Journal of Parapsychology, 77(2), 237-248.

METHODOLOGY FOR CONFIRMATORY EXPERIMENTS ON PHYSIOLOGICAL MEASURES OF PRECOGNITIVE ANTICIPATION

By J. E. Kennedy

ABSTRACT: Research on physiological measures of precognitive anticipation or presentiment is transitioning from exploratory to confirmatory methodology. Appropriate confirmatory practices include (a) using the physiological measures to predict the outcomes of random events with the prediction criteria developed from previous data, (b) prospectively developing and validating the programming for processing the physiological data, (c) using only data prior to the random stimulus on a trial and no data from subsequent trials when developing parameters for adjustments or artifact rejections, and (d) multiple experimenter designs that make misconduct by one experimenter difficult. The physiological measures in precognitive anticipation experiments can be expected to violate the assumption of independence between trials when used as the dependent variable and may produce counterintuitive, false positive biases with standard statistical methods. The most convincing research strategy is to develop prediction criteria using an initial set of data and then apply the criteria to predict the random events on new trials. Other research strategies can be expected to be controversial. In addition, when the physiological values used for analysis are derived after trials with feedback have been completed, data processing must be handled very carefully to avoid bias from retrospective selection of data processing parameters.

Keywords: presentiment, precognitive anticipation, dependent variable dependence, expectation bias, retrospective selection, ESP

Most studies of physiological measures of precognitive anticipation have been exploratory, and the time has come to apply methods for convincing confirmatory research. Mossbridge, Tressoldi, and Utts (2012) recently pointed out the need for the diverse experimental procedures and analysis methods to converge to standard practices that can be directly replicated.

This article focuses on certain methodological issues that require particular attention for precognitive anticipation experiments. The discussion here does not cover all aspects of confirmatory experiments. Important methodological practices that should be implemented for any confirmatory experiment but are not discussed here include power analysis, study registration, documented software validation, and data sharing (Kennedy, 2013a, 2013b, 2013c; Koestler Parapsychology Unit, 2012). The present article focuses on issues that are more specifically associated with precognitive anticipation experiments.

The basic design of a precognitive anticipation experiment is that a participant receives randomly selected stimuli while certain physiological data are collected prior to the stimuli. The physiological data are analyzed to see if the person unconsciously anticipates the specific stimulus that occurs. Various stimuli can be used, such as the display of either an arousing image or a calming image, or either a signal that requires a fast action by the participant or a signal that requires no action. A wide range of physiological measures can indicate anticipation in situations like these, including skin conductance, electrical activity in the brain, pupil dilation, muscle activity, and heart rate. The term *presentiment* has often been used for experiments using emotional or arousing stimuli, but that is a subset of precognitive anticipation research.

Three basic strategies have been used to analyze precognitive anticipation experiments. One analysis strategy is to *predict new events*. An initial set of data is analyzed to develop criteria for using the physiological measures to predict the random events. The methods for developing criteria for making predictions are often called classification or discriminant analysis methods. These methods may be simple such as using a median value, or may involve complex multivariate techniques. The initial process of de-

veloping criteria is often called the *learning* or *training* step. The criteria are then applied to physiological measures on new trials to predict the random events for those trials. Statistical significance for precognitive anticipation can be evaluated with a simple binomial test (or normal approximation) on the proportion of correct predictions for the new trials.

Another analysis strategy is to classify the learning data. Predictive criteria are developed as described above, but statistical significance is evaluated by applying the criteria to the data used to develop the criteria rather than to new trials. This strategy must attempt to adjust for the extent to which the process of developing the criteria incorporates random fluctuations and other properties of the learning data that are not applicable for future random events. These adjustments generally are not straightforward, particularly when multivariate methods are involved. The most convincing way to evaluate the validity of the predictive criteria is to apply the criteria to new trials as described above. Attempts to eliminate that step can be expected to be controversial.

The third analysis strategy is to evaluate the *differences in the physiological measures*. Statistical significance is based on testing the difference between the average physiological measures for the different types of stimuli in the study. This strategy uses the physiological measures (rather than the random events) as the dependent variable. As discussed in the next section, this strategy is prone to false positive biases because the physiological data can violate the assumptions for standard statistical analysis as a dependent variable.

The topics discussed in this article all point to the conclusion that predicting new events is the optimal strategy for analyzing precognitive anticipation experiments. The results are highly convincing if the predictions are made prior to generating the random stimuli, and particularly with appropriate multiple experimenter study design. The analysis strategies of differences in physiological measures and classifying the learning data can be expected to be controversial.

