J. Behav. Ther. & Exp. Psychiat. 54 (2017) 307

Contents lists available at ScienceDirect

Journal of Behavior Therapy and Experimental Psychiatry

journal homepage: www.elsevier.com/locate/jbtep

Erratum to "Behavior as information about threat in anxiety disorders: A comparison of patients with anxiety disorders and non-anxious controls" [Journal of Behavior, Therapy and Experimental Psychiatry, 45, 489–495]

Marcel van den Hout*, Amelia Gangemi, Francesco Mancini, Iris M. Engelhard, Marleen M. Rijkeboer, Marcel van Dam, Irene Klugkist

Utrecht University, Clinical Psychology, Heidelberglaan 1, 3584CS Utrecht, Netherlands

In 2014 we published a paper in this Journal: Hout, M. A. van den, Gangemi, A., Mancini, F., Engelhard, I.M., Rijkeboer, M.M., van Dam, M., & Klugkist, I. (2014): Behavior as information about threat in anxiety disorders: A comparison of patients with anxiety disorders and non-anxious controls. *Journal of Behavior, Therapy and Experimental Psychiatry, 45*, 489–495.

In that paper we reported a replication of an experiment by Gangemi et al. (2012): Behavior as information: "If I avoid, then there must be a danger". *Journal of Behavior Therapy and Experimental Psychiatry, 43*, 1032–1038.

The van den Hout et al. (2014) findings were in line with the Gangemi results but, using traditional null hypothesis testing (NHT), our findings were not significant ($p > 0.05$). There are several problems with NHT and one of them was encountered here: it would be irrational and misleading to argue that a study where the crucial interaction was significant (e.g.: $p = 0.04$) is contradicted by a replication study showing the same pattern of interaction but with a p value of, say, 0.06.

Replication is crucial to sound science, and in our 2014 paper we introduced and reported a novel analysis in this area: Bayesian analysis of constrained hypotheses. We calculated the Bayes factor (BF) for the Gangemi experiment ($BF = 3.31$) and for our replication ($BF = 2.34$), multiplied the two BF's and reported the product ($BF = 7.75$) as the best estimate of the empirical support for the hypothesis after both experiments.

We recently found out that this multiplication, simple as it is, was inappropriate. In the context of testing constrained hypotheses, the error can be explained as follows. The BF is a model selection criterion that combines a measure of fit ('how well do the data fit with the constraints of the hypotheses') and a penalty for model size (to prevent overfitting). By multiplying two BFs of two replication studies, the correction for model size is incorrectly applied twice. As a result, the support for smaller models is overestimated.

In the general context of hypothesis testing with Bayes factors, this is also noted in a recent paper by Wagenmakers, Verhagen and Ly (2016) with reference to the earlier explanation by Jeffreys (1961, p. 332–334) stating that, to combine test results by multiplication, prior distributions need to be properly updated.

The product of BFs without updating prior distributions is interpretable; it can be seen as the empirical support for finding the hypothesized effect in *each* experiment. This, however, is not what we want to measure. What we do want to evaluate, given two or more independent data sets, is whether and to what degree the data, taken together, support the hypotheses. Combining empirical information from replication studies this way is called Bayesian updating.

The right calculation for the 'updated BF' uses the data of both experiments to derive the posterior distribution of the parameters and the subsequent fit measures for each hypothesis. Then, the 'updated fit' is corrected for the model size penalty of each hypothesis to derive the appropriate Bayes factors. In practice, with the BIEMS software, this can be done as follows. First, the data from the first experiment are analyzed with the default, diffuse prior, which is determined by BIEMS. Then the same prior is used with the data from both experiments (i.e., the raw data from both experiments are combined into one file). The resulting BFs in the second analysis are the correct BFs after Bayesian updating.

The conclusion in our 2014 paper was that anxiety patients infer threat from safety behavior, especially in safe contexts. Importantly, this conclusion is not undermined by doing the calculations right. In fact, the combined empirical support for that conclusion is stronger with the right BF: it is not 7.75, which was reported in the 2014 paper, but it is 14.95.

Finally, note that in this paper the name of one of the co-authors was misspelled. The correct name is van Dam, M.

References

Jeffreys, H. (1961). *Theory of probability* (3 ed.). Oxford, UK: Oxford University Press.

Wagenmakers, E.-J., Verhagen, A. J., & Ly, A. (2016). How to quantify the evidence for the absence of a correlation. *Behavior Research Methods, 48*, 413–426.

DOI of original article: http://dx.doi.org/10.1016/j.jbtep.2014.07.002.

* Corresponding author.
E-mail address: m.a.vandenhout@uu.nl (M. van den Hout).

http://dx.doi.org/10.1016/j.jbtep.2016.05.004