

LETTERS TO THE EDITOR

A Puzzling Aspect of the “Global Consciousness Project”

The experiment reported by Bancel and Nelson (Bancel & Nelson, 2008; Nelson, 2001) utilizes a global network of binary random number generators (RNGs). Each RNG produces 200 random bits during each 1-second time interval. These 200 bits are added and the deviation from average is expressed by a Z-score. Thus each second of an experimental period (surrounding some major world event) provides an individual Z-score from each RNG. For their primary evaluation the authors add the Z-scores from all RNGs and thus obtain (after normalizing the sum) one “composite Z-score” for each second. Then they add the squares of these composite Z-scores over all seconds of all experimental periods and find that the sum is significantly larger than expected by chance.

Among many additional statistical evaluations only one is relevant for the following: If, rather than first combining the Z-scores of the individual RNGs to one composite Z-score, one simply squares the Z-scores of the individual RNGs and sums these values over all RNGs and all test seconds, this sum is not significantly large, i.e., the individual RNGs seem to behave quite normal.

The numbers of participating RNGs varied from 3 to over 65. The RNGs were well built and tested, but not all identically designed. And one may assume that the design details of the RNGs are irrelevant.

The Many-Experimenter Puzzle

In order to highlight a puzzling aspect of the experiment, consider a network of N RNGs and imagine for a moment that each RNG has two outputs, a primary output as before and a secondary output that gives the inverse of the primary output: If the primary output gives a 0 or 1, then the secondary output gives a 1 or 0, respectively.

Imagine next that there are two experimenters, (A) and (B). Experimenter (A) uses the primary outputs of the RNGs, whereas experimenter (B) uses the primary outputs of the odd-numbered RNGs, but the secondary outputs of the even-numbered RNGs. Then (A) and (B) use different RNGs (i.e., RNGs with different random number outputs). But since the Bancel-Nelson study had three different kinds of RNGs anyway, experimenters (A) and (B) seem equally justified in their choice.

If, for a particular test-second, experimenter (A) observes the Z-values $Z_1, Z_2, Z_3, \dots, Z_N$, then experimenter (B) observes the Z-values $Z_1, -Z_2, Z_3, -\dots \pm Z_N$, and the resulting composite Z-scores on which the final result is based are

$$Z_{\text{comp}}(\text{A}) = (Z_1 + Z_2 + Z_3 + \dots + Z_N) / \sqrt{N} \quad (1)$$

and

$$Z_{\text{comp}}(\text{B}) = (Z_1 - Z_2 + Z_3 - \dots \pm Z_N) / \sqrt{N}. \quad (2)$$

Additional experimenters could have made other choices in selecting primary or secondary outputs. For a general experimenter the composite Z -score can be written as

$$Z_{\text{comp}}(S_1, S_2, S_3, \dots, S_N) = (S_1 \times Z_1 + S_2 \times Z_2 + S_3 \times Z_3 + \dots + S_N \times Z_N) / \sqrt{N} \quad (3)$$

with $S_k = \pm 1$. Here, S_k is +1 or -1, if the primary or secondary output of RNG # k was used.

This shows that (even in the absence of a physical secondary output) an experimenter is free to choose Equation 1 or Equation 2 or any of the expressions (Equation 3) as the composite Z -score. One choice is as justified (or as arbitrary) as any other.

In agreement with the actual experiment (Bancel & Nelson, 2008), we will assume that the total result observed by experimenter (A),

$$\text{Total}(\text{A}) = \sum_{\text{ALL SECONDS}} \{(Z_1 + Z_2 + Z_3 + \dots + Z_N)^2 / N\} \quad (4)$$

is significantly large, and that the sum based on the variance of the individual RNGs

$$\text{Single}_{\text{Tot}} = \sum_{\text{ALL SECONDS}} \{(Z_1^2 + Z_2^2 + Z_3^2 + \dots + Z_N^2) / N\} \quad (5)$$

is not extra large.

