Journal of Scientific Exploration, Vol. 25, No. 4, pp. 683–689, 2011 0892-3310/11

REPLY

Reply to May and Spottiswoode on Experimenter Effect as the Explanation for GCP Results

Roger Nelson

Global Consciousness Project

This paper is dedicated to the memory of Helmut H. W. Schmidt

I appreciate the opportunity to respond to the article by May and Spottiswoode (hereafter M&S) in which they attempt to identify the source of the anomalous correlations reported by the Global Consciousness Project (GCP). Their aim is to show that the GCP data, like laboratory micro-PK data, can be explained in terms of Decision Augmentation Theory (DAT), and in particular as an experimenter effect. The experimenter they have in mind is Roger Nelson, and while I suppose it is some sort of honor to be perceived as a powerful psi source, I consider it unlikely that the highly significant composite findings in the GCP experiment are attributable to me. In this paper, I will discuss why, and in the process show logical and factual errors that undermine and largely if not completely destroy the case for DAT and the experimenter effect. Before proceeding, however, I want to say that I appreciate the civility of expression and argument M&S bring to bear. I hope the discussion in their paper and mine will be helpful to readers who are interested in the GCP experiment, in the important questions of interpretation it raises, and in the substantial implications it may have for psi research.

Let's begin with some simple mistakes. In their Abstract the authors imply that we propose an asymmetric force or force per bit to explain the deviant GCP statistic. We do not posit or speak of forces at all, so in terms of what the GCP does, this can be seen as a straw man. Of course this is the language M&S are accustomed to using, so we will accept that and deal with the actual issues ab initio. To begin, our primary measure as well as the independent measures we have developed all are correlations or correlation-based. We simply do not make claims about forces. That said, we have found better fits to the empirical data with field-like models than the classic selection models (DAT) that M&S believe should apply (Nelson & Bancel, 2009).

M&S say the "basic idea" of the GCP sprang from Helmut Schmidt's research with RNGs whose behavior was the target of participant guesses or influence, though they don't explain how this leads to the GCP. Schmidt's test trials were typically decisions based on 1 bit (one binary decision) and typically the trials took a few seconds. Crucially, there was a participant with an intention to influence that bit. There is no such participant in the GCP, and to construe the experimenter in that role demands a convoluted argument. Nevertheless, we encourage data-based modeling to test such notions empirically.

We also do not claim, as they assert, that global consciousness is the source of the anomalous effects, rather, we use an operational definition of the object of study. The hypothesis we test simply says that we expect deviations in the data from our global network of RNGs during major world events. Formally:

> **Periods of collective attention or emotion in widely distributed populations will correlate with data deviations in a global network of physical random number generators.**

This general hypothesis is tested via a replication series of completely specified simple hypotheses of the form: *The GCP network variance statistic (or other specified measure) will be greater than expectation from time 1 to time 2 on a given date*. That is, we conduct a series of replications in which the exact data segment is identified along with the statistical test that will be applied. The composite across these replications constitutes a formal test of the general hypothesis (Bancel & Nelson, 2008, Nelson & Bancel, 2011). Over time, we expect to be able to discriminate between models, including one that is physical and might be given a name like "global consciousness" because of the link to collective human attention postulated in the general hypothesis. In the first instance, however, we seek evidence for a correlation, not for a theoretical entity.

It is a surprise to see M&S describe the inception of the GCP as "launched in 1998 in part in anticipation of the then up-coming Y2K." That, I am afraid, is made up. The project as conceived in 1997 was an evolution of FieldREG studies (Nelson, Bradish, Dobyns, Dunne, & Jahn, 1996, Nelson, Jahn, Dunne, Dobyns, & Bradish, 1998), and was concretely modeled on two prototype efforts to expand that concept to a larger scale by combining data from a dozen or more RNGs in Europe and the U.S. (Nelson, 1997, Nelson, Boesch, Boller, Dobyns, Houtkooper, et al., 1998). New Year's was an obvious candidate for a "global event" from the beginning, and the Y2K moment was of course included, but it was not a motivator for the project.

