems plausible that the analysis could be extended also to cover a probabilistic notion of the truth, provided that it remained sufficiently "local". We have also assumed that the observed and predicted values are of continuous variables. It seems plausible that the analysis could be extended to cover non-continuous (discrete) variables with suitably specified PDFs.

7. **Conclusions**

It is suggested that the paradox presented by Miller and Popper has been resolved. The paradox arises because of the neglect of the role of inevitable uncertainties, in observations and in predictions from theories. This supports confidence in the method that scientists normally adopt; science proceeds from one false theory, with a certain fit to observations, to another false theory, with a better fit to observations.

It has been shown how measures of "closeness-to-observation" can indeed be used in the choice between competing theories or models implementing these theories, provided that the closeness metric is grounded in probability theory and uses appropriate estimates of uncertainty. As pointed out by Good (1975), a simple Euclidean metric can be very misleading. (Consider the plight of the hungry traveller for whom the inn many miles down the road may, in practice, be "closer" than the one just across the raging river.)

Of course, comparisons with different observations can lead to conflicting results. This is normal in science; it requires the scientist to suspend judgement until more work has been done. Even then, one approximate theory may be closer in some respects and another in other respects. However, this need not always be true: it is possible for one approximate theory to be closer to observations (in a statistical sense) than is another in all respects, and irrespective of the space in which the closeness is measured.

Popper (1979) and Miller (1975a) both considered the topic of this paper alongside the problem of verisimilitude. The two are clearly different, the latter referring to the content of theories and the former to the closeness-to-observation of predictions from theories. However, it is also clear that there is a link between the two, and it is conjectured that a better understanding of "closeness-to-observation" may shed new light on the problem of verisimilitude.

Acknowledgements

I am very grateful to Mr David Miller for his interest in this work and for pointing me to relevant literature. I thank Dr Antony Galton for his comments, encouragement and advice. I am also grateful to my former colleagues, Professor Andrew Lorenc and Dr Mark Webb (Met Office, UK), and to two anonymous reviewers for their very helpful comments and suggestions.

References

Good, I. J. (1975). Comments on David Miller. *Synthese*, ***30***, 205-206. https://doi.org/10.1007/BF00485307

Miller, D. W. (1966). A paradox of information. *British Journal for the Philosophy of Science*, ***17***, 59-61. https://www.jstor.org/stable/686404

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

Miller, D. W. (1975b). The accuracy of predictions: a reply. *Synthese*, ***30***, 207-219. https://doi.org/10.1007/BF00485308

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

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

Popper, Karl, and Miller, David (1983). A proof of the impossibility of inductive probability. *Nature*, ***302***, 687-688. https://doi.org/10.1038/302687a0

Appendix: Nonlinear transformations

When the two spaces are related in a nonlinear way, (3.1) becomes

$$\mathbf{x} = A(\mathbf{p}) \qquad \text{and} \qquad \mathbf{p} = B(\mathbf{x}) \;, \tag{A.1}$$

where A and B are nonlinear matrix functions. We still assume, for this problem, that $\mathbf{x}$ and $\mathbf{p}$ have the same dimension.

Small departures from $\mathbf{x}$ and $\mathbf{p}$ are related by

$$\delta\mathbf{x} \approx \mathbf{A}(\mathbf{p})\delta\mathbf{p} \qquad \text{and} \qquad \delta\mathbf{p} \approx \mathbf{B}(\mathbf{x})\delta\mathbf{x} \ , \tag{A.2}$$

where $\mathbf{A}(\mathbf{p}) = \nabla_{\mathbf{p}}A(\mathbf{p})$, $\mathbf{B}(\mathbf{x}) = \nabla_{\mathbf{x}}B(\mathbf{x})$ and $\mathbf{B}(\mathbf{x}) = \mathbf{A}(\mathbf{p})^{-1}$. The relations in A.2 become exact as $\delta\mathbf{x} \to \mathbf{0}$ and $\delta\mathbf{p} \to \mathbf{0}$.

