Mathematical note: The bias and sensitivity of statistics used for the analysis of presentiment experiments

Chris Phillips
26 July 2021

Abstract. It is known that the averaging of data from presentiment experiments can produce an artefactual bias which may mimic a presentiment effect when none is really present. Various strategies for dealing with this problem have been proposed, but there is concern that they may reduce the sensitivity of the analysis to real presentiment effects if they exist. To help in determining the best strategies, mathematical expressions are presented here for the expected values and variances of four statistical measures that may be used to analyse presentiment experiments. One of these measures, denoted by W, based on suitably rescaled sums of the data rather than averages, is free from the artefactual bias. It potentially has a sensitivity comparable with that of the biased statistics based on averages, provided that a suitable initial baseline estimate is subtracted from the measured physiological variable, and provided that the baseline does not drift too much in the course of each experimental session. The statistic W can also be straightforwardly generalised to the situation in which the types of the stimuli used are not determined randomly and independently, for example in an experiment using a counterbalanced design.

1. Introduction

Studies designed to detect the phenomenon usually called "presentiment" (or sometimes "precognitive anticipation" or something similar) have been among the most popular in experimental parapsychology over the past few years. It is claimed that collectively such studies have produced strong evidence that physiological processes can be influenced by future events through a form of precognition (Mossbridge *et al.*, 2012; Duggan and Tressoldi, 2018).

Generally these studies involve exposing subjects to a series of stimuli while some physiological variable, such as skin conductance or heart rate, is monitored. Either stimuli of two different types are applied, or else a stimulus is applied at some times but not at others. Normally the sequence of stimuli is determined randomly, so that their type cannot be anticipated by conventional means. Of course, each stimulus will generally affect the physiological variable after it is applied. But the idea of presentiment is that in addition there could be a time-reversed effect, in which the physiological variable would also tend to change before the application of the stimulus.

Typically, in order to try to detect such an effect, a comparison has been made between average values of the physiological variable before the two different types of stimuli (or before the stimuli and in their absence), and any differences found have been analysed statistically to determine whether they are significant. The null hypothesis is that there is no time-reversed influence, and therefore no statistically significant difference between the measurements made before the two types of stimuli.

However, the situation is complicated by the fact that the physiological variable can also be influenced by all the previous stimuli that have been applied. Provided that the types of all the stimuli in the series are determined randomly and independently, then considering any individual stimulus, the measurements made immediately before it should indeed be statistically independent of its type. But difficulties can arise when different measurements from the same series are combined together.

For definiteness, the experiment will be considered to consist of a number of sessions, each comprising a sequence of trials, whose types - either active (or arousing) or control (or calming) - are determined randomly and independently of one another.

Traditionally such experiments have been analysed by first calculating two averages of the measurements within each session - one average for the active trials and one for the control trials. The averages for each session are then averaged again, over all the sessions in the experiment. Finally one is subtracted from the other and it is determined whether the difference is statistically significant. Unfortunately, it is now known that this method of analysis is vulnerable to bias, in the form of an averaging artefact which can mimic a presentiment effect even though none is really present. (Wackermann, 2002; Dalkvist *et al.*, 2002)

Essentially the source of the bias is that the two-stage averaging procedure gives greater weight to measurements for a given trial type when they come from sessions that contain fewer trials of that type. Thus, sessions containing fewer active trials make a disproportionate contribution to the overall average for active trials. The problem is that the physiological quantity being measured may tend to be larger in series containing fewer arousing stimuli (for example through what is sometimes called "expectation bias"). If that is the case, then the averaging artefact will tend to raise the overall average of measurements for active trials. Conversely it will tend to lower the overall average for control trials. In this way, there can be a systematic difference between the averages of measurements for the two trial types, which mimics a presentiment effect.

(As discussed below, another potential source of bias arises from the possibility that by chance all the trials in a session may be of the same type. In that case an average can be calculated for only one trial type. In general excluding such sessions will result in bias. Provided there are a large number of trials in each session, this is found to be a relatively small effect, compared with the other artefact just described. But if bias is to be eliminated completely, both these sources of error must be dealt with.)

What is desitrable is a way of solving the problem of bias without too great a sacrifice of statistical sensitivity - that is, without excessively reducing the power of experiments to detect any genuine presentiment effect that may be present.

Various strategies to achieve this have been proposed, a number of which have been discussed by Dalkvist *et al.* (2014). Essentially they fall into three categories:

(1) The use of alternative statistics which are not biased. For example, it was shown by Dalkvist *et al.* (2002) that if measurements within each session were summed rather than averaged, the bias would be eliminated. An alternative suggestion was the use of measurements for only a single trial in each session, which would also remove the bias (Dalkvist *et al.*, 2014). But the concern about this approach has been that the use of such measures may greatly decrease the statistical sensitivity in comparison with the standard (biased) two-stage averaging technique.

(2) The use of other methods either to reduce or to estimate the size of the bias. For example, the bias can be reduced by using averages taken over all the trials of a given type in the whole experiment, rather than two-stage averages calculated first within sessions and then over different sessions (Dalkvist *et al.*, 2002). Although this reduces the potential bias, it does not eliminate it altogether. This method too would involve some loss of statistical sensitivity.

(3) The use of an alternative approach in which the measurements made before each trial are used to try to predict its type (Kennedy, 2013, 2014). This would eliminate the bias and would permit an exact statistical analysis of the results as a series of binary trials. But it would require a suitable

criterion for the prediction of trial type to be formulated first, on the basis of preliminary studies, and so far it has not found favour with experimenters.

The outline of this note is as follows.

In Section 2, to help in the evaluation of the first two strategies discussed above, four possible measures are considered: (i) the standard statistic *U*, based on the traditional two-stage averaging process, (ii) the pooled statistic $U^{(P)}$, based on averages for the two trial types over the whole experiment, (iii) V_n, based on a single trial from each series and (iv) *W*, based on sums rather than averages. For each of these, assuming the null hypothesis (no presentiment), mathematical expressions are given for the expected value of the statistic (that is, its average over all possible realisations of the experiment) - which determines whether the statistic is biased and how big the bias is - and its variance - which determines how sensitive the statistic would be to a genuine presentiment effect. In addition to exact expressions, approximations are given on the assumption that the number of trials in each session and/or the number of sessions in the experiment are large. The results are summarised in the Table before Section 4.

In Section 3, it is shown that the sensitivity of the fourth statistic, *W*, can be improved by subtracting an initial baseline estimate for the measured physiological variable, obtained by averaging the data from the first few trials. The resulting sensitivity may be comparable to that of the (biased) standard statistic *U,* provided that the baseline does not drift too far during the course of each experimental session.

Some of the mathematical details of the calculations are outlined separately in the Appendix, where (in Section A.4) it is also shown that the definition of the statistic *W* can be straightforwardly extended to experimental designs where the types of the trials are not determined randomly and independently, such as counterbalanced designs.

In Section 4 the results are discussed and some general conclusions and recommendations are given.

2. Expected values and variances of statistics

2.1. Formulation

In the experimental protocol considered here, there are *R* sessions, each consisting of *N* trials. The trials are of two types: *A* - active (or arousing) - and *C* - control (or calming). In the active trials a stimulus of some kind is applied to the participant, and in the control trials there is a stimulus of a different kind, or no stimulus at all. The type of each trial is assumed to be randomly determined, independently of all the other trial types, with the probability of an active trial equal to *a*. (A generalisation to non-independent trial types is discussed in the Appendix, Section A.4.)

