

Kommentare zu Etzold: Wie zuverlässig sind die Ergebnisdaten des Global Consciousness Project?

HOLGER BÖSCH

Wo liegt die Messlatte zur Beurteilung des GCP?

Kaum ein Datensatz ist frei von Unregelmäßigkeiten. Eine, 100 oder gar 100000 signifikante Analysen in einem großen Datensatz wie dem des GCPs sind nicht beeindruckend oder gar wissenschaftlich bedeutsam, wenn sie nicht sorgfältig geplant und systematisch durchgeführt wurden.

Wesentliche Kritikpunkte

Etzold untersucht die, wie er es nennt, Ergebnisdaten des GCPs. Das heißt er übernimmt die dort aufgeführten „Ereignisse“ und deren Analysemethoden so wie sie auf den Internetseiten des GCPs formuliert sind und reanalysiert sie. Zuvor macht Etzold noch auf einige theoretische Schwächen des Projekts aufmerksam. Aber es fehlt ein roter Faden. So präsentiert Etzold in seiner Einführung beispielsweise ein „Gesamtergebnis“, ohne dem Leser zuvor verständlich gemacht zu haben, wie das GCP Daten aggregiert und worauf sich die Analyse bezieht. Der folgende Abschnitt „Methoden“ beschreibt nicht die von ihm eingesetzte(n) Methode(n) sondern geht recht oberflächlich auf die Organisation und Gewinnung der GCP Daten ein.

Etzolds Argumentation ist an einigen Stellen nicht überzeugend. Am weitesten von einer überzeugenden Argumentation entfernt sich Etzold aber bei seinen Analysen zum „Experimentatoren-Effekt“, auch wenn sein Ansatz, den Einfluss des „Experimentatoren-Effekts“ zu den im GCP verwendeten „Ereignisklassen“ in Beziehung zu bringen, sehr innovativ ist. Aber wie kommt Etzold überhaupt dazu einen „Experimentatoren-Effekt“ für das GCP zu postulieren? Die Basis dafür sind eine Hand voll unsauberer Analysen. Die Autoren haben sich verrechnet und erst zu einem späteren Zeitpunkt ihre Fehler bemerkt und behoben, d.h. diese Fehler sind im GCP dokumentiert. Das kann passieren. Einen „Experimentatoren-Effekt“ zu vermuten ist deshalb doch sehr weit hergeholt.

Wer den Artikel weiter liest, erfährt, dass einige Analysen innerhalb des GCPs „Ungereimtheiten“ aufweisen. Was aber bei der Komplexität des Projekts, das im Wesentlichen von nur einer Person – Roger Nelson – betreut wird, nicht verwunderlich ist. Am stärksten sind Etzolds Analysen zum „harten Kern“ der GCP-Daten, auch wenn man sich fragen muss, warum der Leser erst an dieser Stelle erfährt, auf welche Weise die Daten erzeugt werden.

Die von Etzold angerissene Diskussion über die Kontrolldatenproblematik und die anschließend von ihm eingesetzte Methode zur Gewinnung von Kontrolldaten ist grundsätzlich überzeugend, aber nur sehr kryptisch dokumentiert und wenig anschaulich. Der Leser wird mit der Aussage, dass die „Versuchsdaten für 90 ‚Kontrolldaten‘“ signifikant sind,

völlig alleine gelassen. Eine Erläuterung und Diskussion dieses Sachverhalts fehlt. Etzolds Vorgehen ist in vielerlei Hinsicht ebenso kursorisch wie das Vorgehen der Projektteilnehmer des GCPs.

Anspruchsniveau zur Beurteilung des GCP

Vielleicht sollte man die Messlatte zur Beurteilung des GCPs ohnehin nicht zu hoch legen, denn es handelt sich dabei um ein Projekt, das auf systematische Weise Daten aggregiert und auf weniger systematische Weise analysiert. Es ist ein offener und kreativer Prozess, der von seinen Initiatoren und „Imaginatoren“ sicher noch lange nicht abgeschlossen wurde. Es ist ein komplexes Internetprojekt mit all den damit zusammenhängenden Stärken und Schwächen. Der nichtlineare Aufbau macht es äußerst schwierig, einen Überblick zu sämtlichen Methoden und Analysen zu gewinnen und es gehört sicher zu den Verdiensten Etzolds einige der dort vergrabenen Analysen aufgearbeitet zu haben.

Wenn man berücksichtigt, dass es sich beim GCP um ein Projekt handelt, das sich in einem permanenten Wandlungsprozess befindet, um ein Phänomen zu untersuchen, dass es vielleicht gar nicht gibt, darf man nicht davon ausgehen, dass die Ergebnisse den wissenschaftlichen Standards genügen. Ich vermute, dass es um den wissenschaftlichen Status des GCPs wesentlich schlechter steht als Etzolds Artikel vermuten lässt. Das alles entschuldigt aber nicht Etzolds Fehler. Er hat sich vorgenommen die „Ergebnisdaten“ auf ihre „Zuverlässigkeit“ zu prüfen und das hätte meiner Ansicht nach wesentlich systematischer erfolgen müssen. Vielleicht wäre es besser gewesen wenn er sich allein auf die Kontrolldatenproblematik beschränkt und diesen Knackpunkt noch weiter vertieft hätte. Zu viel steht zu weit verstreut nebeneinander.

