



Date:	Sun, 22 Apr 2007 07:51:31 -0700
From:	"Julio Siqueira" <juliocbdsiqueira@YAHOO.COM>  Add to Address Book  Add Mobile Alert
Subject:	Re: Update on Dean Radin - New DATA !
To:	AVOID-L@HAWAII.EDU

Hi Bill,

My comments below [\[Julio\]](#)

Jefferys wrote:

Date: Wed, 18 Apr 2007 09:32:48 -0400

From: "Bill Jefferys" <bill@ASTRO.AS.UTEXAS.EDU>

Subject: Re: Update on Dean Radin - New DATA !

To: AVOID-L@HAWAII.EDU

At 5:46 AM -0700 4/18/07, Julio Siqueira wrote:

>My Conclusion so far:

>

>It seems that the main critics of Radin on Avoid-L have not found

>any technical problem with his statistic calculations regarding the

>Radin vs Good (Nature review) issue.

I thank you for posting the book. BUT...

[Julio] You are welcome to it (you, Brent, Vic, and all the list). Actually, since I came here kind of "supporting" Radin, it is part of my responsibility to try to help in tracking down potential problems related to him. Unfortunately some things lie beyond my depth (and that includes making the statistic calculations from the book to check the mathematical precision of Radin's numbers).

You apparently did not transmit to Radin the information about the filedrawer problem discussed in Jeff Scargle's paper.

[Julio] I did not. But I do know Radin is well aware of this issue. If he decided to leave it aside, you can count that as further problems against Radin... Radin has one paper, if my memory serves me well it is about retro-PK, where he says that he used Scargle's method and still found the results to be statistically significant. But I do not know if Radin only uses it when he... pleases. Scargle made three reviews on www.amazon.com about Radin's book... The man is a skilled hunter!

Scargle shows

that the size of the filedrawer required to neutralize results is much smaller than the usual calculation gives, and in fact can be quite small indeed. Thus, this part of my question has not been dealt with.

Nor does Radin deal with the fact that observed p-values cannot be interpreted as probabilities in any meaningful sense, as Berger and Delampady, and Berger and Sellke have pointed out.

[Julio] Not in any meaningful sense? Can't p values be interpreted as the probability that (assuming that the null hypothesis is true) you would come to the very same result by chance alone? That seems very "meaningful" to me... For example, if I am about to throw a fair die (dice) and someone says to me "Hey, you will get only the number 6 facing up in the first 1,000 throws, and then only the number 2 in the following 1,000 throws" and it happens exactly like this someone says, then there is a p value associated to it. Isn't this p value meaningful in some way?

A p-value is not the Type I error rate for the investigation, for example, though it is often claimed to be such (the Type I error rate, conditioned on the p-value, is generally larger than the p-value, in a conditional frequentist calculation); it is not probability of the null hypothesis being false (that requires a Bayesian approach); nor is it the probability that the "results were obtained by chance" (that is always 100%, by definition, since in any stochastic process the results, whatever they are, are always obtained by chance).

[Julio] You mean that the probability that the results were obtained by chance is always 100%? So, chance is all that there is? No causation, no correlation, nothing?

Indeed,

p-values don't have a reasonable probabilistic interpretation in terms of the actual experiment that has been performed.

Did Radin perform the calculation I suggested, that is, randomly dividing up the experiments into, say 10 groups of 18 or 19, calculating the z-scores, and seeing if they are giving the same basic result, or are (as I suspect will turn out to be the case) all over the map? This is the acid test to see if what he is doing is actually pulling a real signal out of noise, or (as I suspect) pulling noise out of noise.

[Julio] I will forward that. But I am sure Brent Meeker himself will do these calculations and come to us with the results...

:-).

Did Radin deal with the problem of quality control? That is, we know that Rhine's early experiments were poorly controlled, and that the size of the effect declined with improving controls. How do we know that all of the data going into his calculation were so well controlled that biases were so small as to be ignorable? The point is, unless the biases can be demonstrated to be much smaller than any effect being reported, his calculation amounts to GIGO.