At the beginning of each trial, immediately before the application of the stimulus (if there is one), a measurement is made of some physiological variable *X*. For definiteness, it will be assumed that the normal effect of the active stimulus is to cause an increase in *X* after it has been applied. The purpose of the experiment is to determine whether the stimulus can also cause an increase in *X* *before* it has been applied, which would reflect a presentiment effect. For the purpose of statistical analysis, the null hypothesis is that there is no effect - that is, the stimulus cannot influence earlier events. So it is assumed that although X_n, the value of *X* measured in the *n*th trial, can be influenced by the types of all the earlier trials (and can depend statistically on all the values of *X*

measured earlier in the session), it cannot depend on the type of the *n*th or later trials.

The data obtained will vary between experiments for several reasons: (1) the random choice of trial types will cause variations in the physiological quantity being measured; (2) the responses of different participants will vary and (3) the response of each participant will vary on different occasions (even if the choice of trial types remains the same). All these sources of variation must be considered when analysing the experimental data statistically. The expected value (or mean) and variance of the variable X_n will be denoted by μ_n and σ_n^2. It is important to remember that these quantities take account not only of the influence of the different possible trial types, but also of the variations between different participants and between the response of the same participant on different occasions. As already explained, the variables X_n measured within the same session are not assumed to be statistically independent of one another, or of the types of the trials preceding the *n*th trial.

The purpose of this note is to consider several statistics which can be used to analyse the data obtained in such an experiment. First it is determined whether each statistic is biased - that is, whether or not it can produce an artefactual indication of a difference between the measurements made in active and control trials, of the kind explained above.

Then mathematical expressions are obtained for the expected value and the variance of each statistic under the null hypothesis. Where exact expressions are not available, approximations for the expected values and variances of the statistics are given. These approximations require *R* or *N* (or both) to be large.

In each case the statistic is defined so that it represents a difference between averages (or appropriately rescaled sums) of measurements made in active and control trials. If a statistic is unbiased, on the null hypothesis its expected value will be zero. But if there were a genuine presentiment effect, the expected value would reflect the resulting average change in the variable *X*.

For the unbiased statistics, if a genuine presentiment effect of a given size were present, the power of the statistical test would be determined by the variance of the statistic - if the variance were low, the statistic would more likely to produce a significant result, and the statistical power of the experiment would be greater.

The following four statistics are considered:

(1) *U*, obtained by finding the average of X_n over the active trials in a single session, subtracting the average over control trials in the same session, and then averaging the difference over all the sessions (Section 2.2);

(2) $U^{(P)}$, obtained by finding the average of X_n over all the active trials in the whole experiment, and then subtracting the average over all the control trials in the whole experiment (Section 2.3);

(3) V_n, obtained by considering a particular value of *n,* finding the average of X_n over all the sessions in which the *n*th trial is active, and then subtracting the average over all the sessions in which the *n*th trial is a control (Section 2.4) and

(4) *W*, obtained by calculating appropriately rescaled sums of X_n over all the active trials in a single session and over all the control trials in the same session, finding the difference between them, and then averaging the difference over all the sessions (Section 2.5). This statistic can be generalised to experimental designs in which the types of the trials are not chosen independently of

one another (Appendix, Section A.4).

2.2. Statistic *U* based on the difference of averages for each session (biased)

The method originally used to analyse the results of presentiment experiments was to calculate average values of X for each trial type within each session, then to find the difference between these two averages, and finally to calculate the average U of this difference over all the sessions. That is, U is the average over all the sessions of the statistic F, which is defined for a single session as a sum over the trials from $n=1$ to N:

$$F=\sum_{n=1}^{N} Y_n \tag{1}$$

where

$$Y_n = \begin{cases} \dfrac{X_n}{N^{(A)}} & \text{if trial } n \text{ is of type } A \\ -\dfrac{X_n}{N-N^{(A)}} & \text{if trial } n \text{ is of type } C \end{cases} \tag{2}$$

in which $N^{(A)}$ is the number of active trials in the session. Thus the sum of the Y_n over the active trials is equal to the average of the X_n over those trials, and the sum of the Y_n over the control trials is equal to -1 times the average of the X_n over those trials. Therefore F is the difference between the averages of the X_n for the two trial types, for a single session.

The statistic U was used because it was hoped that, if it were found to differ significantly from zero, that would provide evidence against the null hypothesis - and by implication in favour of a presentiment effect, through which the value of the physiological variable could be influenced by a stimulus which had not yet been applied.

Unfortunately, it is now known that the statistic U is liable to be biased because the variable X_n is influenced by the types of all the trials preceding the nth trial. The bias arises because of the particular form of averaging used in defining U. Specifically, values of X_n from sessions with smaller numbers of active trials are given a larger weighting in the averages over active trials, and a smaller weighting in the averages over control trials (and the converse is true for sessions with larger numbers of active trials). The result is that the expected value of U under the null hypothesis can be non-zero, and that measured values of U can differ significantly from zero even when the null hypothesis is true.

This non-zero expected value of U is calculated in the Appendix (Section A.1). In calculating the expected value, care needs to be taken in the treatment of the special cases in which all the trials of a particular session are of the same type. This can occur because each trial type is assumed to be determined randomly and independently of the other trial types. For such a session, the statistic F defined by equation (1) will not represent the difference of two averages, but only an average for the trial type that is present (multiplied by -1 if all the trials are controls). The inclusion of such cases will produce an additional bias in the expected value of U, because (unless $a=½$) it will not

be equally likely for the trials in a session to be all active and or all controls. For this reason, in the calculations for U presented here, sessions in which all the trials are of the same type are considered to be excluded. (But note that this exclusion affects only the exact results for U given in the Appendix, Section A1. The large-N approximations given in equations (3-5) and (7) are unaffected, because the effects of the exclusion can be shown to be exponentially small in N.)

Expected value of *U*

The expected value E(U) is defined as the average of U over all realisations of the experiment (excluding cases in which all the trials in a session are of the same type, as explained above). That is, an average over possible sequences of trial types, over possible participants and over possible responses of a participant to a given sequence of trial types. Because U is the average over all sessions of the statistic F, which is defined for a single session, E(U) is equal to E(F), and an exact expression for E(F) can be obtained from equations (1-2) above and from (A1-A2) in the Appendix. If the number of trials per session, N, is assumed to be large, this expression can be simplified to give an approximate equation for E(U):

$$\mathrm{E}(U) = -N^{-2}\sum_{n=1}^{N}\mu_n' + O\left(N^{-2}\right) \tag{3}$$

in which the final term indicates an error in the approximation which is proportional to N^{-2} when N is large. The quantity μ_n' is defined by equation (A7). Although its mathematical definition looks complicated, it has a simple interpretation. If an expected value of X is calculated for the nth trial, taking into account only realisations of the experiment for which there are k active trials in the session, then in general that expected value will depend on the value of k. The quantity μ_n' reflects the rate of change of the expected value of X_n as k is increased. That is, it reflects how strongly the average variable of the variable depends on the total number of active trials in the session.