Grundsätzliche Überlegungen

Die entscheidende Frage ist doch, ob „Unregelmäßigkeiten“ in den GCP-Daten sind und wenn ja, was die Ursache dafür ist. Zunächst muss also die Frage beantwortet werden, ob die Daten überhaupt Unregelmäßigkeiten aufweisen. Natürlich lässt sich das nicht endgültig klären. Dafür gibt es viel zu viele Analysemöglichkeiten wie beispielsweise im einfachsten Fall unterschiedlich große Zeitfenster vor und nach einem „Ereignis“. Aber in aller Regel gibt es nur eine geringe Anzahl *sinnvoller* Analysemöglichkeiten. Mit deren Hilfe ließe sich untersuchen, ob insgesamt überzufällig viele Unregelmäßigkeiten in einem Datensatz vorliegen oder ob die Unregelmäßigkeiten, die in ihrer Häufigkeit unauffällig sind, an spezielle Ereignisse gebunden sind.

Eine Zeitreihenanalyse mit einem gleitenden Durchschnitt wäre bestens geeignet, Varianzerhöhungen um Ereignisse herum festzustellen. Vermutlich würden die Projektteilnehmer des GCPs argumentieren, dass nicht jedes Ereignis gleich sei und die Auswirkungen unterschiedliche Latenzzeiten hätten. Das ließe sich mit Hilfe von Ereignisklassen und unterschiedlich großen Zeitfenstern lösen. Damit ein solches Vorgehen sinnvoll ist, muss natürlich ein verlässliches Register existieren, das alle prognostizierten Ereignisse systematisch erfasst. Von dessen Qualität hängt viel ab.

Bevor man am Ende einfach schnell ein paar Signifikanztests durchführt, dürfte es für das Verständnis der komplexen Zusammenhänge sehr hilfreich sein, zunächst die Ergebnisse

rein deskriptiv darzustellen. Effektstärken eignen sich zur Deskription weitaus besser als p-Werte (Sedlmeier 1996). Aber die unangefochtene Königin unter den deskriptiven Verfahren bleibt die Graphik (Cohen 1990). Die *Kunst* wissenschaftlichen Vorgehens besteht zum großen Teil darin, komplexe Sachverhalte anschaulich darzustellen. Dafür sind einfache Tabellen und Graphiken bestens geeignet.

Literatur

- Cohen, J. (1990): Things I Have Learned (So Far). *American Psychologist* 45, 1304-1321.
- Sedlmeier, P. (1996): Jenseits des Signifikanztest-Rituals: Ergänzungen und Alternativen. *Methods of Psychological Research Online* 1, (4). Verfügbar unter: <http://www.mpr-online.de>

ROGER D. NELSON

How nearly perfect is the Global Consciousness Project ?

I appreciate Eckhard Etzold's serious and thorough assessment of the Global Consciousness Project (GCP). In his paper, he identifies a number of errors in the registration of the formal predictions, for which I am grateful. This is a very large and complex project, with a continuing, real-time engagement that demands daily input over months and years. In contrast to ordinary scientific research, which is reported *post facto* in judiciously edited papers, the GCP is a living and active process. It is presented to the public on the web (<http://noosphere.princeton.edu/>) as a growing historical transcript, and, as the Americans say, it is presented "warts and all", so that the documentation is completely transparent – nothing is hidden or re-drawn to look better than it is. Over the five years of its growth, there has been a natural accumulation of imperfections. This is regrettable, but in the real world it is arguably unavoidable. The important question, which we address here, is whether the data accumulated in this project are sound, constituting useful evidence for an anomalous effect of human consciousness in the physical world. Unsurprisingly, I have high confidence in the data, based on my deep and continuing involvement in every aspect of the project from its inception as a hugely interesting question, to the design of equipment and software, to the collection and archiving of data, to its analysis and interpretation. For those interested in the most recent and carefully refined assessment of the evidence for structure in the data, apparently associated with world events of importance to humans, I recommend reading Nelson et al. (2002), which is available as a preprint upon email request to rdnelson@princeton.edu.

I have many collaborators (among whom I count Etzold himself), and they have made the project possible. Several people have contributed a great deal of time and energy in various ways, especially by conducting independent analyses. Their work is important, and I deeply appreciate it, but I take responsibility for the documentation, the experimental design and analysis, and necessarily for errors and imperfections. A little over a year ago, I received an inquiry from Eckhard Etzold about mismatches in the prediction registry and the results table. Unfortunately, I was so absorbed in other issues at the time that I was distracted and

did not adequately address his questions. He has, quite properly, taken the route of a formal publication that forces me to pay proper attention.

Although I will not minimize the importance of exact and error-free documentation, I do not believe the imperfections Etzold describes have a deleterious effect on the scientific work, nor do they change the general conclusions or the interpretation of the data. I will discuss here the errors and their resolution, and the effect on the results as a whole, as well as Etzold's more general interpretive remarks. For the most part, he has done an excellent job, and the points he makes are useful. I will begin with an annotated response to the primary concerns he raises concerning the prediction registry.

