

## Response to "Comments on Spottiswoode and May"

Aickin (2003) has published a lengthy commentary on Spottiswoode and May's (2003) paper on skin conductance pre-stimulus response. He appears to have five substantial criticisms to make that we describe below, with our responses.

1. Aickin's principal objection is to our primary analysis. In our protocol, 125 subjects were each exposed to 20 stimuli that were randomly chosen to be either a startling sound, or a silent control, with the probability of either occurring being  $\frac{1}{2}$ . In our analysis we measured whether or not a skin conductance response had occurred in the three-second-long epochs preceding each of these stimuli and compared the proportions of those epochs containing such a response for the pre-sound and pre-control cases.

Aickin states that "It is virtually a cliché in areas such as biostatistics that when one makes multiple observations on individuals, one should use an analysis that takes possible interdependence (within an individual) into account. The test used by Spottiswoode and May does not do this." With this statement, Aickin makes a mistake common among statisticians who are unfamiliar with experiments done in parapsychology (and similar experiments). The error is to misunderstand the source of randomness and independence in the experimental data. As everyone knows, responses in experiments with human subjects are not random and as Aickin indicates, responses within subjects are generally not independent. However, in the analyses commonly used in parapsychology, including ours, what matters is that the *target* or *stimulus* sequence contains independent and randomly chosen elements. Most medical experiments do not have this feature because multiple measurements are taken for each treatment application. Therefore, in the situations with which Aickin is familiar, you do have to worry about lack of independence.

In fact, the experiment we performed eliminated the possibility of interdependence within individuals by design, because the sequence of stimuli, either an audio startle stimulus or silent control, was completely randomized both within and across individuals. Thus, when comparing the subjects' skin conductance responses with the associated stimuli, any interdependence in each subject's responses could not give rise to any spurious correspondence with the sequence of stimuli, because the latter were completely random. No matter what was happening within the individual's skin conductance stream, the times we sampled were equally likely to be stimulus or control periods, and that choice was independent from one time to the next. Aickin's own analysis confirms the correctness of

our argument. He used a statistic that satisfied his requirement to account for the putative lack of independence and he arrived at the same result we did.

An analogy may make the situation clearer. Consider an ESP experiment in which 125 individuals are each asked to guess the result of 20 tosses of a fair coin, which they cannot see. Now it is well known that in such circumstances subjects exhibit various kinds of response bias and certainly do not produce random sequences of guesses. However, in analyzing such experiments it is perfectly legitimate to pool the data across subjects and report an overall statistic for the degree of correctness of the guesses, since the comparison is with a sequence (the results of the coin tosses), which is perfectly random by design. For instance, suppose someone always guessed heads. If the coin were truly random, the probability of a match for each guess would be  $\frac{1}{2}$ , and the outcomes would be independent from one toss to the next, even within the same guesser. It is the target sequence, not the response sequence, that assures randomness and independence. In our experiment the presence, or absence, of a skin conductance response prior to the stimuli in our experiment is analogous to a physiologically based "guess" of the upcoming, randomized stimulus.

2. Aickin describes the term "effect size" as arcane and complains that we do not define it. In fact, effect size measures have been widely used in parapsychological statistics since the 1980's (e.g. Utts, 1991), this move being driven by the increasing use of meta-analyses and an appreciation that small experiments studying a small effect need to compare effect sizes, rather than simply counting cases where  $p < 0.05$ . We agree that we should have specified which effect size measure we employed, namely the commonly used effect size  $Z/\sqrt{n}$ .
3. Aickin argues that "The second oddity is the use of a one sided p-value. . . ." We used a one-sided p-value because we had a one-sided hypothesis, namely that the observed proportions of pre-stimulus epochs containing skin conductance responses would be higher pre-audio stimulus than pre-control. This unidirectional hypothesis was based on our aim of partially replicating, and hopefully improving upon, the earlier cited work that also tested a one-sided hypothesis. We fail to see what is "odd" about this. It is common practice in statistical hypothesis testing. The important thing is that the one-sided hypothesis is specified in advance of data collection.
4. Aickin describes as "unsubstantiated" our pre-declared intention to eliminate a small number of trials with less than six stimuli of either type. As this intention was pre-defined, the justification for it was explained and since the exclusion criterion was entirely symmetric with respect to our hypothesis we felt that it was innocuous. As we point out in our paper the exclusion of the four trials involved made little difference to the results. In fact the observed pre-stimulus effect would have been slightly larger had we included them.

5. Finally, Aickin takes issue with "The Misbehaving Random Numbers." As we make clear in the paper, the random generator that produced these numbers passed the Die Hard suite of tests for randomness (Marsaglia, 1995). We also argue that the hypothesis that the behavior of the random generator was influenced by the experiment protocol is at odds with the extensive literature (Jahn et al., 1997: 350) on random number generator psychokinesis experiments. The bias we observed was some 750 times larger than that reported in a meta-analysis of this literature. Since we published this result, we now have data from a total of 340 subjects. The "bias" we reported for the first 125 subjects has statistically vanished ( $z = 0.7491$ ,  $p = 0.227$ ).

We are disappointed that the review process appears to us to have been less productive than usual in this case. Several of the above points were extensively discussed during the review process without resolution. Aickin states that "I would strongly encourage authors submitting to *JSE* either to employ the aid of a professional statistician, or at least have their work reviewed by one, before submission." While this is doubtless a good suggestion when unusual or complex statistical methods are employed, in our case the statistical analysis seemed so unexceptional that we did not feel it was necessary.

S. JAMES P. SPOTTISWOODE  
*sjp@nrg.com*  
 EDWIN C. MAY  
 JESSICA UTTS

### References

- Aickin, M. (2003). Comments on Spottiswoode and May. *Journal of Scientific Exploration*, *17*, 643-650.
- Jahn, R. G., Dunne, B. J., Nelson, R. D., Dobyns, Y. H., & Bradish, G. J. (1997). Correlations of random binary sequences with pre-stated operators' intention: A review of a 12-year program. *Journal of Scientific Exploration*, *11*, 345-368.
- Marsaglia, G. (1995). The Marsaglia Random Number CD-ROM, with The Diehard Battery of Tests of Randomness. Produced at Florida State University under a grant from The National Science Foundation, [www.stat.fsu.edu/pub/diehard](http://www.stat.fsu.edu/pub/diehard).
- Spottiswoode, S. J. P., & May, E. C. (2003). Skin conductance prestimulus response: Analyses, artifacts and a pilot study. *Journal of Scientific Exploration*, *17*, 617-641.
- Utts, J. (1991). Replication and meta-analysis in parapsychology. *Statistical Science*, *6*, 363-403.