This result can be expressed more tidily as

$$\mathrm{E}(U) = -N^{-1}\overline{\mu'} + O\left(N^{-2}\right) \tag{4}$$

in which the overbar is shorthand for an average over values of n from 1 to N.

The statistic U is defined by subtracting the average value of X_n in control trials from the average value in active trials - so that if the physiological variable being measured differed between active and control trials because of a genuine presentiment effect, that difference would contribute to the expected value of U. But equation (4) shows that even on the null hypothesis - in the absence of a presentiment effect - the averaging process can give rise to a non-zero expected value. The statistic U therefore has a bias, which in this approximation is proportional to N^{-1}, as shown by previous workers (Wackermann, 2002).

From the definition of μ', if the value of X tends to decrease when there are more active trials - or, conversely, to increase when there are fewer active trials - then $\overline{\mu'}$ is a negative number, and E(U) is greater than zero. In those circumstances, the bias of U will mimic a presentiment effect, even though none is present.

Variance of *U*

Although U is a biased statistic, it will still be useful to know its variance, for purposes of comparison with alternative, unbiased statistics. As U is the average of the statistic F, which is defined for a single session, its variance depends on the expected value of the square of F. This is given exactly by equation (A9) in the Appendix, and in the large-N approximation by equations (A10-A11).

These equations are still quite complicated. However, when N is large it is reasonable to make an additional assumption, which simplifies the result further. It will be assumed that when a pair of trials - the mth trial and the subsequent nth trial - are separated by a large number of other trials, then the type of the mth trial (i.e. whether the mth trial is active or control) will have only a small effect on the value of X measured in the nth trial. That is, that the influence of an earlier trial type on later measurements tends to be "screened out" if there are many intervening trials. With this additional assumption, using equation (A12), the variance of U can be approximated by

$$\mathrm{Var}(U) = \frac{R^{-1}N^{-1}}{a(1-a)}\left(\overline{\mu^2}-\overline{\mu}^2+\overline{\sigma^2}-\overline{\gamma}\right)+O\left(R^{-1}N^{-2}\right) \tag{5}$$

As in equations (3) and (4), the overbar is shorthand for an average. For the terms involving the expected value μ_n and the standard deviation σ_n, it denotes an average over all the trials in the session - that is, all the values of n. The remaining term is defined in terms of the covariance γ_{mn} of the values X_m and X_n for pairs of different trials in the same session, and here the overbar denotes an average over all such pairs in the session - that is, all pairs m and n for which $m \neq n$.

In the large-N approximation, Var(U) is proportional to $R^{-1}N^{-1}$. The final term of equation (5) indicates an error in the approximation which is proportional to $R^{-1}N^{-2}$ when N is large.

A simpler approximation for the variance of U can be obtained by working in terms of D_n, defined as the difference between the measurement at the nth trial X_n and the average value for the session, $\overline{X}$:

$$D_n = X_n - \overline{X} \tag{6}$$

From this it is straightforward to show that

$$\mathrm{Var}(U) = \frac{R^{-1}N^{-1}}{a(1-a)}\mathrm{E}\left(\overline{D^2}\right)+O\left(R^{-1}N^{-2}\right) \tag{7}$$

Equations (4), (5) and (7) provide large-N approximations for the expected value and variance of the statistic U.

It is also true that, if the number of sessions, R, is large, then U will be approximately normally distributed, because it is the average of a large number of independent, identically distributed variables.

2.3. Statistic $U^{(P)}$ based on the difference of averages over the whole experiment (biased)

Although the statistic U is biased, it has previously been noted that the bias can be reduced by pooling all the sessions in an experiment into a single series before calculating the averages for the two trial types and finding the difference between them (Dalkvist *et al.*, 2002).

The resulting statistic, $U^{(P)}$, is defined by equations similar to those numbered (1) and (2) above, but with N replaced by the total number of trials in the whole experiment, namely RN, and $N^{(A)}$ replaced with the total number of active trials in all the sessions. This means that the results for U already obtained can be applied directly to give the expected value and variance for $U^{(P)}$.

In addition, when R is large, the expression for Var($U^{(P)}$) can be simplified further, because the covariance γ_{mn} of the values X_m and X_n for pairs of different trials is non-zero only when both trials are in the same session. Thus the average of the covariance over all pairs of trials in the whole experiment is a factor of R^{-1} smaller than in equation (5), and may therefore be omitted from the leading-order approximation. This gives

$$\mathrm{E}\left(U^{(P)}\right) = -R^{-1}N^{-1}\overline{\mu'} + O\left(R^{-2}N^{-2}\right) \tag{8}$$

$$\mathrm{Var}\left(U^{(P)}\right) = \frac{R^{-1}N^{-1}}{a(1-a)}\left(\overline{\mu^2} - \overline{\mu}^2 + \overline{\sigma^2}\right) + O\left(R^{-2}N^{-1}\right) \tag{9}$$

In the same way as for U, an alternative approximation for the variance can be obtained by working in terms of D_n, the difference of X_n from the average value for the session. This gives

$$\mathrm{Var}\left(U^{(P)}\right) = \frac{R^{-1}N^{-1}}{a(1-a)}\left(\mathrm{E}\left(\overline{D^2}\right) + \mathrm{Var}\left(\overline{X}\right)\right) + O\left(R^{-2}N^{-1}\right) \tag{10}$$

Although, like U, the alternative statistic $U^{(P)}$ is still biased, the bias is smaller by a factor proportional to R^{-1}. In this approximation, the variance of $U^{(P)}$ is proportional to $R^{-1}N^{-1}$, like that of U. However, although the dependence on R and N is the same, in practice the variance of $U^{(P)}$ will be somewhat larger than that of U, owing to the additional contribution from $\mathrm{Var}(\overline{X})$, reflecting the variation of the average value of the variable between different sessions.

Provided both N and R are large, again it can be shown that $U^{(P)}$ is approximately normally distributed (see Appendix, Section A.2).

In contrast to the other statistics considered here, it is not clear how $U^{(P)}$ can be used rationally in practice without relying on theoretical expressions for its variance, like those given above. The other statistics can be expressed in terms of sums of statistically independent and identically distributed contributions from each of the R different sessions. Therefore, provided R is large, their significance can be calculated using standard methods that require no assumptions about the probability distributions of the individual contributions. But this is not true of $U^{(P)}$, because the coefficients in its definition depend on the total number of active trials in the whole experiment. This means it cannot be expressed as a sum of independent contributions from the different sessions.

2.4. Statistic V_n based on the difference of averages of single trials from each session (unbiased)

Another proposal for dealing with the problem of bias is to use only a single trial from each session of the experiment - say the *n*th trial - and to discard all the other data. First two averages of X_n are calculated - one for sessions in which the *n*th trial is active, and the other for sessions in which it is a control. Then the difference, V_n, between these averages is found. (Cases where all the *n*th trials are of the same type - either all active or all controls - must be excluded.)

Mathematically, the definition of V_n is similar to that of U in equations (1) and (2) above. But the calculations are simplified because each X_n which contributes to the sums comes from a different session. Therefore, on the null hypothesis, all the X_n are statistically independent and identically distributed.