Therefore (3.2) becomes

$$\boldsymbol{\varepsilon}_p \approx \mathbf{B}(\mathbf{x})\boldsymbol{\varepsilon}_x \ , \tag{A.3}$$

and (3.3) becomes

$$\mathbf{C}_p(\mathbf{p}) \approx \mathbf{B}(\mathbf{x})\mathbf{C}_x(\mathbf{x})\mathbf{B}(\mathbf{x})^T \ . \tag{A.4}$$

Then (3.10) becomes

$$J_p^n(\mathbf{x}) \approx J_x^n(\mathbf{x}) \ , \tag{A.5}$$

and this also becomes exact in the linear limit and in the limit of small errors.

Consider the two examples of nonlinear transformations given by Miller (2006).

Example 1. Transformation from the length and breadth of a rectangle to its area and perimeter.

In this case, (2.1) (ii) becomes $\mathrm{p} = \mathrm{xy}$ and $\mathrm{q} = 2(\mathrm{x} + \mathrm{y})$.
Differentiation gives $\mathbf{B} = \mathbf{A}^{-1} = \begin{bmatrix} \mathrm{y} & \mathrm{x} \\ 2 & 2 \end{bmatrix}$.
Then, if $\mathbf{C}_x = a^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\mathbf{C}_p = a^2 \begin{bmatrix} \mathrm{y}^2 + 4 & \mathrm{xy} + 4 \\ \mathrm{xy} + 4 & \mathrm{x}^2 + 4 \end{bmatrix}$.

Therefore, for a circular $\mathbf{C}_x$, $\mathbf{C}_p$ has an elliptical shape for small uncertainties and becomes a distorted ellipse (oval) for large uncertainties. For the example given by Miller (a rectangle of 34x23), $\mathbf{C}_p$ is highly elliptical, and it also becomes singular when $\mathrm{x} = \mathrm{y}$. Moreover, the elliptical nature of $\mathbf{C}_p$ is such that the ellipse for the prediction that is closer to observation in $\mathbf{x}$-space remains closer in $\mathbf{p}$-space (similar to Fig.2).

Example 2. Transformation from Cartesian to polar coordinates.

In this case (2.1) becomes
(i) $\mathrm{x} = \mathrm{p}\ \cos \mathrm{q}$ and $\mathrm{y} = \mathrm{p}\ \sin \mathrm{q}$,
(ii) $\mathrm{p} = \sqrt{\mathrm{x}^2 + \mathrm{y}^2}$ and $\mathrm{q} = \tan^{-1}(\mathrm{y}/\mathrm{x})$.

Differentiation gives $\mathbf{A} = \begin{bmatrix} \cos \mathrm{q} & -\mathrm{p}\ \sin \mathrm{q} \\ \sin \mathrm{q} & \mathrm{p}\ \cos \mathrm{q} \end{bmatrix} = \begin{bmatrix} \mathrm{x}/\mathrm{p} & -\mathrm{y} \\ \mathrm{y}/\mathrm{p} & \mathrm{x} \end{bmatrix}$,
and $\mathbf{B} = \mathbf{A}^{-1} = \frac{1}{\mathrm{p}} \begin{bmatrix} \mathrm{x} & \mathrm{y} \\ -\mathrm{y}/\mathrm{p} & \mathrm{x}/\mathrm{p} \end{bmatrix}$.
Then, if $\mathbf{C}_x = a^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\mathbf{C}_p = a^2 \begin{bmatrix} 1 & 0 \\ 0 & 1/\mathrm{p} \end{bmatrix}$.

Therefore, for a circular $\mathbf{C}_x$, $\mathbf{C}_p$ has an elliptical shape for small uncertainties and becomes a distorted ellipse (oval) for large uncertainties.

Note that, for a given value of (y/x), $\mathrm{q} = \tan^{-1}(\mathrm{y}/\mathrm{x})$ can take values, $\mathrm{q} + \mathrm{n}\boldsymbol{\pi}$, where n is any integer. This makes the transformation from $\mathbf{x}$-space to $\mathbf{p}$-space multi-valued. However, this ambiguity can easily be resolved when the computing the "closeness" between two points, by considering only values of q with the least difference between them.

The transformation collapses at $\mathrm{p} = 0$. Also, when p is small (smaller than the uncertainty in x and y), the ellipse of uncertainty in $\mathbf{p}$-space distorts such that it no longer surrounds the point $\mathbf{p}$. This issue suggests that some transformations, whilst mathematically possible, are not physically realistic.