Second revision:
Supplementary Material

Linking brain-wide multivoxel activation patterns to behaviour: examples from language and math

Rajeev D. S. Raizada,[1] Feng Ming Tsao,[2] Huei-Mei Liu,[3]
Ian D. Holloway,[4] Daniel Ansari,[4] and Patricia K. Kuhl[5]

[1]Neukom Institute for Computational Science, HB 6255, Dartmouth College, Hanover NH 03755.
[2]Dept. of Psychology, National Taiwan University, Taipei 10617, Taiwan.
[3]Dept. of Special Education, National Taiwan Normal University, Taipei 10644, Taiwan.
[4]Dept. of Psychology, Univ. of Western Ontario, London, ON N6G 2K3, Canada.
[5]Inst. for Learning & Brain Sciences, Univ. of Washington, Box 357988, Seattle WA 98195.

S.1 On the interpretability of widely distributed classifier-weight brain maps

It is certainly the case the the widely distributed classifier-weight brain maps obtained in the present study are harder to interpret than the more localised activation clusters that typically emerge from standard univariate fMRI analyses. Because standard fMRI analyses consider each voxel one at a time, they fail to detect neural coding which is distributed across spatially separated voxels. Contiguous voxels in a cluster do indirectly affect each other's analyses, in virtue of their activations being spatially smoothed into each other. The result of this is that standard fMRI tends to reveal local clusters of activation, an outcome that is further accentuated by the standard step of thresholding-out clusters containing fewer than a specified number of voxels. The field has therefore become accustomed to viewing discrete clusters of activated voxels as being the markers of good-quality and interpretable signal.

In contrast, multivariate analyses are able to reveal combinatorial coding distributed across non-contiguous and possibly quite distant voxels. Although such activation patterns look quite different from what has typically been held to be interpretable signal, this does not mean that they are less faithful depictions of what the brain is actually doing. On the contrary, it may be the localised and discrete clusters of brain activation of the sort which univariate analyses have tended to produce that are misleading. As an analogy: in genetics, the most interpretable results are when diseases turn out to be monogenic, such as Huntington's disease (Bhattacharyya, 2008). However, the vast majority of diseases are polygenic, with fifty or more genes often needing to be considered in concert for genuine predictive power to emerge (e.g., Baker & Kramer, 2006). The brain maps in our manuscript, like those in Marquand et al. (2009), do not consist of just a few discrete clusters. This does indeed make them harder to interpret, but the genetic bases of most diseases

are harder to interpret than the genetics of Huntington's. It could, possibly, be the case that the widely distributed classifier-weight maps found by our study and also by others may be failing to capture a true and more simply localised underlying neural activation. However, it may also be the case that such maps genuinely, but imperfectly, reflect the actual occurrence of much more distributed processing. Which of these possibilities is actually the case is an open question.

S.2 On the effect of zero-meaning the data

In the present study, we subtracted the mean-over-time of each voxel's time-course from every time-point, so that the mean value of the resultant time-course was zero.

Zero-meaning the data does not affect the dynamic range (the difference between the maximum and minimum values of the data), as the subtraction of the mean affects both the maximum and the minimum equally, and hence does not alter the difference between them. Thus, shifting the voxels to a zero mean does not reduce the information. A voxel which has MRI signal of intensity, say, 220 during Condition A and intensity 180 during Condition B carries no more and no less information after its time-course is zero-meaned so that its new intensities are +20 and -20 respectively.

Perhaps the clearest way to show that this zero-meaning step does not affect the information in the data is to note that a mathematically equivalent and alternative step involves leaving the data completely unchanged, and instead appending a regressor to the statistical model with a constant value set to one. This is often described as adding a "bias term" to the model (Bishop, 1995). A corresponding bias-weight multiplies the bias-term to produce a constant offset, which will be equal to the overall mean of the dataset. Examples of such constant terms are the constant-valued columns at the far right side of an SPM design matrix. The beta-values corresponding to these constant-columns capture the mean-value of the signal for each run. Their mathematical effect is identical to that of having subtracted the mean from each time-course before feeding the data into the model. These constant columns do not affect the information contained in the fMRI signal when they are used in SPM (or any other GLM-based analysis package), and in exactly the same way they leave the information equally intact in the linear classifiers used in the present study.

Supplementary References

Baker, S. G. & Kramer, B. S. (2006). Identifying genes that contribute most to good classification in microarrays. *BMC Bioinformatics*, *7*, 407.