This makes it easy to show that the statistic is exactly unbiased:

$$\mathrm{E}(V_n) = 0 \tag{11}$$

Also, from the expression corresponding to equation (5), when R is large, the variance is given by

$$\mathrm{Var}(V_n) = \frac{R^{-1}\sigma_n^2}{a(1-a)} + O(R^{-2}) \tag{12}$$

For comparison with the other statistics, the variance of V_n can alternatively be expressed in terms of the variables V_n and $\overline{X}$ defined above. This gives

$$\mathrm{Var}(V_n) = \frac{R^{-1}}{a(1-a)}\left(\mathrm{Var}(D_n) + \mathrm{Var}(\overline{X}) + 2\,\mathrm{Cov}(D_n, \overline{X})\right) + O(R^{-2}) \tag{13}$$

Although V_n is an unbiased statistic, in this approximation its variance is proportional to R^{-1}, in contrast to U and $U^{(P)}$, which both have variances proportional to $R^{-1}N^{-1}$. Therefore when N is large the variance of V_n will be much larger than the variances of U and $U^{(P)}$. This is not surprising, as V_n makes use of only a single trial from each session, so that a large proportion of the experimental data is discarded.

Provided R is large, again it can be shown that V_n is approximately normally distributed (see Appendix, Section A.2).

2.5. Statistic *W* based on a difference of rescaled sums (unbiased)

Another alternative statistic, *W*, is based on suitably rescaled sums of the measurements, rather than averages (Dalkvist *et al.*, 2002). It may be defined as follows. For each session, two sums of the X_n are calculated - one for active trials and the other for controls. The sum for active trials is then divided by aN, the expected number of active trials per session, and similarly the sum for controls is divided by $(1-a)N$, the expected number of controls per session. Then the difference between the rescaled sums is found, and finally this difference is averaged over all sessions.

(Note that W differs from the statistic U defined in Section 2.2 in that each sum is divided by the expected number of trials of the appropriate type, rather than the actual number. This means it can also be viewed as the difference between two appropriately rescaled sums of the values of X_n over all the active - or control - trials in the whole experiment.)

Thus W is the average over all the sessions of the statistic *H,* defined for each session by

$$H = \sum_{n=1}^{N} Z_n \tag{14}$$

in which

$$Z_n = \begin{cases} \dfrac{X_n}{aN} & \text{if trial } n \text{ is of type } A \\ -\dfrac{X_n}{(1-a)N} & \text{if trial } n \text{ is of type } C \end{cases} \tag{15}$$

In this case both the expected value and the variance of W can be exactly calculated in a straightforward way, without the need for any approximation. On the null hypothesis,

$$\mathrm{E}(W) = 0 \tag{16}$$

$$\mathrm{Var}(W) = \frac{R^{-1}N^{-1}}{a(1-a)}\left(\overline{\mu^2}+\overline{\sigma^2}\right) \tag{17}$$

The statistic W is unbiased, because on the null hypothesis its expected value is zero. It is straightforward to show that if there were a genuine presentiment effect, the expected value of W would be equal to an average difference between the values of X_n in active and control trials.

As above, an alternative expression for the variance can be obtained by working in terms of D_n, the difference of X_n from the average value for the session. This gives

$$\mathrm{Var}(W) = \frac{R^{-1}N^{-1}}{a(1-a)}\left(\mathrm{E}\left(\overline{D^2}\right)+\mathrm{E}\left(\overline{X}^2\right)\right) \tag{18}$$

The variance of W is proportional to $R^{-1}N^{-1}$, like the variances of U and $U^{(P)}$ in the large-N approximation. However, comparison with equations (7) and (10) shows that the variance of W will be larger than that of $U^{(P)}$ if $\mathrm{E}(\overline{X}^2)$ is non-zero, while the variance of $U^{(P)}$ will in general be larger than that of U.

Provided the number of sessions R is large, then because W is the sum of a large number of independent, identically distributed statistics, its distribution will be approximately normal.

Note that the definition of W can be extended straightforwardly to the situation where the types of

the trials in each session are not chosen randomly and independently - for example in a counterbalanced experiment where the number of active trials is constrained to be the same in each session (see Appendix, Section A.4).

3. Statistic *W** based on a difference of rescaled sums, with a baseline estimate subtracted (unbiased)

Comparing equation (18) for the variance of *W* with equation (7) for the variance of the standard statistic *U*, it is evident that the sensitivity of *W* will be decreased if the average per session $\overline{X}$ differs from zero. This is because *W* is the difference of rescaled sums for the two trial types, rather than averages, and the numbers of terms in the sums will vary because the trial types are randomly determined.

For *W* to be useful, it will be necessary to reduce its variance by subtracting an estimate of the baseline value of X_n for each session. The simplest way of doing this is to use an average over the first *M* trials of the session, defined by:

$$\overline{X}^{(-)} = \frac{1}{M}\sum_{n=1}^{M} X_n \tag{19}$$

Then a version of X_n with the baseline estimate subtracted is

$$X_n^* = X_n - \overline{X}^{(-)} \tag{20}$$

and a new statistic *W** can defined by equations analogous to (14) and (15) above, but with X_n replaced by X_n^* and with the sum over trials restricted to the remaining range of trials, $n = M+1$ to *N*. Note that because the estimate of the baseline subtracted is the same for both active and control trials, the new statistic *W** remains unbiased.

In order to obtain a rough estimate of the effect of this baseline subtraction on the variance of *W**, a simple model of baseline drift may be considered, in which the variable X_n is expressed as the sum of one random component B_n, reflecting the variation of the baseline during the session, which will be assumed to change only gradually as a function of *n*, and another contribution $s_n T_n$, reflecting the variation from trial to trial, in which the s_n are fixed constants and the T_n are random variables:

$$X_n = B_n + s_n T_n \tag{21}$$

It will be assumed that each T_n has an expected value of zero and a variance of one, and that the variables T_n are uncorrelated with one another and with the baseline variables B_n :

$$\begin{aligned} &\mathrm{E}(T_n) = 0 \\ &\mathrm{E}(T_n^2) = 1 \\ &\mathrm{E}(T_m T_n) = 0 \qquad \text{for } m \neq n \\ &\mathrm{E}(B_m T_n) = 0 \qquad \text{for all } m \text{ and } n \end{aligned} \tag{22}$$

With these assumptions it is straightforward to show that the expected value and variance of the statistics *U* and *W* (without subtracting a baseline estimate) are

$$\begin{aligned} \mathrm{Var}(U) &= \frac{R^{-1}N^{-1}}{a(1-a)}\left(\overline{s^2} + \mathrm{E}\left(\overline{B^2}\right) - \mathrm{E}\left(\overline{B}^2\right)\right) + O\left(R^{-1}N^{-2}\right) \\ \mathrm{Var}(W) &= \frac{R^{-1}N^{-1}}{a(1-a)}\left(\overline{s^2} + \mathrm{E}\left(\overline{B^2}\right)\right) \end{aligned} \tag{23}$$

In general, the variance of the newly defined statistic *W** will depend on the value chosen for *M*. As shown in the Appendix, Section A.3, when *N* is large, the value of *M* that minimises the variance is much smaller than *N*. This optimal value of *M* is given by equation (A19), and the corresponding minimum value of the variance is

$$\mathrm{Var}(W^*) = \frac{R^{-1}N^{-1}}{a(1-a)}\left(\overline{s^2} + \mathrm{E}\left(\overline{(B-B_1)^2}\right)\right) + O\left(R^{-1}N^{-3/2}\right) \tag{24}$$

Note that this leader-order approximation does not depend strongly on the exact choice of *M,* but remains valid provided *M* is much less than *N* but much greater than 1.