It may be helpful to consider how these three strategies would apply to the evaluation of a person claiming to be able to beat the odds and win money at a gambling game in a casino. The strategy of predicting new events is a direct, unambiguous measure of whether the person actually can reliably beat the odds and win money. The strategy of classifying the learning data is analogous to the person recording some outcomes for the game, then retrospectively developing an algorithm that could have been used to beat the odds for that particular set of data—and then arguing that this shows success even though actually beating the odds and winning money was not directly demonstrated. The analysis strategy of differences in physiological measures is equivalent to the person arguing that larger average bets on the trials that won than on the trials that lost is evidence for success. However, this difference between average winning and losing bets can be achieved with certain betting strategies that do not actually produce net winnings. For example, increasing the amount of the bet on each trial until a winning bet is obtained, and then starting over with a low bet on the next trial tends to produce this difference. With this betting strategy, the largest winning bets occur after a series of losing bets. The average bet for the winning bets may be larger than for the losing bets, but the averages do not indicate whether the total amount won exceeded the total amount lost. This betting strategy does not beat the odds, but tends to produce the differences claimed to indicate success. Here too, actually beating the odds and winning money is not directly and unambiguously demonstrated. The strategy of increasing the bet on each trial until a win is obtained is similar to the gambler's fallacy discussed below.

Dependencies Between Trials

In recent years, the analysis strategy of differences in physiological measures has been frequently used to analyze precognitive anticipation experiments. The evaluation has been based on the difference between the average values of the physiological measures for the different stimuli rather than evaluating the proportion of new trials when the physiological measure correctly predicted which random stimulus occurred.

Physiological measures of anticipation often have dependencies between trials that make them

prone to false-positive artifacts when used as the dependent variable. The gambler's fallacy is the well-known tendency for people to anticipate that the random event on the next trial will have a different outcome than on the previous trial. The degree of anticipation often increases when the same outcome occurs several times in sequence. This introduces dependencies between sequential trials.

Standard statistical methods are based on the assumption that the value for the dependent variable on a trial is independent of the values of the variables on other trials, except for factors that are included in the statistical model. Violation of this assumption can produce misleading results. One of the simplest forms of dependency is a positive serial correlation between the values of the dependent variable on consecutive trials. This dependency causes the *p* values for standard statistical hypothesis tests to be misleadingly significant (false positive) and makes the standard tests inappropriate (Miller, 1986; Neter, Wasserman, & Kutner, 1985, p. 445). It is important to note that the error variance is incorrect if standard statistical methods are applied, and that simply adding a covariate with the value of the dependent variable on the previous trial does not produce accurate hypothesis tests.

With the gambler's fallacy, the anticipation measure on a precognitive anticipation trial depends on an interaction between the stimulus on the previous trial and the anticipation measure on the previous trial. This dependency between trials violates the assumptions for analyzing a dependent variable using standard statistical methods. Unfortunately, the effects of this violation are not easy to discern.

If the physiological measure is used as the dependent variable, determining the effects of the dependencies between trials requires detailed analyses. These dependencies can be considered a type of expectation bias and that term has been used in some writings. Wackermann (2002) used numeric and analytic approaches to investigate the problem and Dalkvist and Westerlund (2006; Dalkvist, Westerlund, & Bierman, 2002) used paper and pencil models and computer simulations. A simple simulation model that offers insight into the nature and implications of the dependencies between trials is described in the Appendix of the present paper.

These analyses found that the dependencies between trials can cause artificial differences between the average physiological values for different stimuli in ways that mimic precognitive anticipation. As described in the Appendix, these differences can create the illusion of precognitive anticipation when the actual anticipation is incorrect for the majority of trails. Certain aspects of these biases are notably counterintuitive.

It is not safe to assume that dependencies between trials can be ignored if the measures of anticipation do not become stronger during a string of trials with the same stimulus. As described in the Appendix, the simulations show that biases can occur in that situation.

Similarly, the currently available analyses do not justify confidence in the assumption that dependencies between trials can be safely ignored if data are pooled from different participants. The available analyses are proof of concept rather than justification for technical guidelines for hypothesis testing. These analyses focus on biases for the mean values, which is useful for demonstrating that biases occur. However, the development of guidelines for hypothesis testing would require evaluating the biases for p values when standard methods are applied. That is a much more complicated evaluation that involves the bias for the error variance as well as the bias for the mean. Also, as noted in the Appendix, different forms of dependencies between trials need to be considered, not just one of the weaker forms as investigated by Dalkvist and Westerlund (2006).