First, he lists events that are in the results table but not in the registry. One of his identifications, "Astrological Moments, 20011209", is in error. The event was in the registry. (I confess that I am consoled that Etzold also has made a mistake of this kind, for it emphasizes the difficulty of managing perfectly such a large and complex corpus of information.) He is correct in the other cases he lists, all of which are predictions made by various people via email. These items are all fully specified in the archived email correspondence from the predictors, but four items were inadvertently not entered into the registry, even though they were fully described in the reports of the analyses. I have now added the original prediction specifications to the registry, with a highlighted note to identify them and acknowledge the delayed entry.

In a second list, Etzold identified five events in the registry that were not included in the formal results. Of these, two could not be analyzed, and three had been analyzed, but never added to the results table. The events and their disposition are as follows:

1. Clinton Affair Grand Jury, 19980911: The analysis was done but inadvertently was not entered into the composite results table. It has now been added to the table as event number 6, with p -value = 0.140. A highlighted notation has been added to the registry.
2. Conjunction of Venus and Jupiter, 19990223: This prediction was incomplete, and too vague for any analysis to be done. A highlighted notation has been added to the registry.
3. TV Broadcast of Oscars, 19990322: The prediction specified a comparison of data taken during presentations vs. commercial breaks. I could not watch the show, and thus could not do the analysis. A highlighted notation has been added to the registry.
4. Dow Jones closing exceeds 10,000, 19990329: The analysis was properly specified, but inadvertently was not completed. It has now been done and is included as event number 24, with p -value = 0.355. A highlighted notation has been added to the registry.
5. Sir Peter Blake murder, 20011206: This analysis was done, but inadvertently was not included in the results table. In addition, the predictor gave an incorrect date (the 7th) but later corrected it. The event has now been added as event number 97, with p -value = 0.923. A highlighted notation has been added to the registry.

As a result of Etzold's extraordinarily diligent assessment, these three additional events are now included in the formal results table and calculations, raising the total from 117 to 120 at the time of this writing. Two of the three have modest positive deviations, but these contributions are balanced by a fairly strong negative deviation in the third case. The net result of these corrections is negligible.

I will now turn to more general points in Etzold's paper. Some of his critiques are correct and useful; in some cases, he is wrong either factually or in his interpretation. Again, I take

responsibility because my descriptions of difficult technical issues have not been sufficiently clear. For example, he says that the data are uploaded daily, which is not correct. For most eggs, the Internet connection is continuous, and the data are sent in 5-minute packets, resulting in a near real-time registration of the data. About 10% of the eggs have automatic dialup modem connections, and these do wait to send the accumulated data for several hours or a full day whenever the connection is up.

He gives an incorrect description of the exclusive or (XOR) process that guarantees an unbiased mean output from the REG/RNG devices. He proposes that a physical source operating at a certain frequency could induce a bias. But his suggestion is not viable. In the majority of the devices the XOR uses a randomized pattern of all bytes with equal numbers of 1 and 0. In the rest, the basic XOR is of two truly random sequences, plus a software XOR that inverts alternate bits. Moreover, we have empirical evidence that the most likely physical and environmental sources do not affect the devices, based on both calibrations and experiments. See, for example, Radin's analysis of diurnal variation in the GCP data (Radin 2002). There is no regular variation of this sort, and there should be if, for example, EM fields matter to the REG devices.

Etzold asserts that analysis of different predictions on the same day, for example using a meanshift and a variance measure, is not appropriate. I can see his point, but on the null hypothesis, mean and variance are independent parameters. Moreover, in all cases, the predictions that specify various different measures are independent, and generally do not use the same data, even though the same dates are involved. Specific examples from his list of "problem" cases should help illuminate this misunderstanding: New Years, Times Square, 1998 looks at *meanshift* in the data from all eggs at one time, rather than assembling the "midnight" data across time zones. The Euro vs US prediction uses a *device variance* measure to compare the behavior of two groups of eggs. The Maxi vs Mini comparison is of two differently defined groups, and uses the *meanshift* measure, which is independent by the null hypothesis. It can be argued that there might remain a tiny correlation via something like a "stacking effect" if there is an anomalous influence. In any case even a brief look at the actual results reveals that these three analyses bear no resemblance to each other. Similar considerations apply to the rest of Etzold's list.

Every effort was made from the beginning to establish a clear overarching hypothesis, namely that we would find structure in the data correlated with engaging global events. However, Etzold seems to believe that we should have been testing hypotheses about what kinds of events, or what particular events would generate an anomalous effect. His tables, especially Tables 1, 4, and 5 and a number of comments in his text address this issue. Of course the reality is that we could not know what to expect from this unique experiment, and we had, in the beginning, to make intuitive guesses that we hoped might allow us to learn enough to make better and more detailed predictions based on accumulating experience. As part of this effort, we have done a great deal of exploratory work that is presented as context for the formal analyses, or presented in a completely separate category of informal analysis in dedicated web pages. Despite the huge database and the long-term research, it is easy to see that we are only at the beginning of this work. Interestingly, when one examines the results in Etzold's Table 4 and 5 carefully, paying attention to the effect sizes, the correspondence is not as bad as he suggests. Consider, for example, correcting the obvious

misplacement of the *Gesellschaftliche Events und Medienereignisse* (Social and Media Events) by the 15 people who answered his poll. As anyone familiar with questionnaire design would observe, the low ranking is likely due to the category name or the form of the question, rather than an intrinsic quality of the events in question. Not only does this group contain the most widely known “global” events *because* of media coverage, it includes prime examples of our hypothesis of global engagement such as New Years and Y2K. But I do want to be clear that I respect Etzold’s work as a valuable effort in a useful direction. It is something that should be pursued, with improved polling instruments and experimental design. I should note that we also have done an assessment of effect size by category, with similar results.