The comparison between the variances of *U* and *W** can be made clearer by using an even simpler version of the model of baseline drift specified by equation (21). If the baseline variables B_n are assumed to depend linearly on *n* - that is, if the baseline is assumed to drift in a linear fashion with time (but at a rate and with an initial value that can vary randomly from session to session), it is found that at leading order the optimal value of *M* is given by

$$M \sim \frac{s_1 N^{1/2}}{\left(2\overline{s^2} - s_1^2 + \frac{1}{6}\mathrm{E}\left((B_N - B_1)^2\right)\right)^{1/2}} \tag{25}$$

and the corresponding variance of *W** is comparable to that of *U:*

$$\begin{aligned} \mathrm{Var}(U) &= \frac{R^{-1}N^{-1}}{a(1-a)}\left(\overline{s^2} + \frac{1}{12}\mathrm{E}\left((B_N - B_1)^2\right)\right) + O\left(R^{-1}N^{-2}\right) \\ \mathrm{Var}(W^*) &= \frac{R^{-1}N^{-1}}{a(1-a)}\left(\overline{s^2} + \frac{1}{3}\mathrm{E}\left((B_N - B_1)^2\right)\right) + O\left(R^{-1}N^{-3/2}\right) \end{aligned} \tag{26}$$

Table: Summary of results for U, $U^{(P)}$, V_n and W

Statistic	Basis	Biased or unbiased	Expected value	Variance
U	Difference between averages for the two trial types over each session, then averaged over all sessions	Biased	$-N^{-1}\overline{\mu'}+O\left(N^{-2}\right)$ when N is large	$\frac{R^{-1}N^{-1}}{a(1-a)}\mathrm{E}\left(\overline{D^2}\right)+O\left(R^{-1}N^{-2}\right)$ when N is large
$U^{(P)}$	Difference between averages for the two trial types over the whole experiment	Biased	$-R^{-1}N^{-1}\overline{\mu'}+O\left(R^{-2}N^{-2}\right)$ when both N and R are large	$\frac{R^{-1}N^{-1}}{a(1-a)}\left(\mathrm{E}\left(\overline{D^2}\right)+\mathrm{Var}\left(\overline{X}\right)\right)+O\left(R^{-2}N^{-1}\right)$ when both N and R are large
V_n	Difference between averages of the nth trial for the two trial types over all sessions	Unbiased	0 exactly	$\frac{R^{-1}}{a(1-a)}\left(\mathrm{Var}\left(D_n\right)+\mathrm{Var}\left(\overline{X}\right)+2\mathrm{Cov}\left(D_n,\overline{X}\right)\right)+O\left(R^{-2}\right)$ when R is large
W	Difference between appropriately rescaled sums for the two trial types over each session, then averaged over all sessions	Unbiased	0 exactly	$\frac{R^{-1}N^{-1}}{a(1-a)}\left(\mathrm{E}\left(\overline{D^2}\right)+\mathrm{E}\left(\overline{X}^2\right)\right)$ exactly

In this table, R is the number of sessions, N is the number of trials per session and a is the probability that a trial is active. The subscript n indicates a quantity related to the nth trial of a session, and the overbar an average over all values of n, from 1 to N. X_n is the value of the variable measured in the nth trial and D_n is the difference between X_n and the average value for the session, $\overline{X}$. The quantity μ' is defined by equation (A7).

The expected value E, the variance Var and the covariance Cov are defined as averages over all possible realisations of the experiment, taking into account different sequences of trial types, different participants, and variations in the response of participants on different occasions.

4. Discussion and practical recommendations

In this note, expressions have been presented for the expected values and variances of several statistics that can be used for the analysis of experimental studies of presentiment. Some of these statistics are known to be vulnerable to an averaging artefact, which can potential produce a statistical bias that may mimic a presentiment effect. The results presented here are intended to assist in understanding and eliminating this artefact, while preserving as much sensitivity as possible to any genuine presentiment effect that may be present.

The initial formulation of the problem outlined in Section 2.1 avoids any restrictive assumptions that might limit the usefulness of the results. In particular, at every stage of the experiment the physiological variable X is allowed to depend in an arbitrary way on all the previous events in the same session, including the randomly determined types of the previous trials in the session and the subject's reactions to them. It is this dependence that can give rise to the averaging artefact being considered.

The main results for the variables U, $U^{(P)}$, V_n and W are summarised in the Table. These include exact results for the expected values of V_n and W, which show that these variables are absolutely unbiased even under the most general assumptions. The same is true of the modified variable W^* considered in Section 3.

In addition, results are given for the variances of the statistics considered. Most of these are approximate results appropriate when N, the number of trials in each session, and/or R, the number of sessions in the experiment, are large. They are intended only as a guide to the sensitivity of the different statistics to any real presentiment effect that may exist, to assist in the choice of analysis method. Because they are only approximations, these values should not be used directly in the statistical analysis of experimental data. Instead, standard statistical methods such as the Student's t test should be used, in which the variance of the statistic used is estimated from the measured data.

The analysis of presentiment data involves a choice between statistics which are vulnerable to some degree of bias that may mimic the effect being studied (such as U and $U^{(P)}$) and those that are exactly unbiased but to some degree less sensitive (such as V_n, W and W^*).

Several techniques have been suggested by which the size of any artefactual bias might be estimated, but they are only approximate. There are also ways in which bias can be reduced (such as the use of the pooled statistic $U^{(P)}$ rather than the statistic U based on averages per session), but essentially bias remains difficult to quantify, except by direct comparison with an unbiased statistic.

There is a sense in which the methods of parapsychology, like Caesar's wife, must be above suspicion, and the potential for an unquantifiable statistical bias is extremely undesirable in a psi experiment. Therefore the recommendation of this note is that an exactly unbiased statistic should be used, even at the cost of some sacrifice of sensitivity.

Of the unbiased statistics considered here, V_n, which is based on data from a single trial in each session, clearly has a much larger variance - and therefore lower sensitivity - than the alternative statistics W and W^*, which are based on sums rather than averages of the measured variable. The recommendation is therefore to use W^*, in which an appropriate initial estimate of the baseline for each session is subtracted from the data, in order to improve the sensitivity.

In practice this will involve two steps:

(1) First, the data from a relatively small number of complete sessions will be used to estimate the optimal value of the parameter *M,* the number of early trials in each session that will be used to estimate the baseline. This can be done by considering in turn each possible value of *M*, calculating the statistic *W** for each of these sessions, and then choosing the value of *M* that minimises the variance of *W** between the sessions. (This is preferable to using the theoretical approximations for the optimal value of *M* presented above.) The data from these sessions must be excluded from further analysis.

(2) Then in each of the other sessions, the baseline will be estimated by averaging the first *M* measurements, and this estimate will be subtracted from the remaining measurements in order to calculate *W** (as explained in Section 4). The *M* data points of each session used to estimate the baseline must also be excluded from further analysis.