Using physiological measures as the dependent variable is particularly problematic for process-oriented research that evaluates factors associated with better precognitive anticipation. In general, the biases from dependencies between trials for anticipation measures are reduced as larger amounts of data are pooled. However, process-oriented analyses evaluate subsets of data and therefore have higher potential for biased results. The unit of analysis in process-oriented research is typically the participant. As Dalkvist and Westerlund (2006; Dalkvist, Westerlund, and Bierman, 2002) pointed out, the number of trials for each participant in a typical precognitive anticipation study cannot be expected to overcome the biases. Determining whether the effects in process-oriented research are due to differences in precognitive antic-

ipation or to differences in the dependencies between trials would be difficult.

Extensive research would be needed to develop a convincing working understanding of the effects of dependencies between trials in precognitive anticipation studies using physiological measures as the dependent variable. The complexity of the biases is significantly increased by the often highly skewed nature of physiological data and the possibility that the dependencies may vary across people and during an experimental session, as well as in response to different instructions and tasks. The effects of these complexities would need to be understood through analyses.

False positive biases also manifest with the analysis strategy of classifying the learning data. The criteria for the predictions or classifications are derived from a group of trials and then applied to the trials in that group. The retrospectively developed criteria were adapted to the specific targets for the group of trials. The dependencies between trials may contribute to the misleading effectiveness of the criteria in this situation.

The safest, most convincing research strategy is to predict new trials using criteria developed from previous data. If randomization is handled properly, the stimulus on a new trial is independent of the stimuli and the physiological measures on previous trials. This analysis strategy does not raise questions about violating the basic assumptions for standard statistical methods and can be used reliably for process-oriented research. The new trials could be with the same participants as the initial learning data or with different participants. Another interesting strategy is to have each participant do only one trial (Dalkvist, Mossbridge, & Westerlund, 2013; Mossbridge, 2013). This eliminates the dependencies between trials, but the validity of the results can still be expected to be controversial unless criteria can be developed that successfully predict new trials.

Realistically, the analysis strategies of differences in physiological measures and classifying the learning data may be best considered as exploratory efforts in the development of methods for predicting random events on new trials. These analysis strategies raise extremely complicated, counterintuitive technical issues that most cautious scientists will find to be more plausible than psi for explaining significant results. The Appendix has additional discussion of these technical issues.

Using the physiological measures to predict the random events on new trials is also more relevant for the practical application of precognitive anticipation. For example, if precognitive anticipation were developed for applications such as airplane pilots anticipating an emergency response, the physiological measures would be used to predict the event.

Processing the Physiological Data

In most precognitive anticipation experiments, the physiological values used in the analysis are derived from relatively complex processing of the raw data. The raw data are seldom used directly. This data processing usually includes modifications described with terms like normalization, baseline adjustment, and/or artifact rejection. Decisions must be made about the parameters for this processing. In addition, decisions must be made about which of many different options will be used to reduce the physiological data for a trial to one number for analysis.

In many cases, the processing of the physiological data to obtain the values for analysis is done after the trials have occurred and feedback has been given. That is very different from a traditional ESP experiment, which has the call or prediction for a trial unalterably registered before the target is revealed. Obtaining the physiological values by data processing after the targets are known introduces significant potential for bias. This potential for bias applies whether the physiological values are used as predictor variables or as dependent variables.

Retrospective selection of data processing parameters that produce favorable results can occur through several mechanisms during data processing. If the processing of the physiological data involves decisions by an analyst, these decisions should be made with the analyst blind to the random events. Blinded evaluations eliminate the almost irresistible temptations to introduce subtle biases, as well as more overt data manipulations. For confirmatory studies, decisions about data adjustments or artifact re-

jection can normally be automated with programming that does not involve the random events.

However, even when the processing is automated, the derived physiological values are susceptible to bias from retrospective data processing. The physiological data after the random stimulus on a trial typically contain information about which stimulus occurred. Analysts need to be vigilant that any normalization, baseline adjustment, artifact rejection, or other processing of the physiological data does not involve data after the stimulus on a trial.

An adjustment can introduce biases if the parameters are derived from a group of trials and the adjustment then applied to the trials in the group. The biases described in the Appendix are applicable when making adjustments to the data as well as in the final statistical analysis. The physiological data for trials with feedback may have dependencies between trials that reflects information about the targets and introduces subtle, counterintuitive biases when parameters are retrospectively derived. For example, if an adjustment to the physiological data is based on the mean of a group of trials, the mean used for a particular trial should be based on prior trials only with no data from subsequent trials.

Any adjustment of the pre-stimulus physiological data that involves post-stimulus data, including from subsequent trials, has the potential to compromise the integrity of the pre-stimulus data. Biases can result from technical details that are difficult to identify from published reports. This issue may be handled more casually in exploratory research, but confirmatory experiments should manage the processing of the physiological data very carefully.