Bhattacharyya, N. P. (2008). Huntington's disease: a monogenic disorder with cellular and biochemical complexities. *FEBS Journal*, *275*(17), 4251.

Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford: Clarendon Press.

Holloway, I. D. & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: the numerical distance effect and individual differences in children's mathematics achievement. *J Exp Child Psychol*, *103*(1), 17–29.

Marquand, A., Howard, M., Brammer, M., Chu, C., Coen, S., & Mourão-Miranda, J. (2009). Quantitative prediction of subjective pain intensity from whole-brain fMRI data using Gaussian processes. *NeuroImage*.

Price, G. R., Holloway, I., Räsänen, P., Vesterinen, M., & Ansari, D. (2007). Impaired parietal magnitude processing in developmental dyscalculia. *Curr Biol*, *17*(24), R1042–3.

S.3 Supplementary Figures

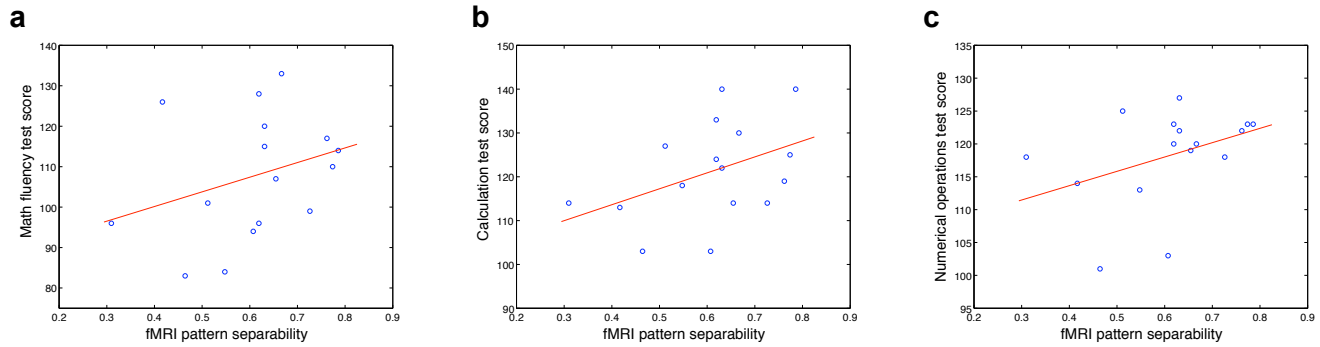

Figure S1: *Scatterplots and regression lines showing the data whose correlation coefficients are plotted in the bar graph of Figure 4 in the main text. The correlations are between fMRI pattern separability in a non-symbolic numerical distance effect task between small-distance and large-distance number pairs and standardised test scores on mathematics tasks.* **(a):** *The correlation with WJ-III (Woodcock-Johnson III) Math Fluency is positive, but does not reach significance: $p = 0.144$, $\rho = 0.382$.* **(b):** *The correlation with WJ-III Calculation is significant: $p = 0.0496$, $\rho = 0.498$, two-tailed.* **(c):** *The correlation with WIAT (Wechsler Individual Achievement Test) Numerical Operations is marginally significant: $p = 0.0674$, $\rho = 0.468$.*