M&S say they have trouble understanding the GCP hypothesis, and knowing what the process is for selecting the events. The latter is a reasonable question, which we have addressed in some detail (Nelson & Bancel, 2011), but valid statistics are not dependent on these issues. It is sufficient that each event is selected and the hypothesis registered prior to examining the data, and that the results are all reported. Whether we are looking at the time of the event or the time when people become aware of the event isn't relevant to the validity of the simple hypothesis tests in the replication series. Fortunately, that question can be empirically explored in the large GCP database, because the event definitions typically cover a time span that includes both aspects. Similarly, many other questions can in principle be asked of the data because we employ a two-stage hypothesis, with a general statement that is flexible to allow explorations, and formal testing via a series of rigorously defined simple hypotheses. In practice, the event definitions are standardized based on experience, with events in each of several categories specified using the same relative starting point and duration.

M&S make a point of disputing an "Orwellian rewrite of history" with regard to the use of PK or a force per bit model, but they are arguing with someone else. The GCP does not use this language; instead, we speak in terms of correlations. Indeed, the primary measure (the only one M&S address) is equivalent to an average pair-wise correlation in the RNG data. As the general hypothesis states, we are asking whether there is a deviation of this inter-node correlation that occurs during (is correlated with) the formally selected events. M&S wonder if the prediction is "constrained to be in real time with the events" and of course it is. To imagine otherwise is to confuse the defined event with the putative effect—an important distinction which, again, can be assessed in the GCP database. For more detail on the point, see Bancel and Nelson (2008) and Nelson and Bancel (2009).

An issue that is never discussed by M&S is the fact that all their calculations and theorizing about the sources of effects derive from a model that was designed to address intention experiments, that is experiments where someone is attempting to change the behavior of an RNG. However, while the GCP experiment uses RNG technology, it is not about intentions to affect the behavior of the devices. Its design is better regarded as an environmental monitor, where the environment of interest is variations in the coherence of consciousness and emotion across large populations. M&S might argue that the experimenter has an intention, but it would be of a categorically different sort. As the experimenter of interest, I would characterize my intention in the GCP as a desire to learn something—I'm not much interested in getting more ones or zeros. (Good thing, too. As a participant in the PEAR REG experiments, I produced a very small, non-significant positive effect over several years and thousands of trials.)

Moreover, the primary metric for evaluating the GCP hypothesis is not an increase or decrease of ones and zeros, but excess correlation among the RNGs. We predict and test for an increase in the composite spatially distributed network variance, which is equivalent to predicting an increase (from zero) in the average correlation between hundreds or thousands of RNG pairs. The

M&S model requires that the experimenter intuit or precognize the outcome of these tests. Of course the experimenter might guess well when looking for events that will be correlated with changes in the GCP network, but that seems an obvious, mundane talent, assuming there are correlated changes. It can most likely be learned, as well, with no experimenter psi required. This is a testable proposition, and is the subject of a program we are developing to define and teach consistent criteria for event selection which can be applied by independent observer/analysts, including skeptics.

In the "Formal DAT Analysis" section, the authors define their procedure: "To determine whether there is a force/bit effect in these data, we created a scatter plot of the stated *Z*-score squared against the number of RNGs that were used to compute the *Z*-score." This states that the N of RNGs is used, not the number of bits, as is usual in the DAT literature. It would have been useful for M&S to explain the switch and show its equivalence. The quoted statement also says they use GCP's stated *Z*-scores for individual events. Thus, their proposition appears to be that the *Z*-score, which represents a spatially distributed variance measure (or increased pairwise correlation), is dependent on the number of RNGs in the force per bit model, but independent of N for the DAT model. They do not further discuss the models or their assumptions, but let's accept that for the moment. Taking the alternative formulation for the GCP effect since it is easier to visualize, we can ask whether the significance of the correlation should depend on the number of RNGs. Since increasing the number of pairs should grow the number of correlations, leading to smaller error bars on the average correlation, the significance represented by the calculated *Z*-scores would be expected to increase. Thus, the discovery of a null relationship of *Z*-scores (representing the correlations) to the N of RNGs would be surprising, and inconsistent with a physical model. We believe the data do not support the M&S claim that the regression has zero slope.