The optimal analysis strategy is for prospectively developed automated programming to predict the random stimulus for a trial using only physiological data prior to the random event. Ideally, the prediction would be made before the random stimulus and any physiological data after the stimulus have been generated or read by the programming. Here too, the most convincing research strategy is to develop predictive criteria with initial data and apply those criteria to predict the random stimuli on new trials. To verify that biases did not occur, the number of trails rejected due to artifacts should be reported and also the number of each type of stimulus on the rejected trials.

Multiple Experimenter Designs

Study designs with procedures that involve multiple experimenters and make misconduct by any one person difficult are highly valuable for confirmatory experiments in a controversial area of research. Experimenter misconduct has occurred many times in parapsychology and is a constant threat (Kennedy, 2013b). Experimenter fraud has occurred in all areas of science. However, the controversial nature of psi research combined with prominent experimenter differences in producing effects make experimenter misconduct particularly salient in parapsychology.

Contrary to what many scientists assume, independent replication and peer review generally have not been effective at detecting or deterring scientific fraud. Most frauds have been detected by coworker whistleblowers. These conclusions are supported by the experiences in parapsychology (Kennedy, 2013b) and in other areas of science (Strobe, Postmes, and Spears, 2012). Multiple experimenter study designs recognize the importance of coworkers in preventing misconduct.

The highly automated methods used in the precognitive anticipation studies typically appear to have been developed without consideration of multiple experimenter designs. For most studies, it would appear to be relatively easy for one experimenter to make a version of the data collection program that fraudulently manipulates the data in a way that would be very difficult or impossible to detect later.

The experimental procedures could be adapted to involve multiple experimenters in the generation of random events and the collection of physiological data. The optimal strategy is to have copies of the random event data and the physiological data held by two different experimenters prior to unblinding. This prevents any one experimenter from easily altering the data in an undetectable manner. Here too, multiple experimenter study designs can be most effectively implemented with the analysis strategy of predicting new events.

One option would be for the computer program that conducts the experiment to obtain random

numbers from a remote site through the internet or a more direct connection. The program would collect the pre-stimulus anticipatory physiological data for a trial, then transfer a copy of that data to the remote site and obtain the random number for the type of stimulus. A copy of the random numbers used in the experiment, or the parameters for a random algorithm, would previously have been transferred to another experimenter or to a third party. With this procedure, the data for both the random stimuli and the physiological measures are held in two independent locations prior to unblinding for a trail. After the initial development of such systems, they could be routinely implemented with little overhead.

Another strategy would include control trials in which the random stimulus was not displayed. The program conducting the experiment would collect data and prepare to display the randomly selected stimulus, but would obtain a random number from a remote site immediately prior to displaying the stimulus. Based on the random number, the stimulus would not be displayed on certain trials that would serve as a control condition. Any unexpected effects on the control trials would presumably indicate experimenter effects of some type rather than precognitive anticipation by the participant. This strategy might be useful for investigating psi-mediated experimenter effects as well as for preventing undetected experimenter misconduct.

Recommendations

Experiments on physiological measures of precognitive anticipation have often had two major methodological differences from traditional ESP experiments. First, traditional ESP experiments use the random events as the dependent variable, whereas precognitive anticipation studies have often used the physiological measures as the dependent variable. Second, traditional ESP experiments have the calls or predictions unalterably registered before the target is revealed, whereas precognitive anticipation studies have often obtained the physiological values used in the analysis by relatively complex data processing after the trials were completed and feedback given. These methodological practices in precognitive anticipation studies introduce significant potential for subtle, counterintuitive biases that are difficult to detect and virtually assure that the results will be controversial.

The recommendations for confirmatory experiments on studies of physiological measures of precognitive anticipation are:

1. The random events can and should be used as the dependent variable, with the prediction criteria developed from previous data.
2. The computer program for processing the physiological data should be programmed and validated prior to collecting the data in order to limit the potential for retrospective selection or modification of data processing parameters.
3. All processing of the physiological data should be very carefully managed to eliminate any criterion, parameter, or adjustment that is derived using any data after the feedback on a trial, including any data from subsequent trials.
4. Experiments should be designed with procedures that involve multiple experimenters and make it difficult for any one experimenter to make unintentional or intentional mistakes that affect the experimental results.

A planned confirmatory experiment by Patrizio Tressoldi (2012) represents a significant methodological advance for precognitive anticipation research. The physiological measures are used to predict new random events, and the programming for the study makes the prediction for a trial before the stimulus is displayed. In addition, the study was prospectively registered at the Koestler Parapsychology Unit study registry, and the planned sample size was based on power analysis.

