Journal of Scientific Exploration, Vol. 25, No. 4, pp. 690–694, 2011 0892-3310/11

REPLY

Reply to May and Spottiswoode's "The Global Consciousness Project: Identifying the Source of Psi"

Peter Bancel

Institut Métapsychique International, Paris, France

I am pleased that May and Spottiswoode have initiated a discussion about the Global Consciousness Project (GCP), and I would like to thank the *JSE* editors for this opportunity to respond to their paper. May and Spottiswoode (M&S) suggest that the source of the statistical deviations reported by the GCP can be attributed to an experimenter effect and that Decision Augmentation Theory (DAT) can adequately model the GCP results. While I disagree with the analysis, their contribution is particularly welcome since they address an essential question that needs to be resolved by any model: Does the GCP measure a real, physical effect?

It is easy to see why the question is pertinent if we recall the experimental methodology. The GCP hypothesizes that data from a network of random number generators (RNGs) will deviate during events of global significance. Testing the hypothesis is a two-step procedure: 1) From time to time, data from the continuously accumulating RNG database are selected according to a blind procedure in which an event is identified from news or other sources and a data segment corresponding to the event is specified. A pre-designated test statistic is then calculated for the selected data. 2) The test statistic is converted to a standard normal *Z*-score and added to a table of *Z*-scores for all events. The formal experimental result is the mean of these *Z*-scores. As of late 2011, the GCP obtains a mean *Z*-score that exceeds zero by 6 standard deviations. This is the hugely significant result that M&S seek to explain.

The explanation proposed by the GCP is that, given the blind selection procedure, the change in the network statistics during events is due to a change in the physical behavior of the RNG devices themselves. This proposal can be tested by developing suitable models of the data. The GCP maintains that models which posit a physical mechanism that acts on the RNGs represent, at the least, a plausible avenue of investigation. What M&S correctly point out is that one cannot exclude, a priori, that a psi-mediated intuition, which informs the experimenter's designation of events, might compromise the blind selection procedure. In such a case, one cannot rely on the formal result to make inferences about the RNG behavior. According to M&S, the freedom of choice in selecting the events and their start/end times, in conjunction with psi-mediated information about the resulting test statistics, sets up a satisfactory explanation of the experiment. For M&S, the experimental result is merely the consequence of fortuitous selection of naturally occurring data deviations.

M&S go a step further by using a DAT model to test their idea against the GCP data. DAT derives from the well-known principle whereby the ratio of signal-to-noise in a sample (i.e. the *Z*-score of a measurement) increases as $DF^{1/2}$ where DF ("degrees of freedom") is the sample size. M&S distinguish between degrees of freedom which are relevant for DAT—those which designate an elemental instance of decision concerning what data to include in a measurement—and irrelevant "internal" degrees of freedom which have no inherent relevance for the decision process. In the GCP event experiment, the elemental DAT degree of freedom is the selection of a data block representing an event. In the DAT picture, a constant effect size, Z_{DAT}, is attributed to each instance of event selection. Z_{DAT} is independent of all internal degrees of freedom, such as the number of seconds or RNGs in the data block. It is thus evident that any physical model which *does* depend on the internal degrees of freedom can be distinguished from DAT models by testing for an association between $DF_{Internal}$ and *Z*. If the DAT model holds, no association will be found, whereas a physical model will yield a positive association between $DF_{Internal}$ and measured values of *Z*. A standard way to test for association is by ordinary least squares regression (OLS). M&S chose to do an OLS for Z^2 versus N, the number of RNGs in the network during the event. Their OLS yields a regression slope within a standard error of zero, and they conclude that this supports a DAT interpretation of the GCP.

This conclusion might have some weight if their regression analysis were done correctly. Unfortunately, M&S make several errors which are fatal to their argument. Here, I briefly sketch their errors and show that a proper test leads to the opposite conclusion from M&S: There is a clear association between Z^2 and $DF_{Internal}$ and reasonably strong grounds for rejecting DAT in favor of a physical model.