With this procedure, the loss of sensitivity incurred by the use of *W** rather than the standard statistic *U* should not be too great, provided that the baseline does not drift too much in the course of each session.

Two of the points made above should be stressed strongly. Firstly all the data used to estimate the optimal value of *M*, and also all the data used to estimate the baseline for each session, must be excluded from further analysis. Otherwise there will still be the potential for an artefactual bias of *W**. Secondly the results for *W**, which is the sum of statistically independent and identically distributed contributions from the different sessions, should be analysed using standard methods such as the Student's t test, which require no assumption about the variance of *W**. (The approximate results for variance presented above are intended only to assist in the choice of analysis method.)

A further advantage of the proposed statistic *W** is that, provided data for individual trials are available, it should be possible to reanalyse the results of previous studies in such a way as to eliminate the possibility of bias. Because the definition of *W** can also be extended to apply to counterbalanced designs (see Appendix, Section A.4), it may also be possible to apply it to a wider range of previous studies than can be analysed using the standard statistic *U*, potentially including conventional psychology experiments whose results may also be relevant to presentiment.

Finally it should be noted that, although *W** should be capable of eliminating bias with an acceptable sacrifice of sensitivity, it is still well worth considering the alternative suggestion made by Kennedy (2013, 2014). In this approach, rather than comparing the results of measurements for the two types of trials, an attempt would be made to use the data to predict the type of each trial in advance. Although this would require some preliminary study to identify and optimise a suitable prediction criterion, it would have several advantages. Like the method proposed here, it would absolutely eliminate the potential for artefactual bias. Also, if a genuine presentiment effect operated, it might be possible to identify a specific "signature" in the data that would increase the statistical power of the experiment. In addition, the statistical analysis could be done exactly using the binomial distribution, in contrast with standard methods that require either an assumption of normality or a large number of experimental sessions.

Appendix. Further details of the calculations

A.1. Statistic *U* based on the difference of averages per session (biased)

Expected value of *U*

Working from equation (1) for the statistic *F* defined for a single session, its expected value can be expressed as the sum of the expected values of the variables Y_n for individual trials:

$$\mathrm{E}(F) = \sum_{n=1}^{N} \langle Y_n \rangle \tag{A1}$$

where $\langle ... \rangle$ indicates an average over all realisations of the experiment.

In order to evaluate the expected value for the *n*th trial, it can first be expressed as the sum of contributions for the two trial types - active with probability *a* and control with probability $1-a$. Then these contributions can be broken down further according to the value of $N_n^{(A)}$, defined as the number of active trials in the session, with the *n*th trial excluded:

$$\begin{aligned} \langle Y_n \rangle &= a \left\langle \frac{X_n}{N^{(A)}} \middle| n \in A \right\rangle - (1-a) \left\langle \frac{X_n}{N - N^{(A)}} \middle| n \in C \right\rangle \\ &= a \sum_{k=0}^{N-2} p_A(N_n^{(A)} = k) \frac{1}{k+1} \left\langle X_n \middle| N_n^{(A)} = k \right\rangle \\ &\quad - (1-a) \sum_{k=1}^{N-1} p_C(N_n^{(A)} = k) \frac{1}{N-k} \left\langle X_n \middle| N_n^{(A)} = k \right\rangle \end{aligned} \tag{A2}$$

In this equation, $\langle ... | n \in A \rangle$ is an average over all realisations of the experiment in which the *n*th trial is of type *A* (and similarly for type *C*), $\langle ... | N_n^{(A)} = k \rangle$ is an average over all realisations in which $N_n^{(A)}$ is equal to *k*, and $p_A(N_n^{(A)} = k)$ is the probability that $N_n^{(A)}$ is equal to k, given that the *n*th trial is of type *A* (and similarly for type *C*).

Because the trial types are assumed to be determined randomly and independently, the probability distributions p_A and p_C are versions of the binomial distribution for the $N-1$ remaining trials, excluding the *n*th, modified to exclude cases in which all the trials in the session are of the same type (see discussion in Section 2.2). If the number of trials in the session, *N*, is large, such cases are extremely unlikely and give rise to modifications that are exponentially small in *N*. So for large *N* we can ignore these modifications and simplify the last equation, to obtain

$$\langle Y_n \rangle \approx \sum_{k=0}^{N-1} f_B(k; N-1, a) \left(\frac{a}{k+1} - \frac{1-a}{N-k} \right) \left\langle X_n \middle| N_n^{(A)} = k \right\rangle \tag{A3}$$

in which the f_B are simply the binomial probabilities that *k* trials are active out of the $N-1$ remaining trials, excluding the *n*th.

The values of *k* in equation (A3) are binomially distributed with expected value $a(N-1)$ and

standard deviation $\sqrt{a(1-a)(N-1)}$. When N is large, this means that the values of k with significant probabilities are concentrated in a region of size $N^{1/2}$, much smaller than the overall range of the sum, which runs from 0 to $N-1$. An approximate form for the quantity in parentheses in (A3) can be found by using a rescaled variable κ, reflecting the smaller scale $N^{1/2}$, defined by

$$k = a(N-1)+\kappa N^{1/2} \tag{A4}$$

This gives

$$\frac{a}{k+1} - \frac{1-a}{N-k} = -\frac{N^{-3/2}}{a(1-a)}\left(\kappa+(1-2a)\left[1-\frac{\kappa^2}{a(1-a)}\right]N^{-1/2}+O\left(N^{-1}\right)\right) \tag{A5}$$

in which the O(N^{-1}) notation indicates that the omitted terms are relatively smaller than the leading term by a factor proportional to N, when N is large.

The conditional expectation of X_n in equation (A3) also varies with k. If N is large, it is natural to expect it to vary as a function of the overall fraction of the trials that are active, that is, as a function of k/N. But as κ varies, k/N changes by only a small amount, comparable with $N^{-1/2}$. We can therefore use a linear approximation, of the form

$$\left\langle X_n \middle| N_n^{(A)}=k \right\rangle = \mu_n+\kappa N^{-1/2}\mu_n'+O\left(N^{-1}\right) \tag{A6}$$

where μ_n and μ_n' are constants. The first, μ_n, is simply the unconditional expected value of X_n, taking all possibilities for k into account, while the second expresses the rate of change of the expected value of X_n as the fraction of active trials increases. Equivalently, μ_n' can be defined as follows, in terms of the limiting value of a sum, as N becomes large:

$$\mu_n' = \lim_{N\to\infty}\frac{1}{a(1-a)}\sum_{k=0}^{N-1} p_B(k;N-1,a)\left(k-a(N-1)\right)\left\langle X_n \middle| N_n^{(A)}=k \right\rangle \tag{A7}$$

Using the approximations (A5) and (A6) in (A3) gives an expression for $\langle Y_n \rangle$ in terms of sums of the binomial probabilities multiplied by powers of κ. As κ is simply a rescaled version of the binomially distributed variable k, these sums are straightforward to evaluate, giving

$$\left\langle Y_n \right\rangle = -N^{-2}\mu_n'+O\left(N^{-3}\right) \tag{A8}$$

Finally, using this approximation in (A1) gives equation (3) for the expected value of the statistic U.