This discussion is intended as a starting point for methodological improvement as the research shifts from exploration to confirmation. The factors discussed in this article can be routinely used for study quality ratings in meta-analyses or other data syntheses. Creative thought will likely produce other options

for enhancing methodology in these experiments.

References

Dalkvist, J., Mossbridge, J., & Westerlund, J. (2013). How to handle expectation bias in presentiment experiments: A recommendation. *Abstracts of Presented Papers: The Parapsychological Association 56th Annual Convention*, 19.

Dalkvist, J., & Westerlund, J. (2006). A bias caused by inappropriate averaging in experiments with randomized stimuli. *Journal of Parapsychology*, *70*, 233–254.

Dalkvist, J., Westerlund, J., & Bierman, D. (2002). A computational expectation bias as revealed by simulations of presentiment experiments. *Proceedings of Presented Papers: The Parapsychological Association 45th Annual Convention,* 62–79. Retrieved from http://uniamsterdam.nl/D.J.Bierman/publications/2002/expectationbias_PA2002.pdf

Kennedy, J. E. (2013a). Can parapsychology move beyond the controversies of retrospective meta-analysis? *Journal of Parapsychology*, *77*, 21–35.

Kennedy, J. E. (2013b). Experimenter misconduct in parapsychology: Analysis manipulation and fraud. Retrieved from http://jeksite.org/psi/misconduct.pdf

Kennedy, J. E. (2013c). [Letter to the editor]. *Journal of Parapsychology*, *77*, 00–00.

Koestler Parapsychology Unit. (2012). Registry for Parapsychological Experiments. Retrieved from http://www.koestler-parapsychology.psy.ed.ac.uk/TrialRegistry.html

Miller, R. G. (1986). *Beyond ANOVA: Basics of applied statistics*. New York, NY: Wiley.

Mossbridge, J. (2013). Single-trial presentiment experiment. Koestler Parapsychology Unit Registry for Parapsychological Experiments, Study Registration ID Number 1005. Retrieved from http://www.koestler-parapsychology.psy.ed.ac.uk/TrialRegistryDetails.html

Mossbridge, J., Tressoldi, P., & Utts, J. (2012). Predictive physiological anticipation preceding seemingly unpredictable stimuli: A meta-analysis. *Frontiers in Psychology*, *3* (Article 390), 1–18. Retrieved from http://www.frontiersin.org/perception_science/10.3389/fpsyg.2012.00390/abstrac

Neter, J., Wasserman, W., & Kutner, M. H. (1985). *Applied linear statistical models: Regression, analysis of variance, and experimental design* (2nd ed.). Homewood, IL: Irwin.

Strobe, W., Postmes, T., & Spears, R. (2012). Scientific misconduct and the myth of self-correction in science. *Perspectives on Psychological Science*, *7*, 670–688. Retrieved from http://pps.sagepub.com/content/7/6/670.full.pdf+html.

Tressoldi, P. (2012). Pupil dilation accuracy in the prediction of random events. Koestler Parapsychology Unit Registry for Parapsychological Experiments, Study Registration ID Number 1001. Retrieved from http://www.koestler-parapsychology.psy.ed.ac.uk/TrialRegistryDetails.html

Wackermann, J. (2002). On cumulative effects and averaging artifacts in randomized S-R experimental designs. *Proceedings of Presented Papers: The Parapsychological Association 45th Annual Convention*, 293–305. Retrieved from http://www.jeksite.org/others/jw_2002_pa.pdf

Acknowledgement

I want to thank George Hansen, Julia Mossbridge, and the reviewers for thought-provoking comments that resulted in extensive revisions and improvements from earlier versions.

Appendix

A simple simulation model shows the nature and effects of dependencies between trials for precognitive anticipation experiments. Assume that a person anticipates the occurrence of a random event with two possible outcomes. For example, the outcomes could be the display of either an arousing image or a calming image. As another example, the outcomes could be display of one of two stimuli, with one requiring a fast action by the participant and the other requiring no action. For purposes of this simple

simulation, the outcomes are labeled "outcome-plus" and "outcome-minus."

Also, assume that a physiological measure has different values prior to the random event, depending on which outcome is anticipated and the degree of anticipation. A wide range of physiological measures can indicate anticipation in situations like this, including skin conductance, electrical activity in the brain, pupil dilation, muscle activity, and heart rate. For purposes of this simple simulation, the physiological measure has negative values -1, -2, -3, etc. for increasing anticipation that outcome-minus will occur, and positive values 1, 2, 3, etc. for increasing anticipation that outcome-plus will occur. The basic effects developed with these assumptions apply widely.