Putting it explicitly, the bottom line drawn by M&S is premature at best: "We are left then to conclude that Dr. Nelson's DAT-like decision capacity drives the GCP result, and it is unlikely that their statistically robust result is due to a variation of their primary hypothesis of some global consciousness connections to the RNG devices." In a personal communication responding to an earlier version of this paper, Peter Bancel stated that "simulations show that it is not possible to distinguish between the models—there's not enough statistical power in the data." He goes on to say that while the data may be consistent with DAT, they are also consistent with a reasonable "force" model. York Dobyns has tested the DAT model against data from RNG experiments and finds it inadequate. He too points out the problem of small effects: "The selection model assumes that the operator somehow becomes aware of the actual run outcomes and assigns intentions to suit, but I also present an argument showing that given the small overall effect size, a standard DAT model would produce the same statistics in the output data as the intention-selecting model . . . " (Dobyns, 1993, 1996, Dobyns & Nelson, 1997).

A little later in their paper, M&S recognize that their analysis doesn't really discriminate alternative models for the GCP data very well, but then say that after all since the GCP does use RNGs the analysis of laboratory intention experiments should still apply: "Why would the GCP data be any different? Thus we call into question the GCP's underlying assumption of variance interaction." This is a very weak argument. Perhaps M&S are confused by differing uses of the term "variance" and perhaps their comments are directed, inappropriately, to the variance of the individual RNGs. In any case they miss the point that, far from being an assumption, we *define* the network variance as our primary measure.

Moving to a different perspective, M&S attempt to compare the success of Nelson vs. other predictors. They state that Nelson "brought 234 events to the attention of the GCP," but their count is based on the assumption that whenever Nelson is included in the "source" column of the formal results table, he is the source. In fact, whenever names other than Nelson are included, they can legitimately be considered the source(s). When others suggest an event, there is frequently a need for collaboration to establish the analysis parameters. For example, an event will be suggested, but not the start and end times required for a formal event specification. Because of their faulty assumption, the counts made by M&S are wrong. In a recent categorization, May 28, 2011, I found that a little more than half of the events had been suggested by one or more others, sometimes including me (N = 188) and that Nelson alone had been the source for the rest of the predictions (N = 177). Looking at the two subsets separately, we see that Nelson's composite Z is 5.188, agreeing pretty well with that calculated by M&S on the smaller database they used, but the composite Z for the other predictors is 3.706, not even close to the M&S calculation. The difference in composite effect size attributable to Nelson vs. others is substantial, but not significant; the difference Z-score is 1.143. What is more important is an obvious logical flaw in the reasoning behind this comparison of outcomes for Nelson vs. others. Since Nelson is involved in registering, analysis, and writeup for every event, and would presumably always have a similar interest, it is clear ex hypothesi that this attempted comparison is artificial and invalid. It cannot tell us whether the GCP effect is due to experimenter psi.

Beyond that, the question whether Nelson is the primary source of the effects is far more complex than M&S apparently recognize. Even if the Nelson vs. other comparison were legitimate and the difference in composite Z were significant, such a comparison selectively ignores other factors. In particular, because the predictions for the GCP formal series are made a priori, they are guesses. They are explicit attempts to specify, without any prior knowledge, a period of time when the data will be found to deviate from expectation in a certain statistic. There is a history of predictions and outcomes, that is a feedback loop that can be expected to educate the predictor as to what factors or features of events are associated with confirmations of the predictions. Does it not seem reasonable that Nelson, who pays more attention than anyone else to the sequence of successful and failed predictions, might learn something along the way about which are the "good bets" to make? That's pretty mundane compared with DAT or the psychic experimenter effect postulated by M&S, but it seems very likely to account for some considerable part of the (non-significant) advantage Nelson has over other predictors. When the details are considered, there are still other reasons why non-Nelson predictions may fail. They are often about local and relatively small items, and many of the ones accepted for registration and analysis are about meditations, peace prayers, earth days, and the like. I'm attuned to the ideas and ideals, and in order to learn about these events I accept many such suggestions, but our categorization studies have shown that they tend to have small effects.