I discuss four separate errors, in order of increasing consequence. The first two have negligible impact, but the others invalidate M&S's calculations and reverse the conclusions one must draw from the DAT analysis.

1. Incorrect values of explanatory and response variables

M&S use values for the explanatory variables, N, that are listed on the GCP website. The website values are only approximate and should be replaced with exact values which account for null data trials. In addition, it would be preferable to calculate exact Z-scores for events 310–313, rather than use the estimates M&S list in their table.

2. Incorrect determination of fit parameter standard errors

M&S take the Z^2 as response variables. This choice yields non-normal fit residuals which impact the reliability of the usual OLS estimators for the fit parameter standard errors. Reliable errors for the regression slope and intercept parameters need to be determined by simulation. A 20,000-iteration Monte Carlo calculation (in which I use the correct values of N and Z) yields standard errors of the slope and intercept of 0.0061 and 0.303, respectively. These are substantially larger than the M&S values of 0.0034 and 0.058.

3. Failure to control for influence points

M&S neglect to perform regression diagnostics. It is well-known that OLS regression is sensitive to outlier and leverage points which may unduly influence the estimation of fit parameters. A common diagnostic is the Cook distance, d, which measures a point's relative influence on parameter estimation. Typically, a cutoff value sets an acceptable level of influence. In the representation I use here, d has a cutoff of 1 and points with $d > \approx 3$ may be considered substantially influential. Data points exceeding the cutoff need to be assessed carefully for experimental errors or other irregularities which might invalidate their inclusion in the regression dataset. For the GCP regression data, five data points have d-values greater than 3 and Event 1 has an exceedingly high value of $d = 42.3$.

A recent paper published in *JSE* (Bancel & Nelson, 2008) assessed the GCP Event Experiment in detail (the paper is cited by M&S in their article). The paper clearly states that, due to network instabilities during the first months of operation, Z-scores for the first 10 events are not reliable and should be excluded from analyses (footnote 21 in the paper). With these Z-scores excluded, a re-calculation of the OLS regression yields a positive slope parameter, increasing from −0.00064 to 0.0063. Correspondingly, the one-tailed Monte Carlo P-value for a test of the DAT hypothesis decreases from 0.51 to 0.13, indicating a much weaker agreement with the DAT model than M&S claim. More importantly, the re-calculation shows that OLS is an ill-suited choice for testing association between Z^2 and N. M&S would do better to use a modern technique of robust regression estimation. Robust methods are far less sensitive to outliers, influential data points, non-normality, and heteroskedasticity, and they frequently provide a power advantage over OLS. A slope estimate using one such robust technique, the Theil-Sen estimator (TSe), is discussed below.

4. Incorrect assignment of the regressor variable

The most serious error M&S make is in their choice of regressor. In Bancel and Nelson (2008), we show in considerable detail that the measured effect can be traced to correlations between pairs of RNGs. If the RNG output is written as *z(i,t)* where *i* labels the RNGs and *t* is the time in seconds, then the average correlation, ξ, is simply

$$\xi = \mathrm{DF}^{-1} \sum z(i,t)\, z(j,t)\ .$$

The sum is over all unique RNG pairs for each second so that $\mathrm{DF}_{GCP} = T(N^2 - N)/2$, where T is the number of seconds during the event. The correlations ξ distribute normally (to high approximation, under the central limit theorem), and the event *Z*-scores are given as $Z = \xi \sqrt{\mathrm{DF}_{GCP}}$. The appropriate regressor is thus DF_{GCP} and not N as M&S propose. Inappropriately selecting a DF of N introduces a large dispersion in the regressor variable, DF_{GCP}. This leads to a partial randomization of the regressor (see Figure 1) and all but guarantees that the regression test will accept the DAT hypothesis.