[Julio] I reproduce here what I told to Meeker:

Brent said (Mar 31, 2007): "You should also remember that the question is not simply one of doing the statistics correctly. There were many procedural problems with Rhine's experiments and as he responded to criticisms and tightened controls the effect steadily declined. One wonders how Radin factored this into his odds."

[Julio] The question was never anything more than "doing the statistics correctly," Brent. This (i.e. the question being more than doing the statistics correctly), sadly enough, is a red herring that both you and Jefferys have been trying to sell to this list (I hope that the two of you are doing this unconsciously...). The question has always been only one, and a simple one: Radin's assertion in his book, Good's challenge, and the final answer from Radin. The two parties (Radin and Good) seem to have acted in good faith. The same cannot be said of you or of Jefferys (though you two may be acting due to unconscious emotional strong bias).

[Julio] What Radin said in his book was: (pg 97 the conscious universe 1996) "If we consider all the ESP card tests conducted from 1882 to 1939, reported in 186 publications by dozens of investigators around the world, the combined results of this four-million trial database translate into tremendous odds against chance—more than a billion trillion to one. If we assume that there is a selective-reporting problem in this database and calculate the number of unpublished, unsuccessful studies required to nullify these astronomical odds, we find that the file drawer would need to contain more than 626,000 reports. That's more than 3,300 unpublished, unsuccessful reports for each published report. This again demonstrates that chance results and selective reporting cannot reasonably explain these results."

[Julio] Radin is only talking in this paragraph about the problem of selective reporting. Therefore, the p value 10^{-2000} is associated ONLY with this issue of selective reporting. And it was to that that Good replied, claiming that the p value he came to was about 10^{-24} or something like it. Good did that by backward reasoning (calculating), changing the odds into p value (something Jefferys stressed that cannot be done, but that he later acknowledged that is one of the areas where Good himself has some works on). To say that "Radin first threw one number and then threw

another" is unacceptable (it was not you who said it. You said worse things though...). It is easy to track down Radin's original calculation by using what he wrote. Not the informal "more than a billion trillion to one", but the more precise "That's more than 3,300 unpublished, unsuccessful reports for each published report." It was this number (3,300) that Good tried to demolish. And it was this number (3,000) that Radin further supported in his reply to Good.

[Julio] Now, we all know that both you and Jefferys have all that it takes to check very carefully (and perhaps very easily and fast...) all this issue. You have the expertise, the tables, etc. Now, Brent, it is your turn...

Are we absolutely certain that none of the data in his calculation involved cheating? (I am not accusing Radin of cheating, but am deeply concerned about the provenance of the historical data he used.)

[Julio] Cheating is, IMO, possible. Even Susan Blackmore was caught cheating by Rick Berger. (shame on her...).

Even a small amount of cheating by the original investigators (now all dead, I presume)

[Julio] Some are already reincarnated, I presume... :-)

would render his calculation meaningless. We know that some highly touted experiments in parapsychology involved cheating (e.g., Soal and Bateman, as exposed by Betty Markwick). How do we verify, to everyone's satisfaction, that no cheating was involved in these data?

As I pointed out, I do not believe anyone who claims to pull results out of data at levels much smaller than 0.01 times the standard deviation of the elemental data, or possibly 0.001 if extremely carefully done (which I doubt in general for parapsychology experiments). In the case of binary choices, whether at equal or unequal probabilities, the standard deviation of a single experiment is of order $1/2$, so I would be very suspicious of any reported effect size smaller than about 0.005 and would not believe at all any effect size smaller than 0.0005 (on a scale of 0 to 1). I would attribute such results to experimental defects, and not to psychic effects.

Why don't parapsychologists devise experiments that pass the interocular traumatic test?

[Julio] It seems that their experiments withstand well Bayesian analysis (Ganzfeld at least).

Let's face it, no one who is not already
convinced will be convinced by experiments of the kind that Radin
uses in his
calculation.

**[Julio] I think many people get convinced. But the important thing is: is it all real? (i.e. psi). I think it is.
But... I do not know.**

**Best Wishes,
Julio**