Variance of *U*

Equation (1) for the statistic *F* defined for a single session gives

$$\left\langle F^2 \right\rangle = \sum_{n=1}^{N} \left\langle Y_n^2 \right\rangle + 2 \sum_{n=1}^{N} \sum_{m=1}^{n-1} \left\langle Y_m Y_n \right\rangle \tag{A9}$$

In the large-*N* approximation, the terms of the two sums in this equation can be obtained using methods similar to those above. To find $\langle Y_n^2 \rangle$ only the leading-order approximation for large *N* is required, and this gives

$$\left\langle Y_n^2 \right\rangle = \frac{N^{-2}}{a(1-a)} \left(\mu_n^2 + \sigma_n^2 \right) + O\left(N^{-3} \right) \tag{A10}$$

The procedure for $\langle Y_m Y_n \rangle$ is rather more complicated. There are four cases to be considered, depending on the types of both the *m*th and *n*th trials, and it is now necessary to work in terms of sums containing the binomial probabilities that *k* trials are active out of the $N-2$ remaining trials, excluding both the *m*th and the *n*th. As above, the sums can be evaluated approximately by using a rescaled variable, this time centred on the expected number of active trials, $a(N-2)$. After some working, this gives

$$\left\langle Y_m Y_n \right\rangle = -N^{-3} \left(\frac{\beta_{mn}^{(A)}}{a} + \frac{\beta_{mn}^{(C)}}{1-a} + \beta_{mn}^{(A)\prime} - \beta_{mn}^{(C)\prime} \right) + O\left(N^{-4} \right) \tag{A11}$$

in which $\beta_{mn}^{(A)}$ is the expected value of $X_m X_n$, conditional on the *m*th trial being active (and $\beta_{mn}^{(C)}$ is defined similarly). The β' quantities reflect the rates of change of $\beta_{mn}^{(A)}$ and $\beta_{mn}^{(C)}$ as the fraction of active trials increases, and are defined by equations analogous to (A7).

Equation (A11) can be simplified considerably by making the reasonable assumption, discussed in Section 2.2, that if the difference $n-m$ is large, then the influence of the *m*th trial on the *n*th becomes negligible, and the difference between $\beta_{mn}^{(A)}$ and $\beta_{mn}^{(C)}$ can be neglected, so that the superscripts (*A*) and (*C*) can be dropped. This gives

$$\left\langle Y_m Y_n \right\rangle = -\frac{N^{-3}}{a(1-a)} \beta_{mn} + O\left(N^{-4} \right) \tag{A12}$$

Combining equations (A9), (A10) and (A12) and noting that $\langle F \rangle^2$ is asymptotically smaller than $\langle F^2 \rangle$, the variance of *F* is found to be

$$\mathrm{Var}(F) = \frac{N^{-1}}{a(1-a)} \left(\overline{\mu^2} + \overline{\sigma^2} - \overline{\beta} \right) + O\left(N^{-2} \right) \tag{A13}$$

Alternatively, Var(*F*) can be expressed in terms of the average covariance of X_m and X_n, to give equation (5).

A.2. Approximately normal distribution of the statistics $U^{(P)}$ and V_n

When the number of sessions, R, is large, it is straightforward to see that the statistics U and W defined in Sections 2.2 and 2.5 are approximately normally distributed, because they are the sums of large numbers of independent, identically distributed variables, namely the contributions from individual sessions.

The same argument cannot be applied directly to the statistic V_n defined in Section 2.4, because its definition contains coefficients which depend on the types of the nth trials in all the sessions. Therefore the contributions of different sessions are not independent.

However, if a situation is first considered in which the types of the nth trials in all the sessions in the experiment are fixed, while the other trial types are chosen randomly, then the contributions of different sessions to V_n are independent. In fact, V_n is the difference between the average of the X_n over sessions for which the nth trial is active, and the average over sessions for which it is a control. Moreover since the types of the preceding trials are chosen randomly, on the null hypothesis the X_n are not only independent, but identically distributed, with expected value μ_n and variance σ_n^2. Therefore if R is large, V_n is the difference between two variables, which are both approximately normal, with expected values μ_n and variances $\sigma_n^2/R^{(A)}$ and $\sigma_n^2/(R-R^{(A)})$, where $R^{(A)}$ is the total number of active trials at the nth position of a session. Thus V_n is approximately a normal variable with an expected value of zero and a variance of $R\sigma_n^2/[R^{(A)}(R-R^{(A)})]$.

It remains to consider all the possible types for the nth trials of the sessions. In effect, this means the variable V_n is chosen at random from one of these approximately normal distributions with variances given by $R\sigma_n^2/[R^{(A)}(R-R^{(A)})]$. In this random choice of distribution, the value of $R^{(A)}$ is binomially distributed (with cases for which $R^{(A)}=0$ or R excluded, but the effect of this is exponentially small in R). Because R is large, significant contributions come only from values of $R^{(A)}$ close to aR. Therefore, when the variation of trial types is taken into account, the overall distribution of V_n remains approximately a normal one with an expected value of zero and a variance of $\sigma_n^2/[a(1-a)R]$.

For the statistic $U^{(P)}$ defined in Section 2.3 the situation is slightly more complicated. As for V_n, the contributions of different sessions are not independent. So, similarly, it is necessary to consider first a situation in which all the trial types in the experiment are fixed, so that the contributions of different sessions are independent, though not identically distributed. If R is large, according to the Lyapunov form of the Central Limit Theorem the sum of these contributions is approximately a normal variable whose expected value is equal to the sum of the expected values of the individual contributions, and variance equal to the sum of the variances. The sum of the expected values is not zero, but a careful analysis shows that if, in addition to R being large, the number of trials per session, N, is also large, then the sum of the expected values is small compared with the standard deviation of $U^{(P)}$. This is sufficient to allow the non-zero expected value to be neglected, so that the same reasoning as above can be applied, showing that when all the possible trial types are taken into account $U^{(P)}$ remains approximately normally distributed.

A.3. Statistic *W** based on a difference of rescaled sums, with a baseline estimate subtracted (unbiased)

From equations (19-23), an exact expression for the variance of the statistic *W** can be obtained:

$$\mathrm{Var}\left(W^*\right) \;=\; \frac{R^{-1}\left(N-M\right)^{-1}}{a\left(1-a\right)}\left(\;\overline{s^2}^{(+)} \;+\; \frac{1}{M}\;\overline{s^2}^{(-)} \;+\; \mathrm{E}\left(\;\overline{\left(B-\overline{B}^{(-)}\right)^2}^{(+)}\right)\right) \tag{A14}$$

In this equation, the average denoted by $\overline{}^{(+)}$ is restricted to the range $n=M+1$ to *N,* by analogy to the average $\overline{}^{(-)}$ over the range $n=1$ to *M* defined by equation (19).

In order to make further progress, an approximation can be sought when *N* is large. Suppose that $1\ll M\ll N$, and that the constants s_n and B_n vary only gradually over the whole session, so that for $n\le M$ they can be approximated by expressions of the form

$$G_n \;=\; G_1 \;+\; G_1{}'\left(n-1\right)N^{-1} \;+\; O\left(\left(n-1\right)^2 N^{-2}\right) \tag{A15}$$

Then from the definition (19) of the average $\overline{}^{(-)}$ it follows that

$$\overline{G}^{(-)} \;=\; G_1 \;+\; \tfrac{1}{2}G_1{}'\,M\,N^{-1} \;+\; O\left(N^{-1},M^2N^{-2}\right) \tag{A16}$$

in which the first error term comes from replacing $\overline{(n-1)}^{(-)}$ by $\frac{1}{2}M$.

