



COMMENTARY

# Response to Bierman and Spottiswoode (2026)

## Patrizio Tressoldi

Science of Consciousness Research Group, Studium Patavinum, Università di Padova, Padova, Italy  
patrizio.tressoldi@unipd.it

## Lance Storm

School of Psychology, Adelaide University, Adelaide, Australia  
lance.storm@adelaide.edu.au

## ABSTRACT

This is a reply to Bierman and Spottiswoode's (B&S) assertion that the between-studies "decline effect" (DE) in psi research is present in all experimental protocols related to anomalous cognition (AC; a.k.a. extra-sensory perception; ESP) and psychokinesis (PK). While B&S utilized non-parametric Spearman  $\rho$  correlations to identify declines across AC and PK databases, we defend the robustness of our original meta-regression approach. Furthermore, we reanalyzed our original results, and the B&S results, taking into account the moderator effect of the number of trials or participants for the presentiment data on the effect sizes. The results were not very different from our original findings, showing that only two of six AC protocols indicated statistically significant declines, while all four PK protocols exhibited statistically significant declines. However, these databases are largely obsolete, lacking data from the last 20 years, and two of them involve very limited time spans. Our conclusion is that B&S's interpretation overstates the observed findings, thus reaffirming our original stance: "there is no evidence of a general DE across different experimental protocols in psi research".

## KEYWORDS

Decline effect, extra-sensory anomalous cognition, psychokinesis

<https://doi.org/10.31275/20264041>

## GOLD OPEN ACCESS



Creative Commons License 4.0.  
CC-BY-NC. Attribution required.  
No commercial use.

Bierman and Spottiswoode's (2026) (B&S) paper presents new empirical evidence related to between-studies decline effects (DEs) of psi phenomena across replicated experiments. Their aim was to confirm the reality of DEs and identify their characteristics.

It would appear that B&S have a more general aim of showing that paranormal effects largely go into decline for some lawful reason, but they offer no theoretical explanation. Indeed, the problem of explanation worsens if the declines are not widespread, since exceptions do not confirm the rule in such cases. Also plausible is

the possibility that psi can sometimes go into decline (or not) depending on how it is tested or how it is categorized. In fact, the various domains and accompanying methodologies suggest psi is governed by situational factors, rendering it possible that even the main categories, ESP and PK, are vague. Our findings in this brief report will give some support to these perspectives.

B&S re-analyzed five of our (Tressoldi & Storm, 2024; T&S) databases, all of which were comprised of studies on anomalous cognition (AC; a.k.a. extra-sensory perception or ESP). They added a new database referred to as 'device-AC'

(i.e., telephone-telepathy), and another four databases related to psychokinesis (PK) phenomena. All databases are available open access at <https://zenodo.org/records/18822509>.

The opening lines of the B&S paper are misleading. The reader will note that Table 1 in T&S (2024) has Ganzfeld, Remote Viewing, Forced-Choice, Presentiment, and Dream-ESP. T&S did not cover DMILS, Remote Helping, Remote Staring, or micro-PK. B&S say their aim is “... to correct the false conclusions that T&S drew in their JSE articles” (p. 269).

It seems to us B&S give the reader the false understanding that we are accountable for claiming there are no widespread declines across ten domains, which we never claimed, because we only covered “five experimental protocols” (T&S, 2024, p. 462), and we did not find any declines in three: Ganzfeld, Forced-Choice, and Presentiment. While Remote Viewing and Dream-ESP showed significant declines, T&S showed no significant decline for Dream-ESP when the two very different databases, Maimonides and non-Maimonides studies, were tested separately.

The B&S main difference in the statistical analysis with respect to T&S, is the use of the non-parametric correlation, Spearman *rho*, instead of the meta-regression used by T&S, which takes into account the fact that larger studies have more influence than smaller studies on the effect-size relationship because each study can estimate its own effect size more precisely. That is, larger  $N \rightarrow$  smaller variance  $\rightarrow$  smaller standard error  $\rightarrow$  higher precision.

Furthermore, Spearman’s *rho* is a rank-based correlation between two variables. It is not really suitable for estimating slopes because it implies only a monotonic association. To go further, we note that B&S use the argument from outliers to justify their using non-parametric

tests. Parametric analyses (such as Pearson’s  $r$ ) do make assumptions about distributions, but parametric tests are robust (Hedges et al., 2010). (See also the Appendix for a comparison between Spearman’s *rho* and Pearson’s  $r$  test values).

Using meta-regression (as T&S did) assumes roughly linear relationships, interval scaling, no extreme skews, and homoscedasticity; assumptions which were all met. If the relationships are curved but monotonic, linear models will underestimate it, even with zero outliers. But are they monotonic? What B&S must show are declines that are monotonic. Their paper fails to show it. Consider our figures (T&S, 2024, pp. 463–464). Visual inspection shows no obvious monotonicity.

Analyzing the correlation between the year of the studies and the effect sizes, B&S observed that all four databases related to PK and three out of five databases related to AC showed statistical signs indicating decline effects. See their Table 1.

However, as we have stated above and as discussed in our original T&S study, if we consider the differences between the Maimonides and non-Maimonides studies that concern AC in a dreaming condition, only two out of six experimental protocols related to AC show a statistically significant decline with both statistical approaches (i.e., Spearman’s *rho* and meta-regression).

Furthermore, even though not related to the DE, B&S found statistically significant correlations between the effect sizes and the number of trials, a sign of an increase in the statistical power in all but three databases (see their Tables 2 and 3). Consequently, to be confident that the DE really is of issue here, it is procedurally wise to partial out the influence of the number of trials or participants on the effect sizes (see B&S, Table 4).