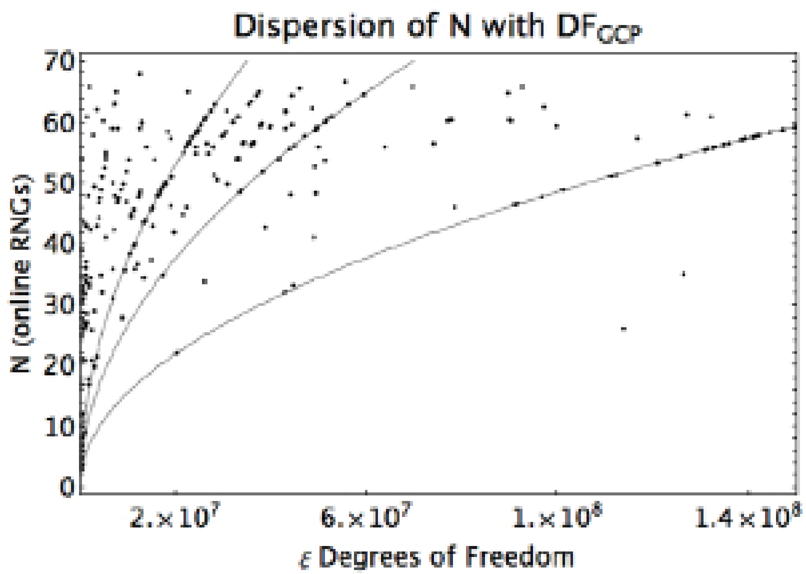

Figure 1. Dispersion of N with DF_{GCP}.
The plot shows the correspondence between DF = N and the correct DF_{GCP} as identified by Bancel & Nelson for the 299 events cited by M&S. At fixed N, there is a broad dispersion in the values of DF_{GCP}. The dispersion greatly reduces the power of Z vs. N regressions. The gray curves are lines of constant event duration (4, 8, and 24 hours, left to right).

To conclude, I show that the DAT model is rejected when a robust estimator and the appropriate regressor are used. I employ the Theil-Sen estimator which has a considerably higher power than OLS for the Z^2 vs. DF regression. The TSe slope estimate is taken as the median slope of all pairs of data points. Confidence intervals can be determined by bootstrap analysis, but a hypothesis test of the DAT model requires empirical determination of the TSe distribution by Monte Carlo simulation. A one-tailed Monte Carlo test of DAT yields a P-value of 0.024. Using the recommended dataset which excludes early events, the *P*-value falls to 0.0053. These correspond to *Z*-scores of 1.98 and 2.56, and indicate that the GCP data reject the DAT model with moderately high confidence. Although it is beyond the scope of this Reply, one can show that a similar procedure which tests the alternate hypothesis of a physical effect accepts that hypothesis as being consistent with the data (in preparation by Bancel).

In summary, M&S highlight a fundamental interpretational issue of the GCP: whether the measured effect has its source in a physical perturbation of the network RNGs. The issue can be addressed by testing for such structure in the event data as would be predicted by a physical effect. The association of Z and DF_{GCP} is one example of this approach, and the analysis presented here supports the GCP proposal. However, the issue is sufficiently important that further, independent tests are needed before a convincing conclusion can be drawn (Nelson & Bancel, 2011). A number of independent tests have been identified, and a report is currently in preparation by Bancel.

Acknowledgements

My thanks to Roger Nelson for many useful discussions. This work is supported in part by a grant from the Parapsychological Association. I would like to dedicate this Reply to the memory of Helmut Schmidt and his pioneering contributions to psi research.

References

Bancel, P., & Nelson, R. (2008). The GCP Event Experiment: Design, analytical methods, results. *Journal of Scientific Exploration, 22*(3), 309–333.

Nelson, R., & Bancel, P. (2011). Effects of mass consciousness: Changes in random data during global events. *Explore: The Journal of Science & Healing, 7*(6), 373–383.