Next we may ask whether perhaps the totals for all possible experimenters

$$\text{Total}(S_1, S_2, S_3, \dots, S_N) = \sum_{\text{ALL SECONDS}} \{(S_1 \times Z_1 + S_2 \times Z_2 + S_3 \times Z_3 + \dots + S_N \times Z_N)^2 / N\} \quad (6)$$

are significantly large, so that we don't have to worry about the ambiguity introduced by the many possible experimenters. That is not the case, however, because the average value of $\text{Total}(S_1, S_2, S_3, \dots, S_N)$, averaged over all S_k choices is easily seen to be equal to $\text{Single}_{\text{Tot}}$ from Equation 5, which was not significantly large.

If the great number of possible experimenters could get widely different results, one certainly wonders why Bancel and Nelson were so lucky as to obtain a highly significant score.

Goal Orientation of Psi as a Possible Explanation

Previous studies (Schmidt, 1987) have emphasized the goal-oriented features of psychokinesis. Take as an example a binary RNG that is activated once per

second to generate a 1 or a 0 that is internally recorded and externally displayed by a flash from a red or green lamp, respectively. Then, a PK subject aiming at an increased generation rate of 1's need not worry about the complex inner workings of the RNG but can succeed by merely focusing on increasing the number of flashes from the red lamp.

Furthermore, PK experiments with pre-recorded random events have shown that the PK mechanism still works when there is a long time delay between the generation (and recording) of the random events and the display of these events to the subject. It seems that the subject's mental effort, or the subject's wish to see red flashes, can retroactively affect the earlier decisions of the RNG to record an excess of 1s.

While no satisfactory theory for such a non-causal mechanism is available, a simple mathematical model (Schmidt, 1975) shows that the apparent action of the subject into the past need not lead to logical inconsistencies, and that we may be dealing with a primary mechanism that defies explanations in more familiar, plausible terms.

The finding that a person's wish for a certain outcome may affect earlier random decisions that lead up to this outcome may be central for the interpretation of the Bancel-Nelson experiment. Certainly the situation here is more complex in that we have a large number of persons who wish for the same outcome. There are, for example, the main experimenters and their assistants, the persons who have arranged the eggs at the different locations, and a large number of people who are emotionally captivated by the concept of a "global consciousness" and may have even financially contributed to the experiment. All these people can view the outcome, accompanied by their good wishes, at least on the Internet.

At this time we have no experimental information on how these good wishes of many people might add up, whether, in particular, the main experimenter is the overriding psi-source, or whether the large number of well-wishers are a vital factor. Whatever the answer to this question may be, however, we can now interpret the many-experimenter puzzle in this way:

After preliminary tests, Roger Nelson decided to base the evaluation on the specific composite score

$$Z_{\text{comp}}(A) = (Z_1 + Z_2 + Z_3 + Z_4 + \dots + Z_N) / \sqrt{N}. \quad (7)$$

Nelson was apparently convinced and able to convince his followers that this was a natural and reasonable measure, whereas from the previous arguments this measure appears arbitrary and as valid as, for example,

$$Z_{\text{comp}}(B) = (Z_1 - Z_2 + Z_3 - Z_4 + \dots \pm Z_N) / \sqrt{N}. \quad (8)$$

After Nelson had made his decision, the goal for the experiment was set. Experimenters and followers were now wishing for a high score based on $Z_{\text{comp}}(A)$, and only this score was prominently displayed at the end. It was this score everybody focused on, and the participants simply succeeded in their specified task.

The concept of goal orientation of PK may serve the Bancel-Nelson experiment in two respects. First, it solves the many-experimenter puzzle and, second, it reduces the phenomenon to something familiar from earlier experiments so that we need not postulate any basically new mechanism.

References

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

Nelson, R. (2001). Correlation of global events with REG data. *Journal of Parapsychology*, 65, 247–271.

Schmidt, H. (1975). Toward a mathematical theory of psi. *Journal of the American Society for Psychical Research*, 69, 301–319.

