Analysis details for an Online Comment Regarding the Claim that Poverty Impedes Cognitive Function

Author: Gregory Francis*[1]

Affiliation:

[1]Department of Psychological Sciences, Purdue University, USA.

*Correspondence to: gfrancis@purdue.edu

This document is part of a set that describes the Test for Excess Significance (Ioannidis & Trikalinos, 2007; Francis, 2012), as applied to the experimental findings in Mani et al. (2013). The Excel file *TESAnalysis.xls* summarizes the probability calculations for each study. It also includes calculations to estimate the means and standard deviations of the scores in Experiments 1, 3 and 4, which were derived from data plots in Mani et al.

The key calculation for each experiment is to determine the probability that an experiment with that design and sample size would produce the specified pattern for success. This is similar to post hoc power calculations, but it considers both significant and non-significant results that were deemed necessary by the original authors as relevant to the theoretical claims. As part of the probability calculations, the analysis assumes that the tests were run properly, were appropriate for the data (e.g., normal population distributions, homogeneity of variance), and were related to the theoretical claims in the way specified by the Mani et al. (2013). If these assumptions are not valid, then readers should be skeptical about the claims in Mani et al. (2013) for reasons other than those derived from this analysis.

Experiments 1, 3, and 4 had nearly identical designs but differed slightly in incentives and in task order. In each experiment participants were placed in either a “rich” or “poor” group and then assigned to either an “easy” or “hard” task that involved making financial judgments. The dependent variables were then measurements of cognitive performance using Raven’s Progressive Matrices and a cognitive control task. A successful outcome for these experiments involved six different comparisons for each dependent variable:

1. A significant interaction between rich/poor and easy/hard. This was measured with an ANOVA.
2. A non-significant effect for rich versus poor in the easy task. This was measured with a t test.
3. A significant effect for rich versus poor in the hard task. This was measured with a t test.
4. A non-significant effect between the easy and hard tasks for the rich. This was measured with a t test.
5. A significant effect between the easy and hard tasks for the poor. This was measured with a t test.

6. A significant difference between the rich and poor across both tasks. This was measured with a contrast derived from the ANOVA.

Mani et al. (2013) reported that both the Raven's and cognitive control measures satisfied the specified pattern for each experiment. Since these two measures are derived from a common sample, they are not independent measures. The probability that both measures satisfy the pattern partly depends on the correlation between these variables. If they are highly correlated, then the probability for one measure to produce the pattern is essentially the same as the probability for both measures to produce the pattern. On the other hand, if the measures are uncorrelated, then the probability for both measures to produce the pattern is the product of their individual probabilities. Mani et al. did not report the correlation between these measures, but an upper estimate of the joint probability is the smaller probability of either measure by itself. For example, if the probability is estimated to be 0.53 for Raven's and 0.65 for cognitive control, then an upper limit on their joint probability is the smaller value, 0.53. An assumption of independence would suggest that the joint probability would be their product, 0.34. In the analysis here, we always used the upper limit, which almost surely overestimates the true joint probability.

The sample sizes across the different conditions in each experiment were derived from the degrees of freedom reported in the text. The means and standard deviations for each condition were derived from the plots in Figures 1, 2 and 3. These values were used to generate 100,000 simulated experiments that sampled values from normally distributed populations having the means and standard deviations identified by the samples in Mani et al. (2013). Each simulated experiment was then analyzed in the same way as Mani et al. and checked whether it successfully matched the necessary pattern of significant and non-significant results described above. As in Mani et al. (2013), an experiment was unsuccessful if it did not match the necessary pattern. Source code (R Development Core Team, 2013) for the simulated experiments is available in the *Simulations* directory.

Table 1 shows the proportion of simulated experiments that reported the necessary significant or non-significant result for each test. The bottom row reports the proportion of simulated experiments that perfectly matched the required pattern. This value is not the product of the other probabilities because the tests are related to each other. Table 2 summarizes the overall probability for each experiment.

Experiment 2 had a similar design as experiments 1, 3, and 4, but the task was modified in a way that was predicted to produce specific non-significant results. Mani et al. (2013) do not describe the means and standard deviations of the data, so the best estimate of the probability of a non-significant set of results comes from the F value reported in the supplemental material that contrasts the rich and poor participants overall. This statistic is given in Table 2 and converted to a Hedge's g effect size that then allows for computation of *post hoc* power (Champley, 2009). The reported probability in Table 2 is the probability of not rejecting the null hypothesis.

Table 1. Proportion of simulated experiments that match the necessary pattern of results. R-rich, P-poor, E-easy, H-hard. Tests described in italics were anticipated null results.

Test	**Exp. 1 (Ravens)**	**Exp. 3 (cognitive control)**	**Exp. 4 (Ravens)**
Interaction	.609	.533	.529
E (R v. P)	.948	.855	.930
H (R v. P)	.890	.969	.887
R (E v. H)	.907	.798	.950
P (E v. H)	.715	.985	.819
R v. P overall	.666	.917	.743
Full pattern	.323	.448	.369

Table 2: Statistical properties of the Mani et al. (2013) experiments.

	n	**Test Statistic**	**p**	**Effect size**	**Probability of Success**
Exp. 1	24, 27, 25, 25	Multiple tests	--	--	.323
Exp. 2	10, 10, 10, 9	1.69	.20	0.408	.763
Exp. 3	29, 24, 23, 24	Multiple tests	--	--	.448
Exp. 4	23, 24, 23, 26	Multiple tests	--	--	.369
Field 2	188	1.715	.088	0.125	.523
Overall					.021

The description of the first field experiment in Mani et al. (2013) does not provide enough information to estimate the probability of the necessary pattern of significance and non-significance. Thus, these findings do not contribute to the current analysis. This approach is conservative because it effectively assigns a probability of 1.0 to these results. In reality, the probability must be less than one.

The second field experiment described a pilot study that measured stress-related markers in a population of farmers pre- and post-harvest. It was important for the theoretical claims in Mani et al. (2013) to establish that stress was different across these measurement conditions. There were two measures of stress that were reported as significant, but an upper limit on the probability of both measures can be estimated with the probability of the heart rate measure. Mani et al. (2013) claimed a significant difference based on P=0.088. This value is above the typical 0.05 criterion, but sometimes researchers use a 0.1 criterion for indicating significance, and this value was used for the probability analysis.

Mani et al. (2013) reported that each of the studies produced a pattern of results that supported their theoretical stance. If the theory is correct, and the population effects are as estimated by the samples, then the probability of getting studies like these to all produce the desired pattern is the product of the probabilities in Table 2, .021. This value is so low that readers should be skeptical about the reported experimental results as they relate to the theory.

References

1. Ioannidis, J. P. A. & Trikalinos, T. A. (2007). An exploratory test for an excess of significant findings. *Clinical Trials 4*, 245-253.

2. Francis, G. (2012). Too good to be true: Publication bias in two prominent studies from experimental psychology. *Psychonomic Bulletin & Review 19*, 151-156.

3. Mani, A., Mullainathan, S., Shafir, E., Zhao, J. (2013). Poverty impedes cognitive function. *Science 341*, 976-980.

4. R Development Core Team. (2013) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

5. Champely, S. (2009). pwr: Basic functions for power analysis. *R package version 1.1.1*. http://CRAN.R-project.org/package=pwr.

05 October 2013