The “experimenter effect” vs. “world consciousness” opposition that is a major focus in Etzold’s paper is a misleading expression of an important issue. I am also interested in this question, and it is a frequent topic addressed in the “Informal Analysis” page on the GCP website. At this point we can do little more than identify conflicting patterns of evidence, indicating that both of these potential sources of effect (and probably others) contribute to the results. There are some persuasive events that Etzold uses to make his point about experimenter effect (see his Table 3). But this is obvious data selection to support his argument, and it is easy in the history of the project to find countervailing outcomes. In some cases (e.g., New Years and September 11 2001) the unexpected patterns in the data are so clearly correlated with the events, and are so obviously generalized across the network, that it stretches credulity to attempt an explanation such as is represented in Etzold’s Table 3. Etzold himself says, and I agree, that a single event in such an experiment as the GCP cannot be expected to answer the primary question: the signal to noise ratio is far too small. It is bemusing to note that he goes on to interpret exploratory reports such as the Brahms’ Requiem example as evidence for an “Experimenter Effect”. Recent papers (Nelson 2002; Radin 2002; Nelson et al. 2002) provide a fuller impression of how unlikely it is that the experimenters’ expectations alone are driving the primary results.

Regrettably, Etzold takes other apparent liberties, including selective quoting, to support his point of view. I suspect this is unconscious rather than deliberate, and indeed it is perhaps a result of reading in a non-native language. For example, in his introduction, he says that the term “weltweite Bewusstseinsresonanz” (world-wide consciousness resonance) is misleading – but that is his term, not mine. I do use the term “global consciousness” but I acknowledge and use language of this sort as metaphor – the homepage introduction to the GCP website includes a paragraph to make this explicit. Etzold goes on to suggest that his term implies that the REG devices we use would have to be distributed over the earth in proportion to the population and states that this is not the case. In fact, there is no justification for this inference from any experimental data. Indeed, the best empirical information we have about anomalous effects of consciousness on REG devices is that they are “non-local” and that distance is not a relevant consideration, much less an implied aspect of an inferred “consciousness resonance”. I concede that in an introduction or a discussion, an author is free to express his assumptions, but I expressly reject these particular assumptions about the distribution of the random sources, and those linking the size of anomalous effects to the population density – we simply do not have any evidence to justify such notions, appealing though they may be from a lay perspective.

Etzold quotes me as saying “... it is tempting to draw the conclusion that a true global consciousness is creating the anomalous effects.” But this is a selective quote: the second following sentence says, “We believe the data are correct and definitely show anomalous departures from expectation, but we can’t claim to have proven the existence of a global consciousness.” I go on to indicate that we must consider alternative interpretations, including a possible anomalous experimenter effect. The latter seems to be the explanation Etzold prefers for the anomalous correlations, and he attempts to provide evidence for it. But as I have indicated above, this is at best premature, and the model can only be supported by selective consideration of the evidence.

Etzold describes some of the 11 September 2001 results incorrectly. The reader should go to the GCP website, or read the publications that describe these results, including Nelson et al (2002), Nelson (2002), and Radin (2002). In particular, the point about a 4.81 sigma spike for one second of data being less unusual than Radin assumed is incorrect on several counts. No assumption is made about this exploratory finding: the contextual explanation gives the probability, which is about one in a million, and further explains that this is equivalent to one such occurrence in about 15 days. Nobody knows whether having such a point fall in the midst of the hour and a half of terror on September 11 is meaningful. Scientifically it is not especially persuasive, but then it is not presented as formal evidence – it is part of the description. The supporting quote from May and Spottiswoode, concluding that the EGG network did not respond, is taken out of context and does not refer to this “spike” (which is fortunate for their critical case since this one-second-spike description is almost irrelevant as evidence for or against the really interesting questions). Similarly, Etzold’s description of the actual formal prediction that I made for 11 September is almost entirely erroneous, even to the probability he cites (it is 0.028, not 0.035). He says that had I predicted an effect over 3 hours instead of 4, the result would not be significant, but the specification was made without knowledge of the results – which cannot be said of this criticism. I will forgo further detailed discussion, since the material is publicly available as noted above.

There are other criticisms that Etzold makes without proper grounds. Most notably he questions the quality of the actual data and the vulnerability of the analyses to biases. He specifically states that the most urgent problem for the GCP database is the lack of control data. He misunderstands or misstates my discussion of control data and asserts that I argue that no controls are available. This is incorrect, and though the usual modes are not applicable, there are several alternatives that we describe and use. In particular, we employ both resampling (which I certainly do not reject, as Etzold believes) and various permutation or Monte Carlo techniques that use the actual data so that we do not have to depend on theoretical comparison standards.

