Journal of Nonlocality Round Table Series
Colloquium #3, December 2013

Tinkering with the Unbearable Lightness of Being:
Meditation, Mind-Body Medicine and Placebo in the Quantum Biology Age

Appendix #2

DMILS SECURELY ESTABLISHED?
(COMMENT ON THIRD JNL ROUND TABLE)

Brian Millar

1. INTRODUCTION

It seems to be a rather common opinion for those studying DMILS (and related biology) that "*there is so much convincing experimental material relating to DMILS that no reasonable person can currently doubt the reality of such effects*."

This seems to me contentious and deserving of further discussion. I offer myself as Aunt Sally: while quite a reasonable person (at least in my own opinion) I nonetheless consider the reality of DMILS as considerably less than established. The following explores why.

2. REPEATABILITY IN MAINSTREAM AND PARAPSYCHOLOGY

2.1 Experimenter Effect in Psi Experiments

Parapsychological experiments in general are not repeatable at will by any experimenter. The motto of the Royal Society is "Nullius in Verba" (Don't believe it just because that's what you are told). I personally spent about a decade (roughly 1975-1985) full-time trying to verify many of the (then) big-time parapsychological claims and failed to find convincing evidence that psi is real.

Nor am I the only serious experimenter with this experience: British workers in particular have been unable to find much evidence of psi. The prototype is perhaps Donald West who indefatigably carried out experiment after experiment, year after year (until he found a co-experimenter (Fisk) who apparently made the magic work). A more recent example, almost forgotten today, is Susan Blackmore who did not encounter a "Fisk", and finally exited the field with an impressive collection of negative results and a severe case of cognitive dissonance. Wiseman, a graduate of the Morris school in Edinburgh, has likewise reported rather consistent negative findings and has subsequently become the most prominent current “card carrying” English skeptic.

From my own “bedrock” of practical experience I know with some confidence that those who claim cross-experimenter repeatability for psi experiments are “talking through their hats”. The American continent may be famed for two reasons: the genesis of the modern horse (although that has recently been cast into doubt) and impressive evidence for psi.

This may be an international example of “experimenter effect”. Even so, it means that experimenters like myself are (permanently) denied first hand experience of psi in their own experiments. The principle of scientific objectivity “Nullius in Verba” cannot be applied. The only alternative left open is to examine the material of other experimenters. This second-hand approach is grossly inferior since an outsider has little access to the myriad details at the fingertips of the actual experimenter. Instead a subjective judgement must be made of the likelihood of all kinds of error for each experiment examined.

2.2 Adequacy of Mainstream Psychological and Medical Research Methodology

Psychology is currently suffering from an acute crisis of confidence (Paschler and Wagenmakers, 2012) in the ability of its standard methods to produce correct results. The problem seems to be not with the methodology as such but with the fact that experiments are carried out by real men and women, often with strong expectation (or hope) of a particular result. Questionable Research Practices (QRP) are rife: indeed the majority of experimental psychologists openly admit to practicing QRP (Leslie, Loewenstein, Prelec, forthcoming). Some questionable practices even seem to be the prevailing research norm. An important QRP is undeclared multiple analyses, in which the researcher, in effect, gives himself several chances of success. There are few experiments indeed with only a single test hypothesis. Unless corrected (by well-known methods) this plays merry hell with resulting statistical estimates. While QRP is the preferred explanation for lack of repeatability in mainstream studies other, more recondite possibilities do exist, including even psi influence (if real) on the results obtained.

Ioannidis (2005b) constructed a simplified mathematical model of QRP and on that basis concluded that “most published research findings are false”. Just what fraction of published research reports in practice yield false positives is currently unknown. Pashler and Wagenmakers (2012) relate that “In 2012, several large pharmaceutical companies revealed that their efforts to replicate exciting preclinical findings from published academic studies in cancer biology were only rarely verifying the original results.” A host of biomedical research seems to fall into the same poorly reproducible boat. Ioanidis (2005a) looked at more than 30 most cited clinical trials and of these only a little over half (59%) replicated satisfactorily. In another study he looked at over 400 claims of genetic studies reporting differences in disease risk between men and women, and found that only a single ONE replicated. On purely methodological grounds this was a particularly poor collection even before looking at replication; but the studies did find their way into the literature. A quarter percent replication is abysmal by any standard. If this proves typical it may be that standard psychological methodology in daily practice is no better than throwing dice (and loaded dice at that).

An optimistic guess is that perhaps only half of studies judged methodologically adequate report replicable results. The best hope of establishing realistic estimates for reproducibility in mainstream psychology are projects for systematic replication (e.g. the Reproducibility Project - http://openscienceframework.org/project/EZcUj/). In parapsychology a partial solution is provided by pre-registration of the details of a forthcoming experiment with some stress on the number and details of analyses to be carried out (most notably the Koestler Registry, in Edinburgh). And in the area of clinical trials there exists a substantial organization (http://clinicaltrials.gov). I strongly urge DMILS workers (as well as parapsychologists) to make use of such registration facilities.

The results of meta-analyses are often taken as strong evidence for the reality of psi, but it may be that what these measure is, in fact, a control condition for the standard psychological methods when carried out by real men and women, but in the absence of any real psi effect.

