Journal of Parapsychology, 78(2), 273–277

CORRESPONDENCE

To the Editor:

The recent article by Dalkvist, Mossbridge, and Westerlund (2014) on expectation bias in presentiment studies discussed an important methodological problem but included a controversial recommendation and two key comments that are not correct. Presentiment studies investigate whether physiological measures indicate that a person can unconsciously and precognitively anticipate a random stimulus. The most common strategy for analyzing the data has been to compare the average values of the observed physiological measures preceding the different types of random stimuli.

This analysis strategy reverses the traditional analysis for a typical ESP experiment, such as a participant pushing a button to predict which light will be randomly selected. The traditional analysis uses the button press or response to predict the random light or target. As described by Burdick and Kelly (1977, p. 93):

> The response array is taken as fixed (in fact, it is immaterial where it came from, and this underlies the great generality of the method). The statistical problem is to evaluate the probability of obtaining a number of hits as large or larger than that observed, *given the response array*.

The random targets are the outcomes that are assumed to be variable and to follow a probability distribution.

This strategy was developed in the 1940s when it was realized that the responses cannot be assumed to be independent as is required for the outcome variable for standard statistical analysis. One of the most well-known examples is the stacking effect that occurs when a single target sequence is used with multiple participants. The responses are not considered independent because "the respondents may tend to possess shared guessing habits" (Burdick & Kelly, 1977, p. 92). The basic principle is that habits and associated nonindependence may occur within any sequence of responses generated by a human, and among any such sequences, including when feedback is not given. As noted above, any habits are immaterial when the traditional analysis strategy is used. The random targets are the outcome variable and are statistically independent (if generated with replacement).

The reversed strategy used in presentiment studies treats the random stimuli (targets) as fixed and uses the physiological measures (responses) as the outcome variable. This strategy results in concerns about dependence among responses. Given the immediate feedback on each trial, the analysis can have biases resulting from the physiological measures reflecting the properties of the particular random target sequence for a participant (Dalkvist, Mossbridge, & Westerlund, 2014; Kennedy, 2013).

Expectation bias as discussed by Dalkvist, Mossbridge, and Westerlund is a type of dependence that can occur in cases with immediate feedback. The traditional assumption for ESP research has been that human responses can also have other forms of dependence that the experimenters may not anticipate and that may or may not involve feedback.

One option for presentiment studies is to apply the traditional analysis strategy by using the physiological measures to predict the random stimuli. When properly done, this eliminates problems of dependence. A previous article recommended this option for confirmatory research and discussed the requirements for proper application (Kennedy, 2013).

Another option is to use the physiological measure as the outcome variable and to build statistical models that attempt to handle the potential dependence. However, the nature, amount, and effects of dependence are difficult to establish, and it is difficult to show that a statistical model adequately handles potentially complicated dependencies among responses (Kennedy, 2013). Dalkvist, Mossbridge, and Westerlund recommended that a simple model be used, but they provided little discussion about methods to evaluate whether a model adequately corrects the means and error variance for all pertinent dependencies.

From my perspective, this modeling approach cannot be expected to provide convincing evidence for a controversial effect (Kennedy, 2013). In addition to making this debatable recommendation, Dalkvist, Mossbridge, and Westerlund made two comments that are not correct.

Dalkvist, Mossbridge, and Westerlund inaccurately stated that bootstrap methods are "free from any statistical assumption" (p. 93) and can be applied in cases with dependencies among observations. Contrary to their comment, standard bootstrap methods are based on the assumption that the original observations are independent (Efron & Tibshirani, 1993, pp. 27, 31, 45, 396; Good, 2005, p. 23). Bradley Efron, the initial developer and promoter of bootstrap methods, commented: "There is no easy solution to problems of dependence … problems of dependence do not appear to be well understood and are an important area for further research" (Efron & Tibshirani, 1993, p. 396). As implied by this comment, simple models cannot be assumed to solve dependence problems.

Dalkvist, Mossbridge, and Westerlund also inaccurately said that my recommendation to use the physiological measure to predict the random stimuli is flawed by very low power because it does not adjust for the effects of the previous stimulus on the physiological measure. However, adjustments for the previous stimulus can be done with this analysis. The physiological values used in the final analysis for presentiment studies typically are derived by relatively complex processing after all data have been collected. The processing is usually described as baseline adjustment, normalization, data reduction, and/or artifact rejection.