One of our current projects is a thorough re-assessment of all formal event analyses, and a complete overview of the empirical behavior of the network and all individual elements. This entails multiple analytical tracks, but the intent is simple: to establish in both depth and breadth the performance characteristics of the EGG network. One of our goals is to create an empirical description that can be compared directly with theoretical distributions, and that will constitute an unambiguous source of “control” data whenever such is actually needed. Enough of this work has already been done to provide confidence in the database of formal hypothesis tests as described on the website and in publications. Our calibration

procedures for qualifying the individual REG/RNG devices also provide high confidence in the theoretical comparisons that are used. Nevertheless, we have much to learn about the fine details, and consider the comprehensive assessment to be an important work. In this context, although it is an *ad hoc* procedure, the example that Etzold gives for assembling and testing what can be regarded as control data is well done. I appreciate the thought and effort, and for the reader, this example gives some indication of the complexity that confronts the analyst interested in the GCP data.

I commend Eckhard Etzold's initiative and extraordinary effort. He is strongly critical, so it is important to note that even his worst-case reading, which excludes over 25% of the replications, still leaves a highly significant result. Although I disagree with his arguments for these exclusions, and with some of his other assertions, I am grateful for his contributions to greater clarity and completeness in descriptions of the GCP database and results. I invite readers who are interested in the questions raised to go to the GCP website at <http://noosphere.princeton.edu/> and to read the published papers describing the project.

References

- Nelson, R.D. (2002): Coherent Consciousness and Reduced Randomness: Correlations on September 11, 2001. *Journal of Scientific Exploration*, in press.
- Nelson, R.D.; Radin, D.I.; Shoup, R.; Bancel, P.A. (2002): Correlation of Continuous Random Data with Major World Events. *Foundations of Physics Letters*, in press.
- Radin, D.I. (2002): Exploring relationships between random physical events and mass attentions: Asking for whom the bell tolls. *Journal of Scientific Exploration*, in press.

Stellungnahme des Autors

ECKHARD ETZOLD

GCP: Konzeption überdenken, Analyse automatisieren

Zunächst möchte ich beiden Kommentatoren danken für die Auseinandersetzung mit meiner Arbeit und dem Global Consciousness Project, das mir, bei aller Kritik, die ich habe, trotzdem sehr am Herzen liegt, und zu dessen Gelingen ich auch gern weiterhin beitragen möchte. Die Stellungnahmen der beiden Kommentatoren sind ausgesprochen konträr. Während Holger Bösch den wissenschaftlichen Status des GCP sehr viel schlechter einschätzt als es meine Untersuchung zu erkennen gibt, behauptet Roger Nelson, die Kritik, die ich formuliert habe, würde in keiner Weise die bisher gefundenen Ergebnisse und getroffenen Schlussfolgerungen schmälern. Zur Begründung führen sie beide eine Reihe von Fehlern an. Da jeweils verschiedene Fehler von den Kommentatoren als Beleg für entgegengesetzte Behauptungen bemüht werden, zeigen sie zunächst eines: das Projekt ist inzwischen zu komplex als dass es von einem Einzelnen noch voll durchdrungen und erfasst werden kann. Das gilt nicht nur für Außenstehende, sondern auch für die Beteiligten und Initiatoren des

GCP. Zum Umgang mit Fehlern in wissenschaftlichen Arbeiten allgemein ist m.E. eine grundsätzliche Bemerkung notwendig: Eine fehlerfreie Vorgehensweise gibt es leider nicht, und sie kann es auch nicht geben, bedingt durch die Grundverfassung des menschlichen Daseins als „ständige(r) Unabgeschlossenheit“ (Heidegger 1979, S. 235). Insofern gibt es auch nichts zu entschuldigen, es sei denn, in der Wissenschaft möchte man sich, was die Umgangsformen betrifft, den Kardinal Richelieu zum Vorbild nehmen: „Man gebe mir fünf Sätze von dem rechtschaffensten Menschen der Welt, und ich werde darinnen etwas finden, um ihn hängen zu lassen.“

Nun zu den strittigen Punkten. Bösch bemerkt: „Etzolds Argumentation ist an einigen Stellen nicht überzeugend. Am weitesten von einer überzeugenden Argumentation entfernt sich Etzold aber bei seinen Analysen zum ‘Experimentatoren-Effekt’,“ und fragt dann: „wie kommt Etzold überhaupt dazu einen ‘Experimentatoren-Effekt für das GCP zu postulieren? Die Basis dafür sind eine Hand voll unsauberer Analysen.“ Ein allgemeiner Experimentatoren-Effekt wurde von mir jedoch gerade nicht postuliert. Die Basis waren schlicht und einfach Diskussion auf der GCP-Website, insbesondere die von Nelson selbst formulierten Bedenken: „One of the possible sources of the apparent ‘global consciousness’ effect is an anomalous influence of the experimenter“¹. Diese Vermutungen lassen sich gerade da besonders gut testen, wo Missgriffe bei der Wahl der Zeitfenster andere Ergebnisse produzieren als bei einer korrekten Analyse. Neben der Annahme eines möglichen „globalen Bewusstseinsfeldes“ war dies die nächstliegende Erklärungsalternative für das Zustandekommen der hier postulierten Effekte, wenn sie wirklich auf anomalen Einwirkungen beruhen sollten.