Schmidt, H. (1987). The strange properties of psychokinesis. *Journal of Scientific Exploration*, 1, 103–118.

HELMUT SCHMIDT
1107 8th Street
Las Vegas, NM 87701
hhwschmidt@msn.com

To the Editor: We are pleased that Helmut Schmidt has taken the time to write a commentary on the Global Consciousness Project (GCP), and we appreciate the opportunity to respond to his letter in the *Journal of Scientific Exploration* (Schmidt, 2009).

Helmut Schmidt's Interpretation

The first part of Schmidt's letter restates our finding, in a somewhat convoluted way, that the primary effect in the formal GCP experiment is in the trial correlations and not in the trial variances. Then he proceeds to an effort to interpret the experiment in terms familiar to his own work. For a number of reasons, we believe that effort fails, though it does provide a focus for us to describe the more coherent and data-driven approach we take.

Schmidt proposes that the GCP aggregate result should be construed as an intentional or goal-oriented PK effect, similar to interpretations of micro-PK experiments in which subjects attempt to change the output of a “target” RNG. Specifically, he suggests that the Project experimenters and RNG hosts collectively exert a PK effect, which results in significant deviations of the chosen target statistics. Schmidt recognizes that, in the case of the GCP, defining goal-oriented PK is complicated by the variable involvement of the many people associated with the project. Accordingly, one needs to ask how a common “target” or goal can be assumed for the pool of individuals who presumably act as PK sources. Schmidt suggests that the Project effectively resolves this “puzzle” by publicly defining the statistic for each global event. He further suggests that, because evidence for

a retro-PK effect has been reported in the literature, there is no inconsistency with individuals learning of the target statistic after data have been recorded (as is generally the case for the GCP). He concludes that goal-oriented PK offers a viable interpretation of the GCP results, and he believes that since this is a familiar framework, it should be preferred over other approaches that may introduce unfamiliar mechanisms.

To begin, given Schmidt's obvious care in reading our papers, we are gratified to see that he agrees that the real question is how we should interpret the result. Unfortunately, he misses or glosses over the essential point of our 2008 paper on the GCP event experiment, which is to describe a research program that begins by identifying data deviations via a formal experiment and then proceeds to characterize the structure of those data deviations (Bancel & Nelson, 2008). Our goal is not to present or defend a theory, but to provide a basis for assessing and comparing different explanatory views, including the one Schmidt proposes.

Schmidt's goal-oriented PK model for the GCP effect is not specific, but appears to depend at least partly on what he calls a "non-causal mechanism" that can reach into the past as well as the imminent future to create anomalous structure in the random data. While there are substantial differences, this bears some similarity to the Decision Augmentation Theory (DAT) idea that psi effects are achieved by making the right choices (for example, timing) based on accessing information about the future outcome of a trial or experiment (May et al., 1995; Schmidt, 1969).

In any case, we find Schmidt's model unsatisfactory on several grounds: It is structurally incorrect, as we show in some detail below, and second, it replaces one mystery with another—instead of a local, concurrent effect of consciousness, it proposes an effect of consciousness reaching back and forth across time. While an observer using precognition (in the DAT version of goal-oriented PK) to "press the button" at just the right time to produce a desired future effect has captured the imagination of various commentators in parapsychology, and has even attracted some modeling efforts, it is possible to do better. The GCP database is deeply dimensioned and it allows us to distinguish models analytically. It is constructed to facilitate adjudication between models that require mental time travel and those that propose immediate effects via a field-like interaction of consciousness and environment.

