Rebuttal of "Voodoo Correlations in Social Neuroscience" by Vul et al. – summary information for the press

Mbemba Jabbi[1], Christian Keysers[2], Tania Singer[3], Klaas Enno Stephan[3],
(authors listed in alphabetical order)

[1]National Institute of mental Health, Bethesda, Maryland, USA.

[2]University Medical Center Groningen, Department of Neuroscience, University of Groningen, The Netherlands. www.bcn-nic.nl/socialbrain.html

[3] Laboratory for Social and Neural Systems Research, University of Zurich, Switzerland. http://www.socialbehavior.uzh.ch/index.html

The paper by Vul et al., entitled "Voodoo correlations in social neuroscience" and accepted for publication by *Perspectives on Psychological Science*, claims that "a disturbingly large, and quite prominent, segment of social neuroscience research is using seriously defective research methods and producing a profusion of numbers that should not be believed." In all brevity, we here summarise conceptual shortcomings and methodological errors of this paper and explain why their criticisms are invalid. A detailed reply will be submitted to a peer reviewed scientific journal shortly

1. The authors misunderstand the critical role of multiple comparison corrections and conflate issues pertaining to null hypothesis testing and effect size estimates, respectively.

Vul et al. argue that it is misleading to identify, in a first step, voxels showing a significant correlation between brain activity and a psychological variable and, in a second step, to report the magnitude of this correlation. Although they are aware that the statistical test underlying the selection procedure is typically combined with a correction for multiple comparisons they misunderstand the consequences of this fact. Correcting for multiple comparisons ensures that the correlations exhibited by the selected voxels do not exceed a certain probability of having occurred by chance. Making this correction increasingly strict has two consequences: (i) significant correlations will be found more and more rarely, but at the same time, (ii) their magnitude will increase. This follows directly from the principles of statistical hypothesis testing, there is nothing magic or surprising about this effect. Once the significance of a correlation has been determined, one can illustrate the effect size by reporting the magnitude of the correlation coefficient, graphically or numerically. This combined reporting of significance (p-values) and effect sizes (e.g. correlation coefficients) is in accordance with the statistical guidelines of the American Psychological Association (Wilkinson et al. 1999).

2. The authors make strong claims on the basis of a questionable upper bound argument.

Vul et al. argue that many of the brain-behavior correlations published in social neuroscience articles are *"impossibly high"* and that *"the highest possible meaningful correlation that could be obtained would be .74"*. This categorical claim is based on a statistical upper bound argument which relies on the questionable assumption that *"fMRI measures will not often have reliabilities greater than about .7"*. However, logically, any theoretical upper bound argument would have to be based on the highest reliability values ever reported for behavioural and fMRI data, respectively (e.g. for fMRI, near-perfect reliabilities of 0.98 have been reported in Fernandez et al. 2003).

3. The authors use misleading simulations to support their claims.

Both simulations provided by Vul et al. are misleading. For brevity, we here comment only on technical flaws of their first simulation. Had this simulation been implemented properly, using family wise error correction at a significance level of 5%, maximally every 20th simulation would have resulted in a single voxel (of the 10,000 tested) which would have falsely been deemed significant. Of course, such a single voxel is still a false positive and would correspond to an impressive correlation, but it would occur extremely rarely. In contrast, Vul et al. convey the impression that "preselecting" voxels by statistical testing will generally result in many false positive voxels.

4. The authors inappropriately dismiss the existence of non-significant correlations.

As the authors admit, their criticism necessarily implies that the criticised studies should not find any non-significant correlations. They dismiss the fact that non-significant correlations have indeed been reported frequently by examining a single case in which the appearance of non-significant correlations can be explained in a way that does not contradict their account. However, several of the criticised studies – including our own ones – and several studies not reviewed in the present paper reported non-significant correlations that cannot be dismissed in this fashion. In these studies, the same indices of brain activation were used to perform fully independent regression analyses with different behavioural sub-scales, but only some of the sub-scales revealed significant results (e.g. Jabbi et al., 2007).

5. The authors' understanding of the rationale behind the use and interpretation of correlations in social neuroscience is incomplete.

In contrast to the interpretation by Vul et al., a key question in social neuroscience is often not *how strongly* two measures were correlated, but *whether* and *where in the brain* such correlations may exist. In the Jabbi et al. (2007) paper, for example, results did not only confirm that brain activity during the vision of other's facial expressions correlated with empathy questionnaires even after applying appropriate corrections for multiple comparisons, but also that such correlations arise in the very same regions involved in the participants' own experience of similar emotions, providing important evidence for simulation theories. This finding stands independently of the absolute magnitude of such correlations and is just one example of many studies in which the *where* and *whether* of correlations are what carries the decisive information.