Des weiteren moniert Bösch, dass ich „recht oberflächlich auf die Organisation und Gewinnung der GCP-Daten“ eingehe und meine Vorgehensweise „in vielerlei Hinsicht ebenso kursorisch (sei) wie das Vorgehen der Projektteilnehmer des GCPs“. Dieser Eindruck mag entstehen aus einer verengten wissenschaftlichen Arbeitsweise heraus, die sich auf das Detail konzentriert und dabei das Ganze aus dem Blickfeld verliert. Bösch mag es gewohnt sein, bei seiner wissenschaftlichen Arbeit sich ins Detail zu vertiefen, und wie er offen zugibt, wäre ihm eine Arbeit zum GCP, die nur ein Detail genau ins Visier nimmt, lieber gewesen. Aber bei einer solchen Arbeitsweise würde andererseits auch vieles ausgeblendet und nicht wahrgenommen. Meine Fragestellung war eine andere. Ich habe das GCP schon in einer frühen Phase kennen gelernt, habe selbst „Predictions“ dafür formuliert und habe ein EGG-Hosting bei der heimischen Technischen Universität Braunschweig vermittelt, das bis heute gepflegt wird. Mir ging es also um das Ganze. Was ist nach den gut vier bis fünf Jahren, seit dem das Projekt läuft, unter dem Strich dabei heraus gekommen? Kann ich den Betreuern des REGs an unserer Technischen Universität weiterhin die Pflege und Unterhaltung desselben empfehlen? Ist die Konzeption noch stimmig? Werden die zuvor festgelegten Kriterien klar definiert und eingehalten? Ein solcher Ansatz erfordert eine andere Vorgehensweise als die Beschäftigung mit dem Detail, denn das Ganze ist mehr als nur die Summe seiner Details. Von daher musste ich mich bei dieser Aufgabe von Anfang an der Herausforderung stellen, das Ganze in den Blick zu nehmen, auch auf die Gefahr hin, dem Detail

¹ <http://noosphere.princeton.edu/brahms.html>

nicht gerecht zu werden. Einem Wissenschaftler, der mit der Arbeit am Detail vertraut ist, mag diese Vorgehensweise suspekt erscheinen, aber umgekehrt wird es diesem nicht gelingen, das Wesen des Ganzen zu erfassen, wenn er sich nicht von der Fixierung auf das Detail löst. Was wie Oberflächlichkeit erscheinen mag, ist wissenschaftlich gesehen nicht negativ zu beurteilen, sondern zunächst einmal neutral. Ob es positiv beurteilt werden kann, hängt von dem Fortschritt in der Sache ab, den solch eine Vorgehensweise produziert.

Missverständlich finde ich auch Böschs Satz: „Was aber bei der Komplexität des Projekts, das im Wesentlichen von nur einer Person – Roger Nelson – betreut wird, nicht verwunderlich ist...“. Ursprünglich gab es einen ganzen Mitarbeiterstamm, der das Projekt entwickelte und mit trug. Nun wird die Betreuung des ganzen Projekts auf die Person Nelsons fokussiert, und damit wird freilich auch er für die Fehler und die Unzulänglichkeiten haftbar gemacht, die im Verlauf des Untersuchungszeitraums offenbar geworden sind. Haben sich die anderen Mitarbeiter im Laufe der Zeit schleichweise von dem Projekt distanziert?

Nelson schreibt bezüglich der in den „Predictions“ nicht aufgeführten und ausgewerteten Ereignisse: „First, he lists events that are in the results table but not in the registry. One of his identifications, „Astrological Moments, 20011209“, is in error. The event was in the registry.“ Da ist Nelson im Recht. Bei einer erneuten Durchsicherung der „Prediction Registry“ ließ sich dieses Ereignis im „Kleingedruckten“ ausmachen: „Dec 9, 3:15 PM EST. Moon occupies 14.1 Libra.“ Um Fehlerquellen zukünftig auszuschließen, würde ich mir allerdings eine klarere Zuordnung der Voraussagen in der „Prediction Registry“ zu ihren Auswertungen in der Ergebnisliste wünschen. Weiter schreibt Nelson: „For example, he says that the data are uploaded daily, which is not correct. For most eggs, the Internet connection is continuous, and the data are sent in 5-minute packets, resulting in a near real-time registration of the data.“ Tatsächlich aber hatte ich geschrieben: „Im Spätsommer 1997 waren Forscher in Europa und den USA technisch in der Lage, mit Hilfe des Internets Zufallsgeneratoren in verschiedenen Zeitzonen miteinander zu synchronisieren und deren Datenströme simultan oder in Intervallen einem Server zuzuführen, um sie dort zu speichern.“ In den Anmerkungen hatte ich ergänzt: „Einige Rechner sammeln zwar kontinuierlich Zufallsdaten, schicken diese aber als ganzes Paket einmal am Tag oder in noch längeren Intervallen an den GCP-Server.“

Nelson weist meine Vermutung, dass die REGs auf Störeinstrahlung mit einem Bias reagieren könnten, zurück: „He gives an incorrect description of the exclusive or (XOR) process that guarantees an unbiased mean output from the REG/RNG devices. He proposes that a physical source operating at a certain frequency could induce a bias. But his suggestion is not viable.“ Das reicht m.E. nicht. Im Zusammenhang mit Hagel und Tschapkes Eisenbahnversuch habe ich bei eigenen Computersimulationsversuchen gesehen, wie schnell und unerwartet sich Kreuzmodulationen einschleichen können, und daher würde ich eine experimentelle Überprüfung dieser Vermutung in Bezug auf die REGs des GCP empfehlen; oder es wäre hier zumindest angebracht gewesen, abgesehen von den notwendigen Kalibrierungen der REGs, auf solche Untersuchungen zu verweisen.²

² Nelsons Hinweis auf eine Studie von Radin (2002) hilft zur Zeit noch nicht weiter, da diese noch nicht veröffentlicht ist.