In addition, assume that the participant anticipates that the outcome on the next trial will be different than on the previous trial and that the physiological measure of anticipation increases when a sequence of trials with one outcome occurs. This is the well-known gambler's fallacy. For example, if outcome-plus occurs, then the physiological measure on the next trial has value -1, indicating that the person anticipates outcome-minus will occur. If outcome-plus occurs on that trial, then the physiological measure of anticipation become stronger to -2 for the following trial. Increasing negative values of the physiological measure continue until an outcome-minus event finally occurs. In general, the person maintains increasing anticipation for the opposite outcome until that outcome occurs.

If a person increases anticipation for an outcome until that outcome occurs, the final trial in that sequence has the most extreme anticipation value and has correct anticipation. For example, a sequence of five trials with outcome-minus ends when a trial with outcome-plus occurs. The participant has increasing anticipation for outcome-plus over the six trials. In this case, the average physiological value for the five outcome-minus trials is $(1 + 2 + 3 + 4 + 5) / 5 = 3$ and the value for the final outcome-plus trial is 6. The average for the outcome-minus trials is less than for the outcome-plus trials. If the analysis is based on the difference between the average values of the physiological measures for the two outcomes, this result is consistent with precognitive anticipation.

In general, for a sequence of trials with the same direction of anticipation, the final correctly anticipated trial will have a more extreme correct physiological value than the previous incorrectly anticipated trials—which appears to support precognitive anticipation.

Note that this positive result for the difference of the averages occurs even though the physiological measure fails to predict the correct outcome on five of the six trials. However, this point is based on a priori knowledge of the prediction criteria—as would happen if the criteria had been developed previously using different data. If the criteria for making predictions were developed from only the data in this example, the fact that most trials have incorrect anticipation would not be recognized. The criteria would incorporate the biases in the data and would be prone to false positive results when applied to the data used to develop the criteria.

The biases for individual sequences like this would be counterbalanced if all possible sequences were present in the analyses. For example, five trials with outcome-plus followed by a trial with outcome-minus produce averages of -3 for outcome-plus trials and -6 for the final outcome-minus trail. This difference of the averages is consistent with precognitive anticipation. However, when the data for this sequence are pooled with the complementary sequence described above, the pooled averages are 1.5 for the outcome-minus trials and -1.5 for the outcome-plus trials—which is appropriately contrary to the precognitive anticipation hypothesis given that only two of twelve trails have correct anticipation. The data for the two sequences are shown in Table 1.

However, an experiment has a finite number of random events and may not include counterbalancing sequences for all sequences. As indicated by these examples, each of the individual sequences alone erroneously tends to support precognitive anticipation if the physiological measure is the dependent variable.

In a simple simulation of a run of 40 trials with these assumptions for anticipation, I found that the average physiological measures for outcome-plus trials were higher than for outcome-minus trails in 69% of 3000 simulations. The degree of bias was reduced if the number of trials in a run increased and became

Table 1
Data for Complementary Sequences

Data for Original Sequence

Trial number:	1	2	3	4	5	6
Stimulus outcome:	minus	minus	minus	minus	minus	plus
Physiological measure:	1	2	3	4	5	6
Correct prediction:	no	no	no	no	no	yes

Average physiological measure for outcome-minus = (1 + 2 + 3 + 4 + 5) / 5 = 3
Average physiological measure for outcome-plus = 6 / 1 = 6
Difference consistent with precognitive anticipation: yes
Correct physiological predictions: 1 / 6 = 17%
Physiological predictions consistent with precognitive anticipation: no

Data for Complementary Sequence

Trial number:	1	2	3	4	5	6
Stimulus outcome:	plus	plus	plus	plus	plus	minus
Physiological measure:	-1	-2	-3	-4	-5	-6
Correct prediction:	no	no	no	no	no	yes

Average physiological measure for outcome-minus = -6 / 1 = -6
Average physiological measure for outcome-plus = (-1 - 2 - 3 - 4 - 5) / 5 = -3
Difference consistent with precognitive anticipation: yes
Correct physiological predictions: 1 / 6 = 17%
Physiological predictions consistent with precognitive anticipation: no

Data for Pooled Sequences

Average physiological measure for outcome-minus = (1 + 2 + 3 + 4 + 5 - 6) / 6 = 1.5
Average physiological measure for outcome-plus = (6 - 1 - 2 - 3 - 4 - 5) / 6 = -1.5
Difference consistent with precognitive anticipation: no
Correct physiological predictions: 2 / 12 = 17%
Physiological predictions consistent with precognitive anticipation: no

worse if only the trials with more extreme values of the physiological measure were evaluated.

