Determining that the GCP is a goal-oriented effect: a short history

Intro

I've worked on the GCP for a long time and have concluded that it's a goal-oriented experimenter effect. Since there was some discussion (apparently) at the recent PA, I thought I'd write a summary to explain how I've come to my conclusions.

How the GCP works

In a nutshell, the GCP goes like this. A synchronized global network of RNGs runs 24/7 with each RNG producing a single data trial per second. The data trials are sums of 200 consecutive bits. Roger, along with an informal group of collaborators, occasionally identifies an "event" that's considered engaging enough and global enough to count as a candidate for a global consciousness (GC) effect. Roger then chooses start and end times for the event, writes these to a formal registry and downloads data for the event period. A standard test is made (basically, for an increase in the network variance) and its p-value is converted to a z-score. The cumulative significance for the on-going experiment is given as the Stouffer Z of the event z-scores. Today, with about 500 events since 1998, the Z is 7, which is huge. This result is taken as evidence for a global consciousness that manifests as deviations in the network when many people react collectively to an event.

Early objections

Early on, two objections were raised. One is that the freedom to choose events and their start and end times is so similar to a free-choice psi task that one can argue that the result is due to experimenter psi. Ed May and others have argued that psi data selection à la DAT provides a better explanation and that GC would only be convincing if the event periods were chosen algorithmically, without experimenter intervention. A second objection, from Jeff Scargle and others, is that since the data are xor'd it's impossible to imagine a mechanism that isn't goal-oriented (GO). If you think this through, it also says that the GCP looks like an experimenter effect.

Deciding between GC and GO

The question for the GCP, then, has long been: is it GC or GO? There are two ways to decide this. One is to provide better evidence for GC. The other is to demonstrate that it's really GO. I've spent more than a decade trying to resolve this and have concluded that the answer is it's (clearly) GO. Most of that time I tried to find evidence for GC. When that eventually proved fruitless, I turned to investigating GO. Tests for GO and direct evidence for psi selection strongly point to GO. My conclusion is that, while there may be something like GC somewhere, there is no evidence for it in GCP the data. But there is good evidence for GO.

Searching for GC

There are three ways to get further support for the GC hypothesis.

One is to implement algorithmic event selection. No one ever did this. If it produced a positive result, it would be a compelling argument for GC. A negative result wouldn't tell us much since one could always move the goalposts and claim that the algorithm wasn't precise enough to identify GC events.

A related approach is to identify untested categories of events and go back in the data to see if there's an effect for these. This would also be compelling if it worked, but as with the algorithmic approach, a negative result could be discounted as having chosen events that aren't sufficiently GC-like. I've looked at about 8 surrogate event sets and all come up null. (some are: a decade's worth of large land-based earthquakes, full moons, registered air plane crashes; 800 very large rock concerts; 164 world cup games; Sunday prayers in observant christian countries; Friday prayers in observant muslim countries).

A third approach, which I spent a long time doing, is to look for data structure in and around events that you wouldn't expect for GO, and which looks like a GC fingerprint. Eschewing details, the standard GCP test finds correlations in the sign of outputs among pairs of RNGs. Working with pair correlations allows us to cut up the data in different ways to look for structure. An example is to see if the correlations decrease as the separation between RNGs gets larger. A distance effect like this could be due to some field-like GC, but you wouldn't expect it for GO. There are other data-minings that can be done to look for structure and Roger and I published a number of them back in 2011. Data-mining is ok, but you have to re-test them on new data in order to draw conclusions. We cautioned about that in the paper. I re-checked those analyses in 2015, which were first done around 2009, by including 6 years of new data. I found that the structure regresses towards the mean. None of it reproduces.

In a poster at the recent PA, Roger says the various data structure can't be explained by GO. That's not right. The data-mined structure has either regressed, or Roger draws a false conclusion. Here's an example of the latter: There is a second, orthogonal correlation that happens to deviate positively, just like the standard correlation (it's basically the 2nd moment of the standard correlation). The claim is that this wouldn't happen in a GO picture, but that's not true. The standard test includes 2nd moment terms too, so GO on the standard test says you *should* see those correlations as well. [for the cognoscenti, the standard test has terms like <z1*z2>, but also z1*z1, which gives the second moment].