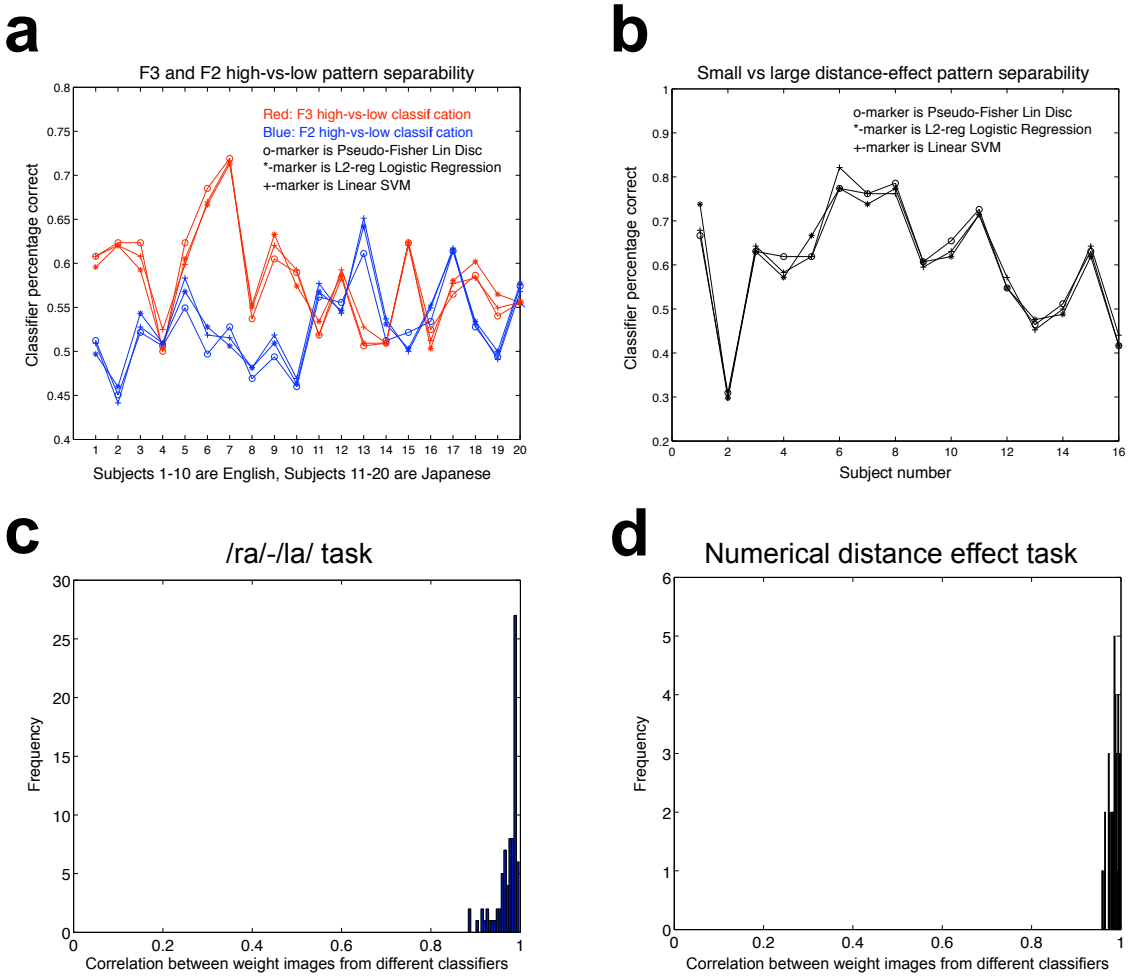

Figure S2: *Despite its simplicity, the Pseudo-Fisher Linear Discriminant classifier produces whole-brain pattern separability results which are extremely similar to those obtained using classifiers which are, in principle, considerably more powerful. It was compared against L2-regularised logistic regression and a linear SVM, with all the classifiers applied to the same whole-brain vectors and using the same cross-validation scheme.* **(a):** *In the /ra/-/la/ data set, the percentage-correct values obtained by all three classifiers for F3-differences and F2-differences are almost identical, as can be seen by the fact that the lines sit on top of each other. The logistic regression and SVM classifiers often, but not always, achieve very slightly higher percentage correct scores than does the simple Pseudo-Fisher Linear Discriminant.* **(b):** *In the numerical distance effect data set, the different types of classifier again yield very similar results.* **(c):** *As well as the percentage-correct scores being almost the same, the brain-wide weight maps produced by the classifiers are also very highly correlated. This histogram of correlations between the weight maps in the /ra/-/la/ task from the three different pairwise classifier comparisons shows that the correlations ranged from very high to almost perfect.* **(d):** *In the numerical distance effect task, the weight-map correlations are equally strong.*