Also, from the definitions of the three averages,

$$\overline{G}\,N \;=\; \overline{G}^{(-)}M \;+\; \overline{G}^{(+)}\left(N-M\right) \tag{A17}$$

Then equation (A16), gives

$$\overline{G}^{(+)} \;=\; \frac{1}{N-M}\left(\overline{G}\,N \;-\; G_1M \;-\; \tfrac{1}{2}G_1{}'M^2N^{-1} \;+\; O\left(M\,N^{-1},M^3N^{-2}\right)\right) \tag{A18}$$

If equations (A16) and (A18) are used to approximate the averages in (A14), after some working it is found that the variance *W** is minimised when

$$M \;=\; \frac{s_1N^{1/2}}{\left(2\,\overline{s^2} \;-\; s_1^2 \;+\; 2\,\mathrm{E}\left(\overline{\left(B-B_1\right)^2}\right) \;-\; \mathrm{E}\left(\left(\overline{B}-B_1\right)B_1{}'\right)\right)^{1/2}} \;+\; O\left(1\right) \tag{A19}$$

When N is large, provided all the terms in the denominator are of comparable size, this is consistent with the assumption made above that $1\ll M\ll N$. Putting this value into equation (A14) gives the leading-order approximation (24) for the variance of *W** when the optimal value for *M* is used. In fact, that leading-order approximation does not depend on the precise value of M used, but remains valid provided $1\ll M\ll N$.

Finally, it may be noted that: (1) if the terms containing B in the denominator in equation (A19) are small compared with those containing s, then the assumption $1 \ll M \ll N$ is still satisfied and equation (24) remains valid at leading order, and (2) if the terms containing B are large compared to those containing s, then although M may fall to a value comparable with 1, and the assumption $M \gg 1$ may be violated, equation (24) will still remain valid at leading order, because the $\overline{s^2}$ contribution will become negligible. In the latter case it will be sufficient to estimate the baseline using equation (19) with $M = 1$.

A.4. Extension of the statistic *W* to non-independent trial types

In deriving most of the results presented here, it is assumed that the type of each trial is determined randomly and independently of the types of the other trials. However, the definition of the unbiased statistic W (in Section 2.5) can be extended straightforwardly to the situation in which the trial types within each session are not independently chosen.

Consider a situation in which the sequence of trial types in each session is drawn from a prescribed pool of such sequences. Either the sequence for each session may be chosen randomly and independently from the pool, or every sequence in the pool may be used consecutively. For example, a counterbalanced design may be used in which every session is constrained to contain the same number of active trials.

To deal with non-independent trial types, equations (14) and (15) defining the statistic W can be generalised as follows.

For any given trial, the probability a that the trial is active, which is simply a constant when the trial types are independent, is replaced by a variable a_n. This depends on the sequence of earlier trial types within the same session, and is equal to the fraction of sequences in the pool which match the earlier trial types and in which the nth trial is active. In other words, a_n is the probability that the nth trial is active, given the types of the previous trials in the session.

It is necessary to exclude data from trials in which the trial type is absolutely determined by the types of the previous trials (for example, in the counterbalanced design mentioned above, if the prescribed number of active trials has already been used up, so that all the remaining trials must be controls). That is, data must be excluded whenever they come from trials for which either $a_n = 0$ or $a_n = 1$. This means that, for each value of n, the data from the nth trial will be included only for a certain fraction of the sequences in the pool, which will be denoted by b_n (that is, the fraction of the sequences for which $0 < a_n < 1$). With these definitions, for a given pool of sequences, a_n depends on the value of n and the types of the previous trials in a particular session, but b_n depends only on the value of n.

Then the definition (15) of Z_n is replaced by

$$Z_n = \begin{cases} 0 & \text{if } a_n = 0 \text{ or } a_n = 1 \\ \dfrac{X_n}{a_n b_n N} & \text{if } 0 < a_n < 1 \text{ and trial } n \text{ is of type } A \\ -\dfrac{X_n}{(1-a_n) b_n N} & \text{if } 0 < a_n < 1 \text{ and trial } n \text{ is of type } C \end{cases} \qquad \text{(A20)}$$

With this definition, it is straightforward to show that the variable *H* defined for each session by equation (14) still has the required behaviour. The presence of b_n in the denominator in (A20) compensates for the omission of data from trials where the trial type is absolutely determined by previous trial types. Therefore in the presence of a real presentiment effect, the expected value of *W* would still reflect the difference between averages of the values of *X* in active and control trials. But in this case the averages are performed only over the trials included in the definition of Z_n, excluding those with $a_n = 0$ or $a_n = 1$.

On the null hypothesis, that is in the absence of a presentiment effect, the expected value of *W* is zero. Thus the generalised form of *W* remains unbiased.

The variance of *W* is rather more complicated than in equation (17). It can be shown that

$$\mathrm{Var}(W) = R^{-1} N^{-2} \sum_{n=1}^{N} \frac{1}{b_n} \left\langle \frac{X_n^2}{a_n (1-a_n)} \right\rangle^{(Z)} \qquad \text{(A20)}$$

The superscript (Z) indicates that the expected value is calculated only over trials included in the definition of Z_n, excluding those with $a_n = 0$ or $a_n = 1$. The omission of these trials increases the variance somewhat, as reflected by the appearance of b_n in the denominator. This expression correctly agrees with equation (17) when $b_n = 1$ and a_n is a constant.

The generalised form of *W* also remains approximately normally distributed when *R* is large.

References

Jan Dalkvist, Joakim Westerlund and Dick J. Bierman (2002). A Computational Expectation Bias as Revealed by Simulations of Presentiment Experiments. *The Parapsychological Association 45th Annual Convention: Proceedings of Presented Papers*, 62-79.

Jan Dalkvist, Julia Mossbridge and Joakim Westerlund (2014). How to Remove the Influence of Expectation Bias in Precognition Experiments: A Recommended Strategy. *Journal of Parapsychology*, 78(1), 80-97.

Michael Duggan and Patrizio E. Tressoldi (2018). Predictive Physiological Anticipation Preceding Seemingly Unpredictable Stimuli: An Update of Mossbridge's et al. Meta-Analysis. *SSRN*, Number 3097702.

J. E. Kennedy (2013). Methodology for Confirmatory Experiments on Physiological Measures of Precognitive Anticipation. *Journal of Parapsychology*, 77(2), 237-248.

J. E. Kennedy (2014). Correspondence. *Journal of Parapsychology*, 78(2), 273-274.

Julia Mossbridge, Patrizio Tressoldi and Jessica Utts (2012). Predictive physiological anticipation preceding seemingly unpredictable stimuli: a meta-analysis. *Frontiers in Psychology*, 3, Article 390.

Jiří Wackermann (2002). On cumulative effects and averaging artefacts in randomised S-R experimental designs. *The Parapsychological Association 45th Annual Convention: Proceedings of Presented Papers*, 293-305.