The optimal strategy for confirmatory research is to have all decisions, derivations, adjustments, and criteria for the physiological data for a trial use only data collected prior to feedback for the trial (Kennedy, 2013). Any incorporation of data after feedback introduces potential for bias. When the physiological measures are used to predict the random stimuli, the trials can be stepped through in the sequence that they occurred, with the prediction criteria developed from previous studies or from previous trials in the current experiment. The prediction criteria can include adjustments for the stimulus on the previous trial. Models similar to those discussed by Dalkvist, Mossbridge, and Westerlund may be useful in developing the prediction criteria, but a confirmatory hypothesis test will be based on applying the criteria to other data that were not used in developing the criteria. This strategy avoids contamination by data after feedback for a trial—and also avoids other dependence problems.

The current situation with presentiment research reminds me of free-response research in the late 1970s. Many free-response studies had been done with methodology that was not optimal. The initial discussions of the methodological issues (e.g., Kennedy, 1979a, 1979b) resulted in arguments that some of the methodological concerns were unlikely to have significant effects and/or could be easily counteracted. However, after a few years of debates (e.g., Hyman, 1985), it became apparent to virtually everyone that the controversial methods should be avoided in future research (e.g., Hyman & Honorton, 1986).

References

Burdick, D. S., & Kelly, E. F. (1977). Statistical methods in parapsychological research. In B. B. Wolman (Ed.), *Handbook of parapsychology* (pp. 81–130). New York, NY: Van Nostrand Reinhold.

Dalkvist, J., Mossbridge, J., & Westerlund, J. (2014). How to remove the influence of expectation bias in presentiment and similar experiments: A recommended strategy. *Journal of Parapsychology*, *78*, 80–97.

Good, P. (2005). *Permutation, parametric and bootstrap tests of hypotheses* (3rd ed.). New York, NY: Springer.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York, NY: Chapman & Hall.

Hyman, R. (1985). The ganzfeld psi experiment: A critical appraisal. *Journal of Parapsychology*, *49*, 3–49.

Hyman, R., & Honorton, C. (1986). A joint communiqué: The psi ganzfeld controversy. *Journal of Parapsychology*, *50*, 351–364.

Kennedy, J. E. (1979a). Methodological problems in free-response ESP experiments. *Journal of the American Society for Psychical Research*, *73*, 1–15. Retrieved from http://jeksite.org/psi/jaspr79a.pdf

Kennedy, J. E. (1979b). More on methodological issues in free-response psi experiments. *Journal of the American Society for Psychical Research*, *73*, 395–401. Retrieved from http://jeksite.org/psi/jaspr79b.pdf

Kennedy, J. E. (2013). Methodology for confirmatory experiments on physiological measures of precognitive anticipation. *Journal of Parapsychology*, *77*, 237–248.

J. E. Kennedy

Broomfield, CO, USA
jek@jeksite.org

To the Editor:

We agree with Kennedy that great care should be taken to avoid use of uncertain methods in future presentiment and other precognition experiments. In particular, in our above-mentioned paper (Dalkvist, Mossbridge, & Westerlund, 2014), we have argued forcefully that expectation bias—a counterintuitive statistical bias caused by the erroneous belief, on the part of a participant, that the likelihood of a forthcoming sequential random event, such as a particular outcome in a sequence of coin flips, is dependent on previous outcomes—must be totally avoided or controlled for. This is particularly important because expectation bias causes systematic—in contrast to unsystematic—errors, thereby leading to invalid rather than unreliable results, the latter being in general less serious than the former, because they do not distort data in any fundamental way. We have also argued that ANOVA is a suitable method (although not the only one) for dealing with expectation bias.

However, as pointed out by Kennedy in the above Letter to the Editor and elsewhere (e.g., Kennedy, 2013), there are also other dependencies among responses that can compromise results from presentiment and other precognition experiments, for example, particular response patterns. However, in contrast to the effects of expectation bias, provided that stimuli are properly randomized, such effects are not systematic and can therefore not by themselves give rise to invalid results, but only affect the alpha level (in an unknown way). Nevertheless, the effects should be addressed. In our 2014 paper, two alternative solutions were discussed. One was to follow Kennedy's recommendation to reverse the statistical roles of the physiological variable and the previous stimulus, letting the former serve as the independent variable and the latter as the dependent variable. Specifically, we recommended doing this within the framework of ANOVA, with the previous stimulus and the physiological variable serving as independent variables and the forthcoming stimulus as the dependent variable—not separately, as recommended by Kennedy. Unfortunately, though, the need to categorize the physiological variable when it serves as an independent instead of dependent variable leads to an expected reduction in statistical power (potentially useful quantitative information is lost). For this reason, the optimal strategy may be to perform two tests—one with the physiological variable serving as the dependent variable (without categorization) and the other with the physiological variable taking the role of an independent variable (with categorization)—to investigate whether they converge, in which case the more powerful test could be accepted.