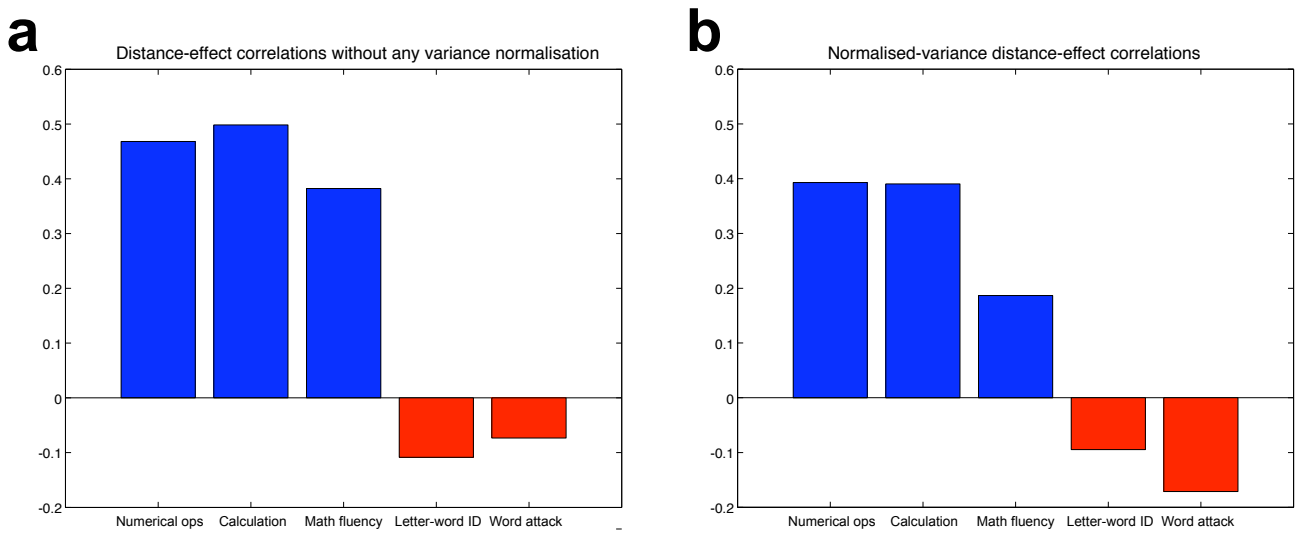

Figure S3: *As is discussed in Section 3.5 of the main text, normalising the variance of the MRI voxels before entering them as input to the classifier has the effect of slightly reducing the resulting correlations between fMRI pattern separability and behaviour.* **(a):** *Correlations without any variance normalisation.* **(b):** *Slightly weaker correlations after applying variance normalisation.*

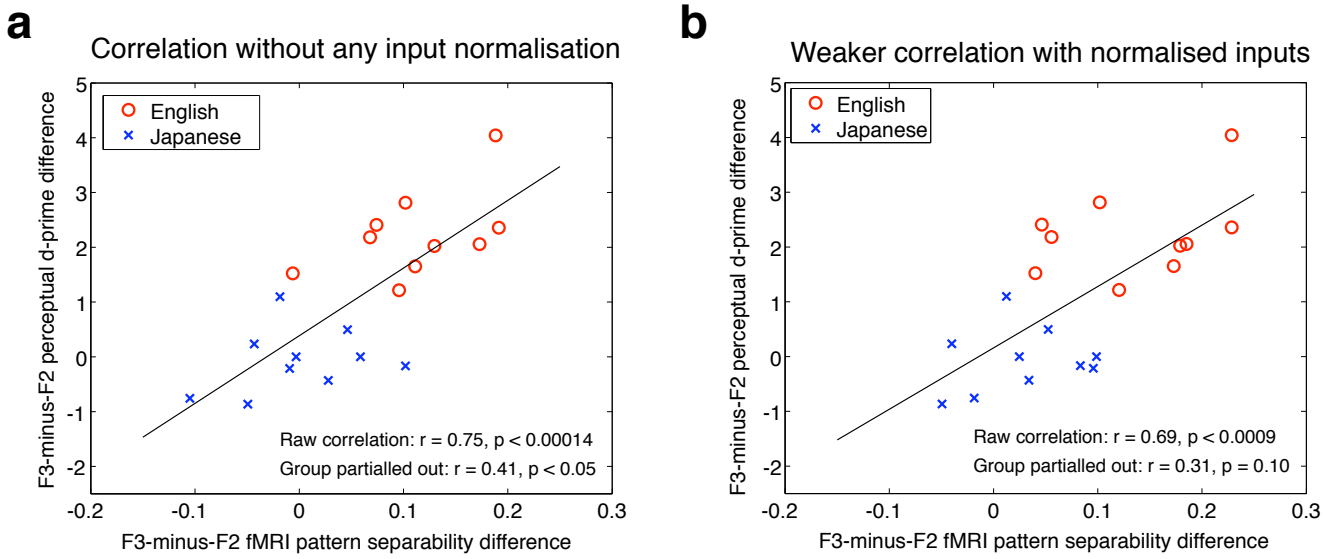

Figure S4: *As is discussed in Section 3.5 of the main text, normalising the variance of the MRI voxels before entering them as input to the classifier has the effect of slightly reducing the resulting correlations between fMRI pattern separability and behaviour.* **(a):** *Correlations without any variance normalisation.* **(b):** *Slightly weaker correlations after applying variance normalisation.*