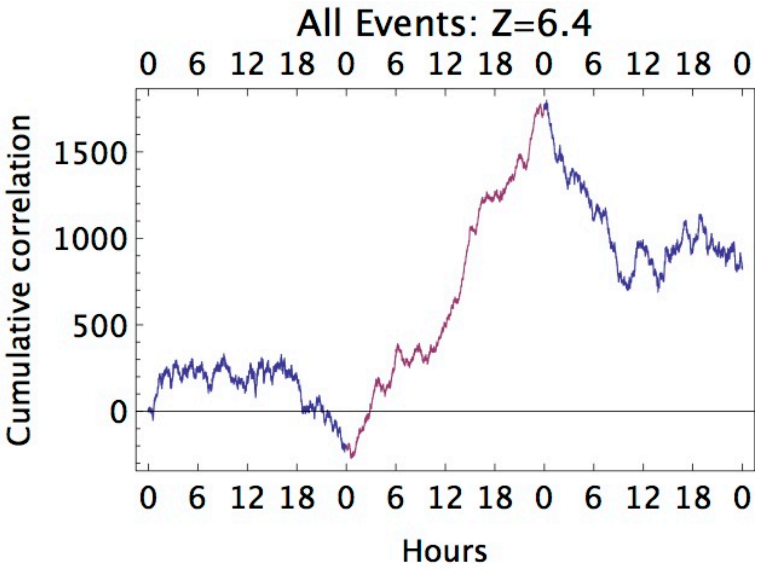
In his PA poster, Roger showed some other data-mining results to further the case for GC. One is the Burning Man data from 1999-2006. This has a large deviation, but Roger doesn't use the standard test (he chooses a somewhat more involved averaging across those years). The standard test gives a null result, and the choice of a different statistic means a re-test is needed. The re-test on the 2007-2014 data comes up null. Another case that gets shown as evidence for a GC effect is the large deviation for a long period after the 9/11 event. While we can't re-test 9/11, we can look at other big attacks like the London and Madrid bombings and others. None of those show a long term deviation, so just choosing to show one event (9/11) out of many doesn't have much evidentiary value. And if you get into post-hoc arguments about how seemingly equivalent events really shouldn't be treated equivalently, then the reasoning really doesn't have any weight.

I could go into more detail, but I think my point is clear.

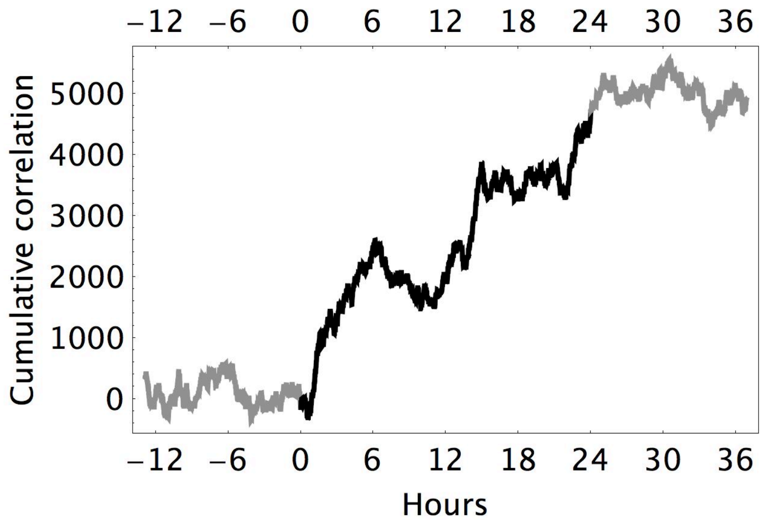
If this all sounds messy, that's because it is. It's easy to fool yourself with data-mining if you don't re-test. We used a lot of weasel words in our publication to ward against over-interpreting, but I fear that message got lost.