Our second suggested approach to the problem of dependent responses was to use a statistical computer simulation—a so-called Monte Carlo method instead of a traditional statistical, so-called parametric method—owing to the Monte Carlo method's freedom from underlying assumptions. Specifically, we recommended using "bootstrapping" because of its general availability. Regrettably, however, it had escaped us that this particular method, despite its freedom from the two other assumptions of traditional parametric statistical methods—equal variances and normal distributions—still requires satisfying the assumption of independent responses if the alpha level is to be reliable. We are grateful to Kennedy for noticing and reporting this oversight (which occurs easily considering the simplicity of the bootstrap method). It should be kept in mind, however, that other Monte Carlo methods exist that are free not only from the assumptions of equal variances and normality, but also from the assumption of independent responses. This holds, for example, for methods in which the stimuli, not the responses, are recursively randomly redistributed to create an "empirical" sampling distribution of a particular response statistic (e.g., the mean) associated with randomly chosen stimulus orders, even though the number of different stimuli limits the statistical power of such methods (for an example, see Dalkvist & Westerlund, 1998). Unfortunately, however, at present these stimulus sampling methods require specifically tailored programming to be done, and they are thus not equally readily available as bootstrapping methods.

Despite the problem of dependent responses in testing presentiment and other precognition results discussed above, we insist on recommending that ANOVA be used. We do this mainly because, by using ANOVA, the expected (huge) effect of the previous stimulus becomes isolated from the possible effect of the forthcoming stimulus in a direct and natural way. It is true, as pointed out by Kennedy, that the effect of the previous stimulus can be controlled for separately before the main test is performed—for example, by extracting residuals from regression analysis in a preparatory phase of the analysis, although such a method seems unnecessarily complicated. Exactly how effects of the previous stimulus (and possibly also some earlier stimulus) are controlled for is not a major concern, however.

Admittedly, from a purely methodological point of view, Kennedy's alternative strategy would seem to be a good one. There may be a psychological problem involved, however: How can we get researchers to accept and do the relatively laborious work the strategy requires?

References

Dalkvist, J. & Westerlund, J. (1998). Five experiments on telepathic communication of emotions. *Journal of Parapsychology, 62,* 219−253.

Dalkvist, J., Mossbridge, J., & Westerlund, J. (2014). How to remove the influence of expectation bias in presentiment and similar experiments: A recommended strategy. *Journal of Parapsychology, 78,* 80–97.

Kennedy, J. E. (2013). Methodology for confirmatory experiments on physiological measures of precognitive anticipation. *Journal of Parapsychology, 77,* 237–248.

Jan Dalkvist, Julia Mossbridge*, and Joakim Westerlund

Department of Psychology
Stockholm University, S-106 91
Stockholm, Sweden
jandalkvist@gmail.com

**Department of Psychology*
Northwestern University
2209 N. Sheridan Road
Evanston, IL 60208, USA

To the Editor:

Professor Caroline Watt's research, supported by the Perrott-Warrick fund, has recently focused on precognitive dreams. Some of her research has basically been anomalistic psychology (Valášek, et al., 2014; Watt et al., 2014), but in one thought-provoking study the psi hypothesis was also tested (i.e., Watt, 2014). The dream ESP studies have recently been reviewed (Sherwood & Roe, 2013), but since Watt has not cited any specific study as being inspirational I refrain from comparing it to prior studies. Some comments do however seem warranted.

Parapsychologists often expect participants in their studies to do things that they have never claimed to be able to do on command. Nevertheless, the researchers seem surprised when their results do not turn out to be significant. Watt argued that her sample was appropriate, since the majority of the participants "believed in precognitive dreaming, had experienced an evidential (according to Bender's criteria) precognitive dream personally at least once in their lifetime, and were able to recall their dreams at least once per week" (p. 121). Nevertheless, evidential precognitive dreams were quite rare experiences for most participants, and only two believed that they experienced this about once a week.