Meine Methode, für ein einzelnes Ereignis jeweils auch nur ein Ergebnis zuzulassen, wird von Nelson in Zweifel gezogen: „Etzold asserts that analysis of different predictions on the same day, for example using a meanshift and a variance measure, is not appropriate. I can see his point, but on the null hypothesis, mean and variance are independent parameters.“ Dass es sich um verschiedene Parameter handelt, ist für mich nicht das Problem. Mein Hauptkriterium war die Vergleichbarkeit. Unter diesem Gesichtspunkt wurden die Ergebnisdaten schon ab 1999 „normalisiert“, d.h. auf 600 df gesetzt. Wenn nun ein Ereignis mehrfach erscheint, weil einmal ein Mittelwert über einen längeren Zeitraum und dann ein Spitzenwert für einen kurzen Zeitraum berechnet werden, dann wird das Prinzip der Vergleichbarkeit wieder ausgehebelt, und ich frage mich, wieso zuerst alle Ereignisse auf 600 df normalisiert werden mussten?

Die Überlegungen und Analysen zu einem „Experimentatoren-Effekt“ in dem Sinne, das die hohen Signifikanzen durch unbewusste, anomale Beeinflussung der Versuchsdaten durch die Versuchsleiter und Projektteilnehmer zustande kommen, werden von Nelson dahingehend verstanden, dass ich sie dadurch erklären will. So weit wie es mir aufgrund der Datenbasis und meiner Möglichkeiten gelungen ist, konnte ein solcher Effekt zunächst definitiv ausgeschlossen werden. Meine Schlussfolgerung lautet: „Würde man den beobachteten anomalen Effekt so verstehen, dass die Erwartung der Versuchsleiter und der Beobachter des Experiments ihre entsprechenden Ergebnisse produziert, dann scheinen die hier vorgestellten Untersuchungsergebnisse gegen einen solchen direkten Effekt zu sprechen.“

Was ergibt sich aus dieser Diskussion nun für die Zukunft? Böschs Behauptung, dass es um den wissenschaftlichen Status des GCP schlechter gestellt sein mag als es meine Untersuchung vermuten lässt, ließe sich dadurch entkräften, dass die Analysen im GCP ausgehend von der Formulierung der „Predictions“ bis hin zur Ergebnisberechnung und Darstellung automatisiert werden. Als Vergleich steht mir das RetroPsychoKinesis-Project³ vor Augen, das von John Walker betreut wird. Hier geschieht die Datengewinnung sowie die tägliche Berechnung der Ergebnisse rein automatisch. Entsprechend ließe sich die Software des GCP auch um entsprechende Steuermodule erweitern, so dass nach dem Eintrag einer Voraussage in der „Prediction Registry“ die Berechnung des Ergebnisses automatisch vollzogen wird, wenn die Daten zur Verfügung stehen. Das würde nicht nur Nelson entlasten, der die Analysen manuell durchführt und auch die Ergebnisse noch manuell auf der Ergebnisseite eintragen muss, sondern auch die Zuverlässigkeit der Ergebnisdaten erhöhen. Manuelle Analysen blieben dann nur noch den explorativen und hypothesengenerierenden Verfahren vorbehalten.

Über die Aufnahme einer „Prediction“ in die „Prediction Registry“ sollte wenigstens ein Dreiergremium entscheiden, das per Abstimmung aus der Mitte der Projektteilnehmer jeweils zeitlich befristet gebildet wird.

Die wissenschaftliche Zuverlässigkeit der Ergebnisdaten scheint mir zur Zeit das wichtigste Ziel zu sein. Zahlreiche Projektteilnehmer, Interessierte und Institutionen sind am GCP beteiligt (und ich gehöre im Rahmen meiner Möglichkeiten mit dazu). Im Interesse derer, die beteiligt sind durch das Hosting eines REGs (EGG) oder durch die Formulierung von

³ <http://www.fourmilab.ch/rpkp/>

Vorhersagen, sollten die Ergebnisse des GCP gegenüber jedem Zweifel erhaben sein. Die jetzige Vorgehensweise, in der explorative und konfirmatorische Analysen nicht sauber voneinander zu trennen sind, zerstört das Vertrauen in die Ergebnisse. Und auch wenn die Effekte noch so ungewöhnlich oder unerwartet sind, kann das nicht die Verrechnung derselben in das Gesamtergebnis rechtfertigen, wenn es nicht vorher für diese eine klar formulierte Hypothese gegeben hat.⁴

Zur Datenaufbereitung: Es hat sich im Laufe der Zeit herausgestellt, dass die übertragenen Zufallsdaten höchst unterschiedliche Qualität aufweisen. Auf der GCP-Website sind die REGs aufgelistet mit den jeweiligen Zeiträumen, in denen sie unbrauchbare Daten produziert haben. Auch hier wäre eine automatisierte Datenkontrolle sinnvoll, die die hereinkommenden Daten auf Fehlerhaftigkeit scannt und ggf. aus dem Datenpool entfernt. Die ausgesonderten Daten sollten nicht gelöscht werden, sondern in einem Extra-Datenpool gespeichert werden, um für spätere Analysen zur Verfügung zu stehen.