**Table 1.** Comparison between T&S Updated, and B&S Table 4 Results.

Protocol	Type of psi	Years span	<i>N</i> of studies	T&S slope	T&S $p$ (one-tailed)	B&S <i>rho</i>	B&S $p$ (one-tailed)
Remote Viewing	AC	1974–2020	40	–0.009	0.01	–0.27	0.05
All Dreams [Post-Maimonides only]	AC	1966–2014	50 [36]	–0.005 [–0.02]	0.055 [0.21]	–0.19 [–0.24]	0.096 [0.08]
Forced Choice	AC	1987–2022	137	0.0004	0.075	0.10	0.88
Ganzfeld	AC	1974–2020	113	0.001	0.29	–0.04	0.33
Presentiment	AC	1997–2017	62	0.047	0.18	–0.01	0.35
Device-AC	AC	2003–2023	26	–0.01	0.0009	–0.59	<0.001
Remote Helping	PK	1995–2006	11	–0.036	0.012	–0.62	0.02
Remote Staring	PK	1989–1998	15	–0.035	0.06	–0.60	0.01
RNG	PK	1969–2004	380	–0.0003	<0.001	–0.15	0.001
DMILS	PK	1977–2000	36	–0.014	0.0037	–0.33	0.02



In Table 1 above, we compare the T&S and B&S results after the inclusion of the number of trials or the number of participants in the case of the presentiment data as a covariate in the different statistical analyses, meta-regression and Spearman  $\rho$ , respectively.

The differences between the results reported by T&S and B&S in their Table 4 are small. Protocols in bold are those where both T&S and B&S statistics agree on the statistical evidence of DE.

To summarize Table 1, only *two out of six AC experimental protocols* showed a statistically significant DE. In contrast, *all four PK experimental protocols* showed a statistically significant indications of DE with both statistical approaches.

The results reported by T&S original study are confirmed even with these new analyses.

However, to truly understand this DE 'size', we must refer not to  $p$ -values alone, but to the slope and correlation values (see T&S slope and B&S  $\rho$  columns in Table 1), possibly with their confidence intervals for a better estimate. For example, there is a large difference in both the slope *and* the correlation between RNG-PK and Remote Helping,  $-0.0003$  vs.  $-0.15$ , and  $-0.036$  vs.  $-0.62$ , respectively.

We also note that B&S's Table 1 reports a cumulative figure of  $r_s = +0.07$ , which is not significant but suggests a general increase, albeit for a heterogeneous dataset this is not entirely justified if the argument is that the dataset should be made homogeneous. But under what principle do we accept either assumption?

Also, it must be taken into account that all these PK databases are quite obsolete. B&S databases provide no data for approximately *the last 20 years*. Furthermore, the Remote Helping and Remote Staring databases include only 11 and 15 studies, respectively, and cover only 12 and 10 years, respectively, which is a very limited number of

studies and time span and therefore does not justify calling the DE estimates robust; they are tentative at best.

B&S also ran a simulation using the arbitrary number of the first 20 studies of each experimental protocol and pooling them together into a single database (see B&S Figure 2). They based this procedure on the unproven hypothesis that all AC and PK phenomena derive from a common cause. An added problem is that the choice of testing 20 studies is a rather arbitrary one. How is it justified? Should B&S not also model for 10, 30, 40 studies, and so on?

To conclude, B&S's overall interpretation is that "*after using all available meta-analytic data and non parametric methods to quantify the decline, there is a decline over years within protocols, with the remarked exception of the presentiment protocol*" (p. 272).

We contend that this interpretation overstates what the results show (see Table 1) and that the more appropriate conclusion is as we have previously stated in our original paper: "*There is no evidence of a general DE across the different experimental protocols*" (Tressoldi & Storm, 2024, p. 461).

## REFERENCES

- Bierman, D. J. (2024). The decline of effect size in psi research: An evidence-based commentary on Tressoldi and Storm (2024). *Journal of Scientific Exploration*, 40(2), 268–273. <https://doi.org/10.31275/20264003>
- Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, 1(1), 39–65. <https://doi.org/10.1002/jrsm.5>
- Tressoldi, P., & Storm, L. (2024). The myth of the decline effect in psi research: The empirical evidence. *Journal of Scientific Exploration*, 38(3), 461–465. <https://doi.org/10.31275/20243313>

## APPENDIX

To come back to an earlier point about assumptions and test suitability, we feel it would be helpful to make a comparison between parametric and non-parametric test results using values derived from Pearson's  $r$  and Spearman's  $\rho$  tests, respectively, for the 10 protocols. Table A1 shows that the relevant protocol pairs are mostly similar.

Be that as it may, justification for a specific test has to be based on assumptions appropriate to the data. It is arguable which test suits the data best, and it matters not that values are the same in strength and direction. Crucially, what matters most is how the data distribute for the various protocols, and we may know better as databases accumulate in size over time; we may see discrepancies between values based on which test is used; or we may not.

**Table A1.** Correlational Values: Spearman's  $\rho$  vs. Pearson's  $r$ .

Protocol	Spearman's $\rho$	Pearson's $r$
Remote Viewing	-0.45*	-0.48*
Dreams	-0.28*	-0.28*
Forced Choice	0.11	0.09
Ganzfeld	-0.05	-0.09
Presentiment	0.02	0.04
Device-AC	-0.62*	-0.59*
Remote Helping	-0.64*	-0.59*
Remote Staring	-0.70*	-0.66*
RNG-PK	-0.26*	-0.22*
DMILS	-0.36*	-0.48*

Note: Spearman's  $\rho$  values come from the B&S paper Table 1. Pearson's  $r$  are obtained from all the meta-analytic databases discussed in the paper, and available at <https://zenodo.org/records/18822509>.