Precognitive dreams range from trivial to profound unforgettable experiences, but often concern deaths or accidents. The 50 participants in Watt's study were expected, during five nights at home, to dream about a randomly selected short video clip (around 1 min) that they would get to watch at some *unspecified time* in the future. Access to the clip was not given immediately. The task can thus be compared to trying to predict what one will briefly watch on television at some randomly determined time.

In addition, the participants were limited to 300 words to summarize their dreams (this letter is 632 words). The literature suggests that precognitive dreams are in some ways distinctive, but five nights can generate quite extensive dream reports. Unless the participants had learned to distinguish precognitive dreams they had little chance of excluding irrelevant dreams in their summaries. Signs of psi may have disappeared during the transfer to the summary, drowned by irrelevant dreams. Understandably, most participants doubted that their dream summaries would relate to the target clips.

Correspondence to the target clip was rated on a 1–100 scale by a judge and then by the participants themselves. In general, both the judges and the participants thought that there was very little resemblance. None of the two judges had any prior experience of evaluating this kind of material. Surprisingly, there were still 64 direct hits (32%) despite that "there was nothing qualitatively special about the hits compared to the misses" (p. 123). Similarity ratings had no significant relationship to hit rate. A possible explanation of the strange results is that Watt unintentionally tested the judges' ability to predict the correct targets. Some would however argue that their unexpected

success was likely more due to an experimenter psi effect rather than due to their own ability.

References

Sherwood, S. J., & Roe, C. A. (2013). An updated review of dream ESP studies conducted since the Maimonides dream ESP program. In S. Krippner, A, J. Rock, J. Beischel, H. L. Friedman, & C. L. Fracasso (Eds.). *Advances in parapsychological research 9* (pp. 38–81). Jefferson, NC: McFarland.

Valášek, M., Watt, C., Hutton, J., Neill, R., Nuttall, R., & Renwick, G. (2014). Testing the implicit processing hypothesis of precognitive dream experience. *Consciousness & Cognition, 28*, 113–125.

Watt, C. (2014). Precognitive dreaming: Investigating anomalous cognition and psychological factors. *Journal of Parapsychology, 78*, 115–125.

Watt, C., Ashley, N., Gillett, J., Halewood, M., & Hanson, R. (2014). Psychological factors in precognitive dream experiences: The role of paranormal belief, selective recall and propensity to find correspondences. *International Journal of Dream Research, 7*, 1–8.

Nemo C. Mörck

Torggatan 6A lgh 1203
282 31 Tyringe, Sweden
nemomorck@hotmail.com

To the Editor:

I am grateful to Nemo Mörck for his comments but would like to correct a couple of inaccurate statements. First, Mörck notes the relevance of Sherwood and Roe (2013), but states that "Watt has not cited any specific study as being inspirational." Sherwood and Roe (2013) had not been published at the time our study was conducted. However we do cite Sherwood and Roe's (2003) review and call for further research as motivating our study. Second, Mörck states that our participants were asked to dream about a video clip "they would get to watch at some *unspecified time* in the future." In fact, our participants were informed: "after we have received your dream summary you will be sent a 'target' video clip to view." Because the study aimed to have each participant conduct one trial per week, participants knew that they would quickly receive feedback on the target after they submitted their dreams, usually within hours. Finally, Mörck correctly notes that there is some ambiguity over the source of extra-chance scoring in Watt (2014). Due to the way that psi is defined, it is not possible to be absolutely certain about the source of any putative psi effect. I hope that, in time, improved theoretical understanding of psi will help to minimise such ambiguities.

References

Sherwood, S. J., & Roe, C. A. (2003). A review of dream ESP studies conducted since the Maimonides dream ESP programme. *Journal of Consciousness Studies, 10,* 85–109.

Sherwood, S. J., & Roe, C. A. (2013). An updated review of dream ESP studies conducted since the Maimonides dream ESP program. In S. Krippner, A. J. Rock, J. Beischel, H. L. Friedman, & C. L. Fracasso (Eds.). *Advances in parapsychological research 9* (pp. 38–81). Jefferson, NC: McFarland.

Watt, C. (2014). Precognitive dreaming: Investigating anomalous cognition and psychological factors. *Journal of Parapsychology, 78*, 115–125.

Caroline A. Watt

University of Edinburgh
7 George Square
Edinburgh EH8 9JZ, UK
Caroline.Watt@ed.ac.uk