Searching for GO

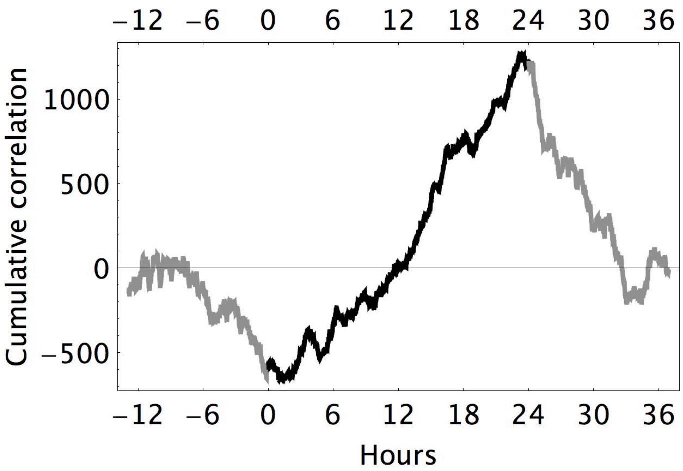
I subsequently spent a few years fleshing out evidence for GO. That has worked out nicely. The smoking gun can be seen in this picture. It shows the cumulative of all events (I stretch them out over 24 hours so all events have the same duration – easier to visualize it that way...). The plot also shows data 12 hours before and 12 hours after the events. You can see that there's a strong negative effect in the before/after data. This is exactly what you would expect if you allow freedom of choice in the start/end periods of events: psi-choosing a start time to give a little more positive effect means that the earlier nearby data will, on average, be more negative. I first saw this in a data-mining I did in 2006. I was expecting *more positive* effects near the events since that's what you'd guess assuming GC. I didn't go back to it until much later when the data structure approach was starting to look squirrelly. Re-testing on new data found that it reproduced. This was the one data-mine that held up!



In the plot, the red trace is the GCP event data and the blue traces are before/after data. There are actually two things here that support GO data selection. One is the negative-going effect both before and after. The other is that it aligns so closely to the actual start and end times that were chosen for the event predictions. In my mind, this is so striking as to be almost pornographic (for scientists). You can see the mechanism at a glance, but there's a data split that's even more telling. A large subset of events are fixed at a 24 hour GMT day. These events only have freedom to accept the event, but not to choose start/end times. That subset should *not* have proximate negative deviations because there's no psychic probing of the nearby data. That's indeed what is found as this plot shows.



All the remaining events, those that *do* have freedom to choose the start/end times, are shown in the last figure, below. Here we see that the negative deviations fully cancel the selected positive effect during events. This balancing is what you'd expect for GO selection (there could also be selection models were the balancing isn't perfect). Together, the two figures can explain why the balancing is less complete in the first plot, which takes all the events together.



A principled approach

I recently wrote a paper that looks at GO for the GCP data from a different angle. Some have found the paper a bit dense, so I'd like to make a couple of comments.

The paper does two things. One is that it shows that a simple field-like GC can't work because of the RNG xor's and network synchronization problems (see the paper for what I mean by "simple field-like"). The analysis is a bit involved because you need to get into some technical details. Nevertheless, some have suggested that addressing a field model like this is a strawman argument (presumably because we know psi doesn't work in such a simple way), or that I'm assuming this is how GC should work. This completely misunderstands my motivation for the analysis.

I don't believe psi would behave as a simple field, but as GC is a novel proposition, it's important to resolve possible explanations where we can, whatever one's prior might be. The analysis provides an answer to someone asking: How do you *know* a simple field model won't work? At least it's useful to have a definitive answer to that question in hand.

The second, and more substantial part of the paper deals with setting up a definition for GO and using it to make general tests for GO, independent of any mechanism. I like this approach because it states GO as a sort of principle that can be tested. The tests I do all favor GO. One may quibble with some of the details, but the tests are self-consistent and all provide signatures of GO. Taken together with the negative before/after effect, which GC can't explain, the conclusion of GO seems to me unavoidable (remember, the proximate negative effect isn't always there; it only appears when there's freedom to choose event times). If one still thinks that some GC is *also* operating and contributing to the effect, some evidence will be required to make that claim. Right now, there isn't any.

Some have felt that my definition of GO as self-referential fine-tuning is needlessly jargony. I disagree. Fine-tuning is a powerful principle that has appeared in various guises throughout the history of science. I think it's useful because it connects with sophisticated usage in methods of causal determination, and it has been applied fruitfully in many fields. Closer to home, the fine-tuning principle is lurking in H. Schmidt's equivalence principle for RNG experiments. It also finds echo in Occam's razor, Leibnitz's principle of indiscernibles and Einstein's strong equivalence principle. Considering fine-tuning allows us to think differently about GO and the tests in the paper come directly from that.

I've thoroughly enjoyed working on the GCP and my collaborations with Roger. I would have loved to find hard evidence for GC, but that hasn't turned out to be the case. There are many data-minings I've left unpublished and after looking high and low, I'm convinced that the only 'there' there is GO and not the GC many of us were hoping for. What would be useful, I think, would be to draw lessons from the project and think hard about how GC might be detectable, if it in fact exists.