As expected, precognitive anticipation based on the physiological measure predicting the random events using the a priori prediction criteria was at chance in these simulations.

A more counterintuitive result is that biases also occur when the degree of anticipation does not increase during a string of trials with one outcome. My simulations found that about 57% of the simulations were in the direction supporting precognitive anticipation even when the degree of anticipation remained constant over a string of trials with one outcome. The physiological measures were set to +1 or -1 and remained at that level until switched to -1 or +1 after the random outcome changed. This situation retains some of the dependency between trials and apparently also retains some of the false positive bias. I verified this result with three different sources of random numbers. Dalkvist and Westerlund (2006) used similar assumptions for their simulations and reported corresponding effects—which also brings into focus the fact that their findings are based on a type of dependency that produces much weaker effects than when anticipation increases for sequential trials with the same stimuli (57% versus 69% of the simulations in the direction consistent with precognitive anticipation).

One way to gain some insight into this counterintuitive result is to recognize that when the random

events happen to alternate outcomes for some trials, the strategy of anticipating a different outcome from the previous trial produces completely correct results. On the other hand, longer strings such as a sequence of five trials with outcome-minus followed by a trail with outcome-plus results in average physiological values of (1 + 1 + 1 + 1 +1) /5 = 1 for the outcome-minus trails and 1 for the outcome-plus trail. The averages are tied, and with finite amounts of random data the longer strings apparently do not fully offset the results for the trials with alternating outcomes.

It may be useful to comment on the use of permutation tests to evaluate precognitive anticipation studies with the physiological measures as the dependent variable. A permutation test assumes the trials are independent and tests the hypothesis that the physiological value on any trial could have occurred equally likely with the random stimulus on any trial. A significant result indicates that this hypothesis is false and there are differences in the data. However, the test does not indicate why the differences occurred. Significant results could be due to precognitive anticipation or to the trials not being independent. A permutation test intended to exclude biases from the dependencies between trials would include for a trial only the permutations that have the same values for the stimulus and anticipation measure on the previous trial.

Dalkvist and Westerlund (2006; Dalkvist, Mossbridge, & Westerlund, 2013) propose handling the dependencies between trials with a 2-way ANOVA that includes a factor for the stimulus on the previous trial. Putting the values in Table 1 into the cells for the ANOVA and examining the means suggests that the ANOVA is susceptible to biases. It appears to me that attempts to adjust for or partial out the effects from the previous trial would need to include the physiological measure on the previous trial as a term (covariate) as well as the stimulus on the previous trial. However, as with serial correlation between trials, simply adding terms for the values on the previous trial cannot be assumed to provide accurate hypothesis tests.

In general, statistical tests for evaluating the effects of the dependencies are also subject to methodological problems from the dependencies. In addition, for any conclusions based on statistical tests with nonsignificant outcomes, the power of the test is of central importance.

Even mild to moderate skeptics of psi will likely find biases from dependencies between trials to be a plausible explanation for significant results when the analysis is based on differences in physiological measures or classifying the learning data. Great effort would be needed to obtain a useful understanding of the effects of the dependencies with these analysis strategies, and probably greater effort to convince other researchers that adequate understanding had been achieved. Studies that develop prediction criteria on initial learning data and then successfully predict the random events on new trials are much simpler and more convincing. The evidence for precognitive anticipation can be expected to remain controversial until this strategy is reliably demonstrated.

Broomfield, CO, USA
jek@jeksite.org
http://jeksite.org/psi.htm

Abstracts in Other Languages

Spanish

Metodología para Experimentos Confirmatorios sobre Medidas Fisiológicas de Anticipación Precognitiva

RESUMEN. La investigación sobre las medidas fisiológicas de anticipación precognitiva o presentimiento está pasando de ser una metodología exploratoria a una confirmatoria. Las prácticas confirmatorias apropiadas incluyen (a) el uso de medidas fisiológicas para predecir los resultados de eventos aleatorios con criterios de predicción desarrollados a partir de datos anteriores, (b) desarrollo y validación de forma

prospectiva de la programación para el procesamiento de los datos fisiológicos, (c) usar solamente los datos previos al estímulo aleatorio de una prueba y no datos de pruebas posteriores durante desarrollo de parámetros para los ajustes o rechazos de artefactos, y (d) diseños con varios experimentadores para hacer que la mala conducta de un experimentador resulte difícil. Se puede esperar que las medidas fisiológicas en experimentos de precognición anticipatoria violen el supuesto de independencia entre las pruebas cuando se usen como variable dependiente y pueden producir sesgos contra-intuitivos positivos falsos con el uso de métodos estadísticos estándar. La estrategia de investigación más convincente es el desarrollo de criterios de predicción utilizando un conjunto inicial de datos y luego aplicar los criterios para predecir los acontecimientos aleatorios en nuevas pruebas. Se puede esperar que otras estrategias de investigación sean controversiales. además, cuando los valores fisiológicos usados para el análisis se derivan después de que se ha dado feedback después del experimento, el procesamiento de datos debe ser manejado con mucho cuidado para evitar el sesgo de la selección retrospectiva de los parámetros de procesamiento de datos.

