Correspondence

Cardeña, Etzel

Published in:
Journal of Parapsychology

2016

Citation for published version (APA):
Cardeña, E. (2016). Correspondence. *Journal of Parapsychology*, *80*(1), 120-122.

Total number of authors:
1

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science, 349*, 1095–9203. Retrieved from: http://www.sciencemag.org/content/349/6251/aac4716

Parker, A. (2000). A review of the ganzfeld work at Gothenburg. *Journal of the Society for Psychical Research, 64*, 1–15.

Parker, A. (2013). Betraying the future by distorting the present: A reply to Bauer et al. *Journal of the Society for Psychical Research, 77*, 125–127.

Parker, A. (2015a). The jungle of hypnotic psi: Part 1. Research on hypnosis relevant to psi. *Journal of Parapsychology, 79*, 20–36.

Parker, A. (2015b). The jungle of hypnotic psi: Part 2. Research on relationships between psi and hypnosis. *Journal of Parapsychology, 79*, 37–52.

Parker, A., Grams, D., & Pettersson, C. (1998). Some further variables relating to psi in the ganzfeld. *Journal of Parapsychology, 62*, 27–45.

Wright, D. B., & Loftus, E. F. (1999). Measuring dissociation: Comparison of alternative forms of the Dissociative Experiences Scale. *American Journal of Psychology, 112*, 497–519.

Adrian Parker

Department of Psychology
Gothenburg University
Box 500
40550 Gothenburg, Sweden
adrian.parker@psy.gu.se

To the Editor:

Dr. Parker continues to attack our 5-year-old publication (Marcusson-Clavertz & Cardeña, 2011) and also my reply (Cardeña, 2016) to his criticisms. Regarding his remark that his papers were "intended to be constructive critiques," anyone who is aware of his more than decade-long public and private history of attacks against me will be able to evaluate his likely intention.

Now to more substantial issues:

1. Dr. Parker maintains that it is not "so meaningful" to spend time pointing out the various typos, misinterpretations, and inaccuracies in his original papers, but besides correcting the record they provide a potential indicator of how careful and accurate may be the rest of the content.
2. His statement that "some of the main hypotheses were evaluated using five to nine participants," is highly misleading. We had five main hypotheses (Marcusson-Clavertz & Cardeña, 2012). Hypothesis 1, belief in self-success in the psi task will be positively related to psi scoring, was confirmed with all 26 participants. The same was true of Hypothesis 2, which postulated that previous psi experiences (not the same as the "sheep-goat" effect, contrary to what Dr. Parker states) would also correlate with psi scoring. For Hypothesis 3, we conducted a t test between high and low hypnotizables (a correlational analysis would be inappropriate because the medium hypnotizables were excluded), comparing cells of 14 versus 12. For Hypothesis 4, contrary to Dr. Parker's claim that it "was confirmed," we actually wrote on page 246 that "Neither was there a significant interaction between hypnotizability and dissociation." This is the *only* hypothesis-testing analysis for which we had cells of between five and nine. Incidentally, Table 6 of one of his papers shows that Dr. Parker (Parker, Grams, & Pettersson, 1998, p. 330) analyzed cells of $n = 8$ and $n = 10$, besides carrying out many analyses on multiple variables. Finally, for Hypothesis 5, which proposed that greater alterations in consciousness would relate to high psi performance and hypnotizability, the significant result with a very strong size effect was with the group of 14 high hypnotizables, not a cell of "five to nine."
3. As for not predicting that participants overall would significantly psi hit, we wrote on p. 244, "Evaluating overall psi was not a target of this process-oriented research." Choosing participants

that we thought would be psi-hitters along with those we thought would be chance-scorers or even psi-missers made it oxymoronic to predict overall success. As Bem, Palmer, and Broughton (2001, p. 215) wrote "genuine progress in understanding psi rests on investigators' being willing to risk replication failures by modifying the procedure in any way that seems best suited for exploring new domains or answering new questions."

Regarding other misleading statements by Dr. Parker:

1. He mentions that we did not specify "which were the post-hoc findings," but on page 247, we stated that analyses with other PCI dimensions than the *altered state* one "can be considered exploratory."
2. As for preregistering our study, at that time it was not common practice and the PKU registry was not even started until the fall of 2012, so if Dr. Parker wants to blame us for that lack of precognition, perhaps he could start listing his own previous studies with preregistered analyses. And, as I mentioned in my previous reply, the samples for each one of his studies (Parker et al., 1998) was 30. Whether he combined his data later for some analyses is irrelevant to my statement.
3. Contrary to what Dr. Parker seems to hint, an indication that we were aware of the strengths and limitations of the DES and other dissociation instruments is that the American Psychiatric Association tasked me some years ago to write a review chapter of dissociation measures (Cardeña, 2008).
4. Dr. Parker questions that the *Journal of Experimental Psychology: Learning, Memory, and Cognition* is a rigorous journal because a recent Open Science Collaboration paper found problems with replicability of some studies (incidentally, these findings have also been disputed, see Gilbert, King, Pettigrew, & Wilson, 2016), despite its 78% rejection rate (http://www.apa.org/pubs/journals/features/2013-statistics.pdf) and 2.86 impact factor. Sadly Dr. Parker offers no references to journals in which he has published with better statistics than these.
5. Finally, Dr. Parker might cite and agree as much as he wants with the antipsi people here in Sweden, but his statement that their criticism "might easily have led to the loss of the Lund chair" goes against the public statements of support I received in the media by Lund University's then-President Per Eriksson (Stiernstedt, 2012), the current Dean of my College Ann-Katrin Bäcklund (Anonymous, 2012), and the then-Chair of the Department, Per Johnsson (Fagerström, 2012; see also Cardeña, 2015).