Schmidt appears to believe that we propose a global consciousness effect. This is not the case, formally. Instead we propose an experiment designed to test for such an effect, defined operationally. To this end, the GCP experiment evaluates the hypothesis that our data will show correlations with global events. In particular, we test for internal correlations in the network at the time of global events, and then proceed to assess physically relevant aspects of those correlations, such as their spatial and temporal structure. These are among the design aspects that make the GCP unique. Hence, we disagree with Schmidt's unstated but essential assumption that the formal result by itself provides enough information for choosing between approaches. That is far too narrow a stance for a subject

we understand so poorly, especially given the richness of the 10-year database and the fact that the GCP design explicitly allows for detection of effects that may be other than goal-oriented. As we indicated in our 2008 paper, we have made significant progress in identifying correlation structure that is inconsistent with simple goal-oriented models, whether of the PK variety proposed by Schmidt or selection-type models such as DAT. A series of papers detailing our findings is in preparation.

We strongly disagree that familiarity with a past point of view is a reason for preferring it in a new experimental situation, the more so if this requires stretching the “familiar” parameters. Indeed, should the GCP effects prove to be of general relevance to psi phenomena, it may help illuminate what models make sense for more conventional parapsychological experiments such as micro-PK and remote viewing.

The preceding comments aside, Schmidt’s proposal is inconsistent with the experimental procedures. Formal registration of event parameters is made by Roger Nelson, with contributions by one or a few other “predictors” in about half the cases. Only infrequently are other collaborators (analysts and egg hosts) aware that an event has been registered prior to the calculation and posting of its outcome. Results are posted to the GCP Web site after analysis, and only then do most of the people associated with the GCP know about the event. We can determine from site usage statistics that there is little or no traffic to the prediction registry before this public posting of the results. If collaborators and associates are not aware that an event is registered as a formal test of the GCP hypothesis before the result is calculated, they cannot influence the outcome of the calculation as Schmidt proposes, certainly not by intention. Thus the only reasonable model for Schmidt is that the GCP’s result is an experimenter PK effect due solely to the project’s director, aided perhaps by a retro-PK influence from assorted people who look at the results. While this attribution may appear to be compatible with Schmidt’s position, we reiterate that it begs a principal motivation of the GCP: the details of data structure can be used to distinguish between competing interpretations, including experimenter PK and DAT models, or some hitherto unknown capacity of consciousness.

One can attempt to refine Schmidt’s approach and we encourage such efforts. For example, one could argue that it is not necessary that people involved with the project be informed of the precise statistic for each event in order to exert a PK influence. The similarity of statistics from event to event (the statistics vary chiefly by the assigned duration of event periods) allows a sufficient commonality of goal-oriented targets for an extended group of people. But this does not resolve the prediction ignorance issue, nor does it address the fact that many skeptics also follow the project closely (they would presumably expect or even wish for a null outcome). Given the speculative nature of the arguments, there is a risk of complicating an already vague theoretical approach beyond reason. In the end, we believe that maintaining the course of analytical modeling is the most productive approach. We need to continue with the characterization phase of the experiment

that examines spatial, temporal, and correlation structure in the 5-sigma data set and use the insights gained to constrain models.

Structure in GCP Correlations

We present a model based on the empirical data and show analytically that it is superior to Schmidt's proposal. This data-based modeling produces concrete arguments for or against well-formed alternative interpretations, but there are also more general arguments worth mentioning. Schmidt suggests that the anomalous GCP effects are due to the main experimenter, who knows the predictions, and perhaps to the rest of the people who are involved in the project via retrocausal influence. His model considers only the simple primary effect and ignores the possibility of other structure in the data. However, it is the case that in addition to the primary statistic that is specified in the GCP hypothesis registry and posted to the results page, there are several other independent effects and correlations. These, as we will see, show structure that was not expected by anyone involved in the experiment, including the main experimenter, prior to about 2007, some 9 years into the project. It requires an extraordinary stretch of the imagination to see how any kind of experimenter effect model could apply to or accommodate these unpredicted and unexpected effects, which were literally unknown for most of the project's history.

For the record, we might add that Roger Nelson, the main experimenter, generally is not aware of any intention to affect the world-spanning network of RNGs during the events or during the analysis. Rather, he envisions the network as an instrument set up to monitor what is happening, not as a target for anyone's intention. Indeed, the GCP has a policy of refusing occasional offers from well-meaning people and groups to intentionally affect the network.