6. The authors ignore that the same brain-behaviour correlations have been replicated by several independent studies and that major results in social neuroscience are not based on correlations at all

The authors suggest that major work in social neuroscience has been published in high-impact journals only because of the reported size of brain-behaviour correlations. This claim is not true. First, many studies based their main results on other analyses than correlation analyses (e.g. Singer et al. 2004, 2006) and these main findings, for example in the domain of empathic brain responses, were subsequently replicated many times by different groups across the world (e.g., Gu & Hahn 2007; Jackson et al. 2005, 2006; Lamm et al. 2007a, 2007b; Saarela et al. 2007; for a review see Singer and Leiberg, 2009). Moreover, correlations between these empathy-relevant brain areas and individual difference measures have now been reported in a large number of studies using different methods including parametric regression analysis in single-subject analyses (e.g., Jackson et al. 2006; Lamm et al. 2007a; Saarela et al. 2007; Singer et al. 2008). These multiple replications, using different methods, mean that it is very unlikely that the reported results are merely false positives and demonstrate that many of the results obtained by social neuroscience have already passed the most stringent test in science: replication by independent studies.

7. The authors used an ambiguous and incomplete questionnaire.

As basis for their survey, the authors sent a questionnaire to a selective set of scientists. This questionnaire contained questions that were ambiguous and incomplete. The authors neither asked which correction method was employed nor whether a secondary statistical test was actually really applied to the data of the significant voxels. This is important because, as explained above, there is no problem per se in reporting correlation coefficients as effect sizes of significant results, as long as these significant results survived multiple comparison correction and if no dependent secondary statistical tests were applied to the selected voxels (in some studies this may have been unclear due to the complexity of the description of data analysis). We suspect that due to the ambiguous and incomplete questionnaire multiple studies may have been misclassified by the authors as "invalid".

8. The authors make flawed suggestions for data analysis.

The authors advocate the use of "independent" data and suggest that data from the same subject could be split up in half where "half of the data are used to select a subset of voxels exhibiting the correlation of interest, and the other half of the data are used to measure the effect (examining the same voxels, but looking at different runs of the scanner)." This suggestion is flawed since it is common knowledge in statistics that repeated measures from the same subject are never independent; the activity in a given voxel during the first half of the data will be correlated with its activity during the second half of the data. It is this well-known phenomenon that led to the development of covariance component models, e.g. repeated measures ANOVA, in statistics in order to deal with multiple measurements from the same subject.

Conclusions

In this summary, we have provided a very brief summary that exposes some of the flaws that undermine the criticisms by Vul et al. We have pointed out that brain-behaviour correlations in social neuroscience are valid, provided that one adheres to good statistical practice. It has also been emphasized that many analyses and findings in social neuroscience do not rest on brain-behaviour correlations and have been replicated several times by independent studies conducted by different laboratories. A full analysis of the Vul et al. paper and a detailed reply will be submitted to a peer-reviewed scientific journal shortly.

References

Fernández, G., Specht, K., Weis, S., Tendolkar, I., Reuber, M., Fell, J., Klaver, P., Ruhlmann, J., Reul, J., Elger, C.E. 2003. Intrasubject reproducibility of presurgical language lateralization and mapping using fMRI. Neurology 60, 969-975.

Jabbi, M., Swart, M., Keysers, C., 2007. Empathy for positive and negative emotions in the gustatory cortex. Neuroimage 34, 1744-1753.

Jackson, P. L., Brunet, E., Meltzoff, A. N., and Decety, J., 2006. Empathy examined through the neural mechanisms involved in imagining how I feel versus how you feel pain. *Neuropsychologia* 44: 752-761.

Jackson, P. L., Meltzoff, A. N., and Decety, J., 2005. How do we perceive the pain of others? A window into the neural processes involved in empathy. *Neuroimage* 24: 771-779.

Lamm, C., Batson, C. D., and Decety, J., 2007a. The neural substrate of human empathy: effects of perspective-taking and cognitive appraisal. Journal of Cognitive Neuroscience 19: 42-58

Lamm, C., Nusbaum, H. C., Meltzoff, A. N., and Decety, J., 2007b. What are you feeling? Using functional magnetic resonance imaging to assess the modulation of sensory and affective responses during empathy for pain. PLoS ONE 2: e1292. doi: 10.1371/journal.pone.0001292

Saarela, M. V., Hlushchuk, Y., Williams, A. C., Schurmann, M., Kalso, E., and Hari, R., 2007. The compassionate brain: humans detect intensity of pain from another's face. *Cerebral Cortex* 17: 230-237.

Singer, T., & Leiberg, S. (in press). Sharing the emotions of others: The neural bases of empathy. In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences IV.*

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R.J., Frith, C.D., 2004. Empathy for pain involves the affective but not sensory components of pain. Science 303, 1157-1162.

Singer, T., Seymour, B., O'Doherty, J.P., Stephan, K.E., Dolan, R.J., Frith, C.D. 2006. Empathic neural responses are modulated by the perceived fairness of others. Nature 439, 466-469.

Singer, T., Snozzi, R., Bird, G., Petrovic, P., Silani, G., Heinrichs, M., Dolan, R.J. 2008. Effects of oxytocin and prosocial behavior on brain responses to direct and vicariously experienced pain. Emotion 8, 781-791.

Vul, E., Harris, C., Winkielman, P., Pashler, H. (in press) Voodoo correlations in social neuroscience. Perspectives on Psychological Science.

Wilkinson, L. and the Task Force on Statistical Inference. 1999. Statistical Methods in Psychology Journals: Guidelines and Explanations. *American Psychologist* 54, 594-604.