3. ADEQUACY OF DMILS METHODOLOGY

If even the best evidence for psi cannot be taken uncritically at face value, then how does it stand with DMILS? In this writer's opinion this is likely on average considerably worse than even in parapsychology, and that for three (overlapping) reasons:

a) DMILS experiments are typically more complex and (at least a few critical) operations must be carried out by hand. DMILS experimenters do not have the luxury of some PSI experimenters, where (almost) everything is carried out automatically by computer.

b) A typical DMILS experiment is often understandably of an exploratory nature. Not just a handful of variables but possibly hundreds are trawled in a single experiment. Some attempt may be made at an omnibus measure of effect (often ANOVA) but this all too often gives the appearance of being an unplanned pro-forma afterthought for the sake of publication, rather than a genuinely self-critical examination.

c) DMILS researchers are frequently physicians at heart. They exist to heal people and if a good clinician has reason to believe that dancing naked around a bonfire at midnight at full moon may benefit the patient, then it's "off with the lab-coat" ! As (potential) patients we must be grateful that physicians are not just scientists. The clinician is involved while the pure scientist is dispassionate. This (McCoy/Spock) difference results in a distinctive "Research Culture" for DMILS. Should an oncologist find some sub-group in an experiment to be completely tumour-free then it may quite miss his mind to mention that this was not a planned analysis. Not all medical researchers can instantly "switch hats" between physician and scientist. Medical researchers tend to do their DMILS experiments in just the same way they do any other kind of medical research ("the RIGHT way, the way my daddy and granddaddy did it"). I hasten to add that this is only a general tendency and that there are those (particularly young workers) who are currently labouring to introduce more blindness and controls.

The old “Gold Standards” of psychology, inherited whole by biomedical research have, in practice, been found wanting and the world awaits higher standards of implementation. Parapsychologists (a few anyway), under the goad of extensive criticism, have travelled this road somewhat earlier than the mainstream and have adopted higher standards than are usual in psychology.

DMILS researchers largely belong to a distinctive Research Culture, which is closely modelled on mainstream medical research. Compared to parapsychology this methodology is often inferior, particularly in the sense that it is less self-critical. And the critical parapsychologist may shake his head at much DMILS. In the modern world the bottom line is that the standards required to establish an extraordinary result are higher than needed for ordinary work. DMILS workers NEED higher standards than mainstream medical work.

It is decidedly not the purpose of this note to declaim with Jeremiah “Woe to DMILS !” I suggest rather that DMILS workers may be able take advantage of methodological lessons that parapsychology has already painstakingly learned. I take the opportunity here to make a single concrete suggestion.

It has been noted in 3b above how many DMILS experiments use a wide exploratory net, with many variables, in order to catch any effect which might occur: afterwards, a half-hearted attempt is made to justify this, as a whole, with some omnibus test. The “Pilot/Confirmation” design provides a much more satisfactory and greatly more sensitive solution here. An initial pilot experiment allows identification of the relevant effects and variables. There is no attempt to claim (post hoc) significances for the pilot: it is used solely to make detailed predictions for the confirmation part and the significances calculated from this can then be relied upon.

A practical objection to the pilot/confirmation design is that it is cumbersome; furthermore, factors of importance can change in the interval between the two phases. It is possible under some circumstances to combine both pilot and confirmation into a single unit: the “pilot” might be designated (ahead of time) as EVEN and the “confirmation” as ODD trials, or some randomized order can be used. The important limitation is that whoever analyses the “pilot” and makes detailed predictions for the “confirmation” must be (at that moment) blind to the confirmation data. Together with colleagues, I long ago introduced this method in Beloff's lab, under the title of “Edinburgh Split” design.

4. CONCLUDING REMARKS

Parapsychology is essentially moribund as a university study: it is virtually impossible to earn a living with it. One of its strengths was an attempt (by some) at “methodological purity”; a weakness was that little attention was paid to neural and molecular underpinnings. This was likely (at least in part) a consequence of the basically Idealist (Dualist) philosophy early introduced by Rhine.

DMILS experimenters are primarily interested in mechanism – but, conversely, display some lack of methodological rigour.

When I was active in parapsychology, some 30 years ago, the claim of “irrefutable evidence” was also in the air; but with the march of time this has eroded. I suspect that such talk functions as a psychological defense mechanism: as Shakespeare put it “The lady doth protest too much !” It is devoutly to be wished that DMILS experimenters can leave this kind of bombast behind faster than parapsychology, and concentrate instead on accumulating high quality evidence both for and against preconceived notions.

REFERENCES

Ioannidis, J.P.A. (2005a) Contradicted and Initially Stronger Effects in Highly Cited Clinical Research, JAMA 2005;294(2):218-228, http://jama.jamanetwork.com/article.aspx?articleid=201218

Ioannidis, J.P.A. (2005b) Why most published Research Findings are False, PloS Med. 2005 August; 2(8): e124

Leslie, K, Loewenstein, J.G, Prelec, D. (forthcoming, Psychological Science), Measuring the Prevalence of Questionable Research Practices with Incentives for Truth-telling, http://ssrn.com/abstract=1996631

Pashler, H and Wagenmakers E-J (2012) Editors' Introduction to the Special Section on Replicability in Psychological Science: A Crisis of Confidence?, Perspectives on Psychological Science 2012 7:528, http://pps.sagepub.com/content/7/6/528