French

MÉTHODOLOGIE POUR LES EXPERIMENTATIONS CONFIRMATOIRES SUR LES MESURES PHYSIOLOGIQUES D'ANTICIPATION PRÉCOGNITIVE

RESUME : Les recherches sur les mesures physiologiques d'anticipation précognitive ou de pressentiment sont en train de passer d'une méthodologie exploratoire à confirmatoire. Les pratiques confirmatoires appropriées incluent (a) l'utilisation de mesures physiologiques pour prédire les résultats des événements aléatoires avec les critères de prédiction développés à partir des données antérieures ; (b) le développement prospectif et la validation du programme de traitement des données physiologiques ; (c) l'utilisation uniquement des données avant le stimulus aléatoire sur un essai et non des données des essais subséquents lors du développement des paramètres pour l'ajustement ou le rejet d'artefacts ; et (d) des dispositifs avec de multiples expérimentateurs qui rendent difficile d'expliquer les résultats par la mauvaise conduite d'un seul expérimentateur. On s'attend à ce que les mesures physiologiques des expérimentations d'anticipation précognitive puissent violer l'hypothèse de l'indépendance entre essais lorsqu'elle est utilisée comme variable dépendante, ce qui peut produire des biais contre-intuitifs de faux-positifs avec les méthodes statistiques standards. La stratégie de recherche la plus convaincante est de développer des critères de prédiction en utilisant un ensemble initial de données et d'appliquer ensuite ces critères pour prédire les événements aléatoires sur les nouveaux essais. D'autres stratégies de recherche seront nécessairement controversées. De plus, lorsque les valeurs physiologiques utilisées pour l'analyse sont dérivées après que des essais avec feedback aient été effectués, le traitement des données doit être réalisé de façon très prudente pour éviter les biais de sélection rétrospective des paramètres de traitement des données.

German

EINE METHODOLOGIE FÜR BESTÄTIGUNGSEXPERIMENTE VON PHYSIOLOGISCHEN MESSUNGEN ZUR PRÄKOGNITIVEN ANTIZIPATION

ZUSAMMENFASSUNG: Die Forschung zur physiologischen Messung der präkognitiven Antizipation oder Presentiment befindet sich im Übergang von der explorativen zur Bestätigungsmethodologie. Angemessene Techniken zur Bestätigung umfassen (a) die Verwendung physiologischer Messungen zur Vorhersage von Zufallsereignissen, wobei die Vorhersagekriterien aufgrund früherer Daten formuliert wurden, (b) die prospektive Entwicklung und Validierung von Programmen zur Verarbeitung der

physiologischen Daten, (c) die Verwendung nur solcher Daten, die dem zufälligen Stimulus bei einem Trial vorhergehen und keiner Daten darauffolgender Trials, wenn es um die Entwicklung von Parametern zur Anpassung oder zum Ausschluss von Artefakten geht, und (d) Versuchspläne mit mehreren Experimentatoren, wodurch Fehlverhalten eines Experimentators erschwert wird. Es steht zu erwarten, dass die physiologischen Messungen bei Experimenten zur physiologischen Antizipation die Annahme der Unabhängigkeit der Trials untereinander verletzen werden, wenn sie als abhängige Variable verwendet werden und bei den statistischen Standardmethoden zu contra-intuitiven falsch-positiven Biases führen werden. Die überzeugendste Forschungsstrategie besteht in der Entwicklung von Vorhersagekriterien unter Verwendung eines ursprünglichen Datensatzes, um dann diese Kriterien zur Vorhersage von Zufallsereignissen bei neuen Trials heranzuziehen. Andere Forschungsstrategien dürften sich als kontrovers erweisen. Sobald die in die Analyse eingehenden physiologischen Werte, die von Trials mit Feedback gewonnen wurden, endgültig vorliegen, muss die Auswertung der Daten äußerst sorgfältig geschehen, um einen Bias aufgrund der retrospektiven Selektion der Parameter der Datenauswertung zu vermeiden.