Is there anything I can agree with Dr. Parker on? Well, of course our study had a number of limitations, which we listed on page 252 and which Dr. Parker repeats, including a small *N* and using the measure of alteration of consciousness with the same people under ganzfeld but not from the psi session. As with most research in any field, it should not be considered definitive but as evidence supporting some hypotheses and not others, and offering some promising new leads. So I hope that he will agree with me that it will be more constructive for everyone concerned to spend time conducting their own preregistered new research with large *N*s rather than pursuing this debate.

References

Anonymous (2012, October 31). Experiment om tankeöverföring [Experiment on telepathy]. *Studio Ett.* Retrieved from http://sverigesradio.se/sida/artikel.aspx?programid=1637&artikel=5330277

Bem, D. J., Palmer, J., & Broughton, R. S. (2001). Updating the ganzfeld database: A victim of its own success? *Journal of Parapsychology, 65,* 207–218.

Cardeña, E. (2008). Dissociative disorders measures. In A. J. Rush, M. B. First, & D. Becker (Eds.), *Handbook of psychiatric measures* (2nd ed.) (pp. 587–599). Washington, DC: American Psychiatric Press.

Cardeña, E. (2015). The unbearable fear of psi: On scientific censorship in the 21st century. *Journal of Scientific Exploration, 29,* 601–620.

Cardeña, E. (2016). [Letter to the Editor]. *Journal of Parapsychology, 79,* 242–245.

Fagerström, E. (2012, April 19). Ett decenium i vetenskapens gransland. [A decade on the borders of science]. *Sydsvenskan.* Retrieved from http://www.sydsvenskan.se/2015-04-19/ett-decennium-i-vetenskapens-gransland

Gilbert, D. T., King, G., Pettigrew, S., & Wilson, T. D. (2016). Comment on "Estimating the reproducibility of psychological science." *Science, 351,* 1037–1038. doi:10.1126/science.aad7243
Marcusson-Clavertz, D., & Cardeña, E. (2011). Hypnotizability, alterations in consciousness, and other variables as predictors of performance in a ganzfeld psi task. *Journal of Parapsychology, 75,* 235–260.
Parker, A., Grams, D., & Pettersson, C. (1998). Further variables relating to psi in the ganzfeld. *Journal of Parapsychology, 62,* 319–337.
Stiernstedt, J. (2012, July 14). "Vi studerar tomtar och troll också." ["We also study brownies and dwarves"]. *Svenska Dagbladet.* Retrieved from http://www.svd.se/vi-studerar-tomtar-och-troll-ocksa

Etzel Cardeña

Department of Psychology
Lund University
Paradisgatten 5P
Lund 223 50, Sweden
etzel.cardena@psy.lu.se

To the Editor:

Few recent parapsychological experiments have given rise to as much acrimony as a study in this journal by Dr. David Marcusson-Clavertz and Prof. Etzel Cardeña (2011; hereafter DM/EC). They reported that, for high hypnotizables *only*, there was a significant correlation between psi *z* scores and being in an altered state of consciousness. Later in an interview, EC was quoted as having said that this correlation is ". . . a very strong indication that telepathy has really occurred" (Oredsson, 2012, p. 17). Vociferous exception was registered by a number of Swedish academics (Halle et al., 2012). They emphasized that the overall hit rate was in fact nonsignificant (27% direct hits, chance = 25%) and considered the correlation an artifact of multiple analyses: The study was profiled as pseudoscience. For most English-speaking readers this controversy lies largely behind the (Swedish) language barrier (but see Cardeña, 2013a, 2013b).

More recently, in the course of a major review of psi and hypnosis research in this journal Prof. Adrian Parker (2015) characterized the DM/EC study as "overloaded with variables and hypotheses" (p. 41). Cardeña (2015) has disputed aspects of Parker's criticism but did not really settle the underlying issue. Is the DM/EC inference from the data really justified?

The recent statement by the American Statistical Association (Wasserstein & Lazar, in press) highlights the gulf between "statisticians' statistics" and "researcher's statistics" and brings the vexing problem of multiple analyses into the limelight. Every additional analysis is basically an extra shot at the significance jackpot and the strict logic of statistical significance testing requires that a "family-wise" (or similar) analysis is performed per study; only then is the probability value calculated correct for the *set* of analyses, rather than for the individual analysis. This does not mean that only a single variable should be investigated per study; rather the *set* of *p* values must be corrected appropriately (e.g., Bretz, Hothorn, & Westfall, 2011).

Most researchers do not formally correct probabilities for multiple analyses. If a *large number* of analyses are performed this has unfortunate consequences: (Virtually) every such study contains sufficient (spurious) "significances" to be published as "evidence of some effect." It is unlikely that such a study will come up totally empty with "no evidence for an effect." Karl Popper (1959 /1992) proposed that the essential characteristic of scientific theories is that they are *falsifiable* (demarcation criterion). Excessive uncorrected analyses "immunize" against the possibility of falsification.

With 100 analyses per study the (binomial) probability of some "significances" is as high as 99.4% and the corresponding empty hands (no "sigs") is a negligible .6%. For 14 independent analyses per experiment just more than half of the studies are expected to be "publishable" merely by capitalizing on chance. Stacking the odds in the researcher's favour by multiplying analyses is decidedly not "playing the game."

In lieu of formal correction for multiplicity, researchers early devised a rough and ready workaround. Two broad classes are distinguished—confirmatory and exploratory studies. For confirmatory studies the number of analyses is typically restricted to a few and significance testing can be meaningful (if tak-