In their discussion, M&S argue that the difference in success rate for Nelson compared with other predictors cannot be attributed to practice and experience. "To realize that, say earthquakes would be an effective event while sporting events would not, would require an independently supported model which predicted, and hopefully explained, why these classes of event would show differing GCP effects. No such model has been offered." In fact, though it doesn't have the status of a formal model, categorical analysis reveals characteristics which *do* help identify types of events that produce larger and smaller effects (Nelson, 2008). While these are general and descriptive findings, they are adequate to provide the sort of advantage Nelson's predictions show, simply as a matter of experiential learning.

I am pleased that May and Spottiswoode took the time to attempt an explanation of the GCP data deviations, though it seems to me they should have thought more deeply about various issues. They confuse or conflate various levels of description, and they make unexamined assumptions. I don't have a fundamental problem with an "experimenter effect" as a contributor to deviations from expectation in psi experiments, even in the GCP data. But there are no good reasons to think it is all or even most of the source. May and Spottiswoode make two separate attempts to persuade us otherwise. Their DAT explanation, even if applicable, fails because it is unable to discriminate between appropriate models. Their attribution to Nelson as experimenter fails because their assumptions about who is the source of predictions is faulty. It is, however, useful to think through these issues. They help us understand the experiment, and stimulate efforts to make it a better research vehicle.

References

Bancel, P., & Nelson, R. (2008). The GCP Event Experiment: Design, analytical methods, results. *Journal of Scientific Exploration, 22*(3), 27.

Dobyns, Y. (1993). Selection versus influence in remote REG anomalies. *Journal of Scientific Exploration, 7*(3), 259–269. Comment, RetroPsychoKinesis Project, http://www.fourmilab.ch/rpkp/dobyns.html

Dobyns, Y. (1996). Selection versus influence revisited: New method and conclusions. *Journal of Scientific Exploration, 10*(2), 253–267.

Dobyns, Y., & Nelson, R. (1997). *Empirical Evidence against Decision Augmentation Theory.* Technical Note PEAR 97005, Princeton Engineering Anomalies Research, Princeton University, School of Engineering/Applied Science.

Nelson, R. (1997). *Multiple Field REG/RNG Recordings During a Global Event, Parts I & II.* http://noosphere.princeton.edu/ejap/gaiamind/1997_2a.html [Originally published in *The Electronic Journal of Parapsychology, eJAP*]

Nelson, R. (2008). The Emotional Nature of Global Consciousness. *Proceedings of the Bial Foundation 7th Symposium, Behind and Beyond the Brain*; Porto, Portugal; 26–29 March 2008.

Nelson, R., & Bancel, P. (2009). Response to a letter from Helmut Schmidt. *Journal of Scientific Exploration, 23*(4), 510–516.

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.

Nelson, R., Boesch, H., Boller, E., Dobyns, Y., Houtkooper, J., Lettieri, A., Radin, D., Russek, L., Schwartz, G., & Wesch, J. (1998). *Global Resonance of Consciousness: Princess Diana and Mother Teresa*. http://noosphere.princeton.edu/ejap/diana/1998_1.html [Originally published in *The Electronic Journal of Parapsychology, eJAP*]

Nelson, R., Bradish, G., Dobyns, Y., Dunne, B., & Jahn, R. (1996). FieldREG anomalies in group situations. *Journal of Scientific Exploration*, *10*(1), 111–141.

Nelson, R., Jahn, R., Dunne, B., Dobyns, Y., & Bradish, G. (1998). FieldREG II: Consciousness field effects: Replications and explorations. *Journal of Scientific Exploration*, *12*(3), 425–454.