But let us proceed to the quantitative modeling.

We can show that there are exactly two orthogonal measures of linear correlations in the GCP data. One of these is the primary statistic we specify in most of the formal hypothesis tests. It is a measure of network variance and can be regarded as a correlation of means. This is the only statistic Schmidt considers. The second has a similar mathematical structure and represents a correlation of variances. Both measures are significantly large compared with expectation. Moreover, these two measures are correlated with each other during the formal events, and in addition they both show a decline in magnitude as the separation of the RNGs increases. A simple model such as Schmidt's addresses only one of these five separate manifestations of structure, and so is obviously not viable. But we can go further, to a direct calculation of model fits to the data.

The primary GCP effect can be understood as non-zero correlations between the 65 or so RNGs (the number grew during the first years, but stabilized in about 2004). We show that these pair correlations in the GCP data depend on degrees of freedom (df), and this result helps to identify and distinguish competing models. The bottom line is an argument in favor of a field-type model and against DAT-like or experimenter effect scenarios.

The network variance measure used in most of the formal hypothesis tests can be expressed as an average pairwise correlation or a sum across all pair products. In this analysis we look at the dependence of the pair-correlation sums per event (“ppsum”) on the df by using regression analysis. The basic regression model is $ppsum = df^K + n(df)$, where df^K is the effect and $n(df)$, a noise term. We can think of these as signal and noise. We know $n(df)$ very well because a small signal-to-noise ratio assures that the data are essentially normally distributed. Thus, $n(df)$ is $\text{Sqrt}[df]$, where df is the total number of pair-products for the event. The model says that the effect causes $ppsum$ to increase as some power of df . We seek to determine the exponent K from a (nonlinear) regression. Two cases stand out: $K = \frac{1}{2}$ and $K = 1$.

$K = \frac{1}{2}$ is the prediction of DAT and similar trans-temporal models including Schmidt's. It says that because the noise and effect both increase as $\text{Sqrt}[df]$, the results will show a constant Z across events. Accumulating more data (through increasing the event duration or the number of eggs) does not increase the signal-to-noise ratio in this model. In contrast, $K = 1$ is the prediction of a field-type model. It says the effect accumulates linearly with df , thus beating out the noise, which is scaled by $\text{Sqrt}[df]$.

How can we decide which model best represents the actual data in the formal GCP event experiment? Linear regressions for $K = \frac{1}{2}$ and 1 both give reasonable fits. But we can do better. Rather than comparing regressions using fixed guesses for K , we can calculate a non-linear fit directly on K . This gives $K = 0.59 \pm 0.12$. If we go no further it looks like the $K = \frac{1}{2}$ model is favored. However, we now need to fold in what we have learned from other analyses about structure in the GCP data.

We know that there is a distance effect where the inter-RNG correlations (ppsums) decrease as the geographic distance between RNGs increases, with a distance scale on the order of 8,000 to 10,000 Km. We also have evidence from time correlations between the two orthogonal measures of network structure that the effect duration is usually a few hours, and we know it is typically shorter than the event prediction. The question is: How do these parameters affect our determination of K ? For our purposes here, we will look at the time case in some detail.

If the effect lasts only a few hours on average, then long events (that is, the prediction specifies an event longer than the effect) will add noise but not signal as the event length increases. This means that a straight-line regression against duration alone will have a weaker slope than a regression against the full df .¹ We find a positive slope for both, but the regression slope against the total df has a $Z \approx 4$, while the regression slope against only the duration part of df has $Z \approx 2.4$.

Now we can ask: What is K for a field-type effect when the event continues much longer than the actual effect? The answer is $K = 0$. This is because the model sees no increase in the effect as the event duration lengthens. The effect is constant (because the effect is fully “within” the event), while the noise still grows

as duration (and hence df) increases. To see this, consider that K is the exponent for the absolute signal strength as a function of df . If you add more df —but those data don't contain any signal—then there is no increase in the absolute signal. Therefore, $K = 0$ in that regime, i.e., once the event duration surpasses the 'actual' duration of the effect.

