In light of differences in findings presented in our meta-analysis (1) and a previous publication (2), Poeppl et al. (3) hypothesize that the discrepancies are potentially due to the inclusion of region-of-interest (ROI)-based studies, the choice of a larger cluster threshold, or different inclusion criteria. In order to accurately address these concerns, we tested the validity of each hypothesis statistically. We agree with Poeppl et al. that, in general, the inclusion of ROI-based studies may potentially lead to a skewness in null space. Thus, to ensure the validity of our findings in a general context, we have excluded all ROI-based coordinates and reduced the cluster threshold to 10 mm$^3$, which is sufficient to detect even small brain regions such as hypothalamus. We repeated the activation likelihood estimation (ALE) meta-analyses using the new dataset and parameter. The analyses suggest that our findings remain robust, and these hypotheses were statistically rejected. Moreover, we have included the coordinates of the studies that our keyword-based search missed (4–6), and our findings remain unchanged.

It is noteworthy that meta-analyses are highly sensitive to the input space, and analyses based on different datasets are hardly comparable even if they attempt to address the same question. This is an intrinsic dynamical feature of any meta-analysis. Poeppl et al. (3) advocate for a restrictive inclusion of data from studies with a primary objective to investigate sexual arousal. In turn, we advocate for inclusion of data from any cue reactivity study that investigated response to visual sexual cues in contrast to neutral stimuli in healthy, drug-naïve individuals. This difference in study selection leads to a larger input space in our study and, naturally, to different outcomes.

We further simulated different scenarios to investigate how different control strategies for false positives may affect the final outcomes of the meta-analyses. Our findings indicate that using an uncorrected, relaxed $P$ value of $P < 0.05$ increases the probability of discovering contrast effects, whereas corrections for multiple comparisons or $P < 0.001$, as applied in our paper, do not converge to any sex-specific effects.

In summary, we are confident that the discrepancies in the findings are explained by differences in inclusion criteria and control for false positives. With a growing number of studies investigating the relevance of biological sex on cognitive processes, all meta-analyses, including our work, have to be revised. Nevertheless, robustness is often achieved by approaching the statistical problems in a conservative manner, and a strong yet reasonable control of false positives is necessary to ensure the validity of the findings.


*Department of Physiology of Cognitive Processes, Max Planck Institute for Biological Cybernetics, 72076 Tübingen, Germany; and Imaging Science and Biomedical Engineering, University of Manchester, Manchester M13 9PL, United Kingdom

Author contributions: H.R.N. designed research; R.K. and H.R.N. analyzed data; and E.M., N.K.L., and H.R.N. wrote the paper.

The authors declare no competing interest.

Published under the PNAS license.

To whom correspondence may be addressed. Email: hamid.noori@tuebingen.mpg.de.