Die Kontrolldaten sollten auf dieselbe Weise gewonnen werden wie die Ergebnisdaten, allerdings aus Datensegmenten von Zeitfenstern, für die keine Effekterwartung formuliert wurde. Nur so ist meines Erachtens eine „saubere Baseline“ für das GCP zu gewinnen bzw. können schnell Irregularitäten in der Datengenerierung, die es trotz der REG-Kalibrierungen noch gibt, erkannt werden. Auch die Kontrolldatengewinnung sollte automatisiert werden. Jedes andere Verfahren halte ich für zweifelhaft, insbesondere auch die Monte-Carlo-Simulationen, die nur dann eingesetzt werden sollten, wenn keinerlei Vergleichsdaten zur Verfügung stehen.

Das alles erfordert sehr viel neuen Programmieraufwand. Aber der Schaden, der entsteht, wenn das inzwischen nun schon fünf Jahre laufende Projekt und dessen Ergebnisse in nachfolgenden wissenschaftlichen Arbeiten, Studien und in Diskussionsgruppen zitiert wird, um die Aufsehen erregenden Erfolge und Fortschritte in der Parapsychologie zu demonstrieren, ist größer.

Ähnlich wie die experimentelle Laborforschung zur Psychokinese steht das GCP unter dem Verdacht, dass sich die hochsignifikanten Ergebnisse nicht nur einem möglichen anomalen Effekt verdanken sondern auch Fehlern in der Datengewinnung, der Analyse und der Datenauswertung. Um diesen Verdacht auszuschließen, sind neben einer noch weitgehenderen Automatisierung des Projekts Verfahren zu entwickeln, die es möglich werden lassen, quasi „Fehlertypen“ in der Datenauswertung zu erkennen und von möglichen „Signaturen“ anomaler Effekte zu unterscheiden. Die Untersuchung der Effektstärke könnte

⁴ Inakzeptabel ist es daher für mich, wenn Nelson ausführt: „In some cases (e.g., New Years and September 11 2001) the unexpected patterns in the data are so clearly correlated with the events, and are so obviously generalized across the network, that it stretches credulity to attempt an explanation such as is represented in Etzold's Table 3.“ – Es mögen die Muster und Effekte noch so sehr mit dem Ereignis korrelieren; wenn es dafür keine Hypothese oder keine Theorie gegeben hat, die ein solches Zusammentreffen vorher vorausgesagt hat, sind sie bedeutungslos und nichts weiter als Zufall. In diesem Fall liegt m.E. ein wirklich wissenschaftlicher Beleg erst dann vor, wenn bei einem zweiten ähnlichen Ereignis nach einer klar formulierten Voraussage der selbe Effekt sich in der selben Weise zeigt. Es wird aber etwas nicht dadurch wissenschaftlich, dass es einfach in der Grafik oder Tabelle nur schön aussieht.

in diese Richtung führen. Das Modell der Pragmatischen Information (MPI) von Walter von Lucadou (2002) macht ebenfalls Aussagen über die Entwicklung eines Effekts und seiner Veränderung. Im GCP gibt es einmalige Ereignisse, aber auch welche, die sich wiederholen, z.B. der Jahreswechsel. Nach dem MPI ist zu erwarten, dass der vorher beobachtete Effekt in diesem Fall entweder langsam verschwindet oder sich in sein Gegenteil verkehrt. Doch beim GCP finden wir in diesem Fall eine annähernd gleich bleibende Effektstärke. Die im GCP von den Erwartungen des MPI abweichende Effektentwicklung ist vor diesem Hintergrund m.E. erklärungsbedürftig.

Eine Zäsur ist m.E. nötig, in der das Konzept und die Methodik des GCP erneut zur Diskussion gestellt wird. Ähnlich wie bei der FAMMI-Meta-Analyse könnte eine solche Diskussion per E-Mail oder in den einschlägigen Mailinglisten geführt werden, mit dem Ziel, Kriterien und Ideen für die zukünftige Arbeit des GCP zu sammeln. Der jetzige Zustand des GCP lässt keine Zufriedenheit aufkommen, und es werden immer noch Hoffnungen und Erwartungen geweckt, die mit den Fakten nicht mehr im Einklang stehen. Aber um das zu ändern müsste bei allen Projektteilnehmern und -Begleitern erst einmal die Bereitschaft wachsen, wirklich zu verlässlichen Ergebnisdaten kommen zu wollen.

Literatur

- Heidegger, M. (1979): *Sein und Zeit*, 15. Aufl., Tübingen.
- Lucadou, W. v. (2002): „Mitlaufen“ des MPI ist zu begrüßen. *Zeitschrift für Anomalistik* 2, 82-84.
- Radin, D.; Nelson, R. (2002): Meta-analysis of mind-matter interaction experiments: 1959 to 2000. <http://www.boundaryinstitute.org/articles/rngma.pdf>.
- Steinkamp, F.; Boller, H.; Bösch, H. (2002): Experiments Examining the Possibility of Human Interacting with Random Number Generators: A Preliminary Meta-Analysis. Proceedings of the 45th Annual Convention of the Parapsychological Association. Paris, 5-8 Aug. 2002, 256-272.