Using this model we can make a prediction: If we have a field-type effect with average duration of several hours, then $K = 1$ for short events and $K = 0$ for long events. This is because the event psums increase when the event is shorter than the effect ($K = 1$) and then are constant for events longer than that ($K = 0$). To test the prediction, we split the data into events longer and shorter than 12 hours. The non-linear regressions on K then give $K = 0.84 \pm 0.24$ for short events and $K = -0.07 \pm 0.16$ for long events.

This result is obviously consistent with the predictions of a $K = 1$ field-type model, even though we have not yet fully exploited the information in the actual data. A similar, but more involved, analysis shows the distance effect also tends to affect K , by an amount that is equivalent to about 0.2. Combining the two effects, duration and distance, we end up with a measured K pretty close to 1. We do not present the details for the distance calculation here, but the modeling approach is similar to that for duration. Both of these analyses are fully documented in a new *JSE* article that is in preparation.

The duration analysis by itself yields a powerful distinguishing result. For events shorter than 12 hours, it closely matches the predictions from a field-like data anomalies model, and indicates that longer durations add nothing to the effect. It also explains why a DAT model looks viable when we use the whole data set: because the whole set ignores the question of (real) event durations, it averages the crossover from $K = 1$ to $K = 0$, thus accidentally (in this interpretation) agreeing with DAT. Our result says that events longer than 12 hours will fully bracket the effect, on average. To see the argument, consider a simplified example. If all GCP effects were to persist for 6 hours or less, events longer than 6 hours in duration would all contain the same full effect; we would see no difference in the average deviation of the absolute psum as duration increased from 6 to 12, or 18, or 24 hours. On the other hand, a 6-hour event *would* have more signal than a 3-hour event. Thus, we would expect to see the effect grow from 0 to 6 hours and then flatten out for longer events.

Simply stated: we expect a crossover from $K = 1$ to $K = 0$ if the average effect duration lies between our shortest and longest events. The result shown in our regression analysis is consistent with this model. Independently, we have determined that deviations in the event data show a characteristic time structure via correlation of the two independent network measures (addressing RNG means and variances). This correlation maximizes for event durations of approximately half an hour to 2 hours, then diminishes to non-significance over the next few hours. Thus, we have both empirical evidence for our assumption of an intermediate effect duration and a fully consistent modeling result based on that assumption.

The formal event data contain structure that points to true data anomalies and a field-like model as the most competent approach of those we have tested, empirically superior to a DAT model or a goal oriented, retrocausal model as proposed by Schmidt.

ROGER NELSON

Director, Global Consciousness Project

Princeton, New Jersey

rdnelson@princeton.edu

PETER BANCEL

Global Consciousness Project & Institut Métapsychique International

Paris, France

Note

¹ The full df in this case is the total number of pair-products in an event = #seconds \times N \times $(N - 1)/2$, where #seconds is the duration, S . So the full df depends on both S and N . However, increasing N will always add some signal (modulo an N dependence coming from a distance effect), whereas increasing S adds only noise once S is longer than the ‘real’ effect. So at large df, a regression on duration alone will be most sensitive to the crossover, while a regression including the full df will get more weight from the number of pair-products at high df.

References

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

May, E. C., Utts, J. M., & Spottiswoode, S. J. P. (1995). Decision Augmentation Theory: Toward a model of anomalous mental phenomena. *The Journal of Parapsychology*, 59(2), 195–220.

Schmidt, H. (1969). Precognition of a quantum process. *The Journal of Parapsychology*, 33(2), 99–108.

Schmidt, H. (2009). A puzzling aspect of the “Global Consciousness Project” [Letter to the editor]. *Journal of Scientific Exploration*, 23, 507–510.