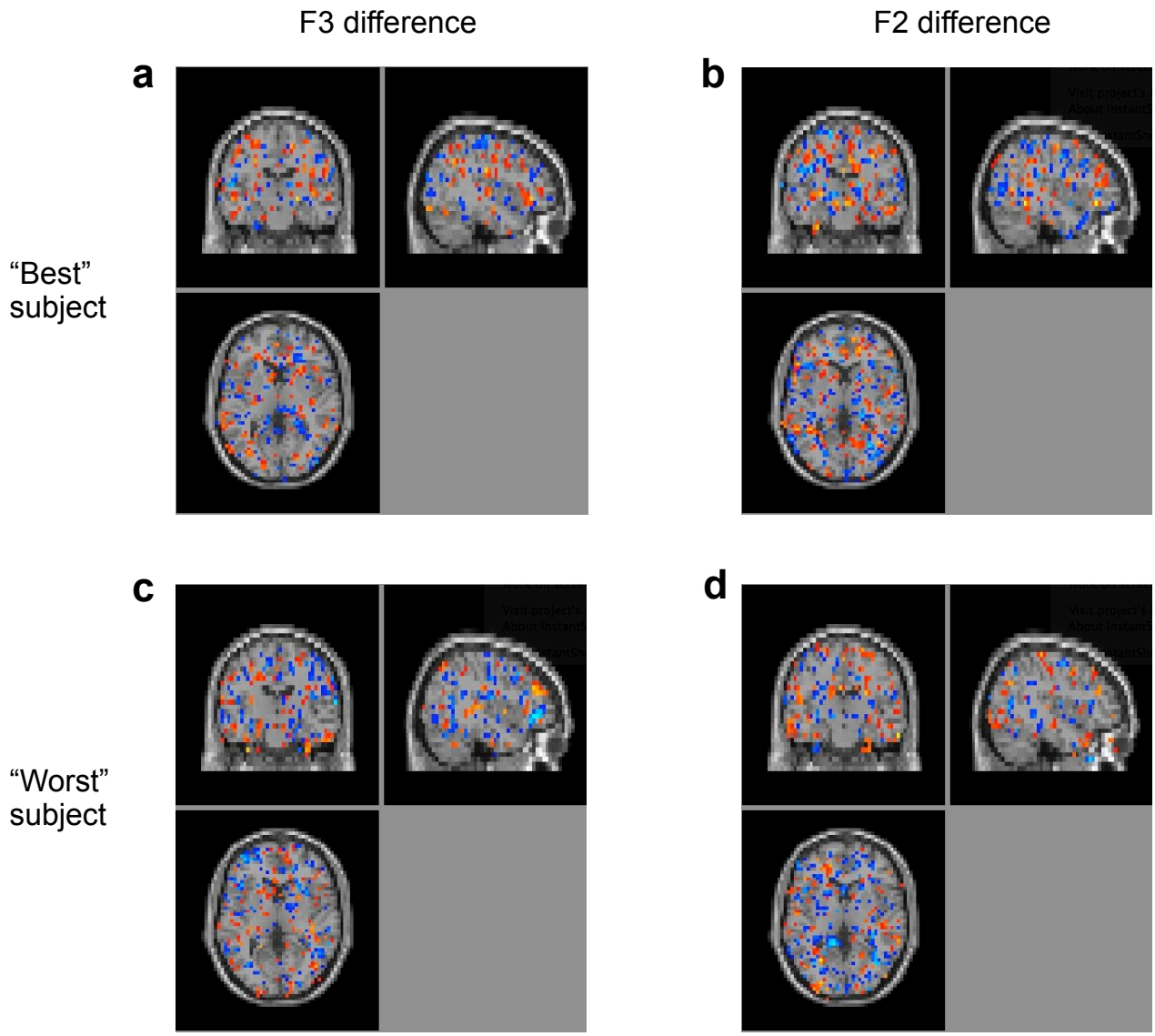

Figure S5: *The positive and negative F3 and F2 weight maps of individual subjects, specifically the "best subject", namely the person whose perceptual ability to hear F3-differences most exceeds their ability to hear F2-differences (this happens to be the English subject En06), and the "worst subject", who shows the opposite behavioural pattern (Japanese subject Jp07). It can be seen that the weights are highly distributed throughout the brain. Although these individual-level maps are interesting to inspect, it is hard to gain much interpretive information from them. For that, the group-level maps shown in Fig. 6 in the main text are more appropriate. Positive weights are shown in orange, and negative weights in blue, with maps thresholded at the weight-strength of* ± 0.001.