MIS Quarterly

RESEARCH ARTICLE

NEURAL CORRELATES OF MULTIDIMENSIONAL VISUALIZATIONS: AN FMRI COMPARISON OF BUBBLE AND THREE-DIMENSIONAL SURFACE GRAPHS USING EVOLUTIONARY THEORY

Eric Walden, Gabriel S. Cogo, David J. Lucus, and Elshan Moradiabadi
Rawls College of Business, Texas Tech University, Box 42101,
Lubbock, TX 79409 U.S.A.
{eric@ericwalden.net} {gabriel.cogo@ttu.edu} {david.j.lucus@ohio.edu} {elshan.moradiabadi@ttu.edu}

Roozmehr Safi
Department of Management, University of Missouri–Kansas City, 5110 Cherry Street,
Kansas City, MO 64110 U.S.A. {safir@umkc.edu}

Appendix

In this appendix, we include a variety of nontechnical information, technical details, explanations, and robustness tests. We follow the methodology of Dimoka (2012). We will offer a brief outline here, but for a more detailed and complete explanation, see Dimoka's study. We try to explain concepts and procedures in laymen's terms where possible, but some details are, by nature, technical. We offer those details in the technical section and omit some of them from the nontechnical section.

Nontechnical Section

The variable that we measure is the blood oxygen level dependent contrast, or BOLD contrast for short. BOLD is simply the amount of oxygenated blood in a region of the brain. We must use a contrast because there is always oxygenated blood everywhere in the brain. A lack of oxygenated blood in the brain causes death. So, we measure the difference in oxygenated blood during different tasks. Specifically, for our work, we measure the differences between the task of answering a question about a bubble graph and a three-dimensional surface graph. There is always oxygenated blood in all regions of the brain when solving bubble graphs and when solving three-dimensional surface graphs, but there is more in some regions and less in others.

The scanner divides the brain into voxels, which in our case are 3 mm cubes. For each voxel, we measure the BOLD signal every 2.5 seconds. Note that the scanner measures in slices rather than measuring the whole brain at once. It can cause measurement issues if the onset of a task and the onset of a scan occur at the same time because one part of the brain is being measured when the task starts and another part of the brain 2.5 seconds later. We control for this in two ways. First, between each trial we add a random time delay that is uniform between 0 and 2.5 seconds. This is equivalent to starting each scan at a random place in the brain and completely removes correlation between stimulus onset time and when a certain region of the brain is scanned. Second, rather than scan from top to bottom in 2.5 seconds we scan every other slice (e.g., slice 0, 2, 4, 6, …) in 1.25 seconds then scan the other slices (e.g., 1, 3, 5, 7, …) in the next 1.25 seconds. Along with smoothing that we do, this helps remove temporal artifacts based on slice order. Ideally, we would scan slices at random, but the physics of scanners makes that intractable at short scan times of 2.5 seconds.

When neurons do work, they require oxygenated blood to replenish their ability to do more work. However, oxygenated blood does not flow instantly to neurons in a brain region. The oxygen level builds up, peaks at about 6 seconds, then tapers off when demand decreases. This is

called the hemodynamic response function (HRF). Thus, while a task might start at time t1 and continue to time t2, the oxygenated blood flow (which we measure) will begin to ramp up at t1, peak at t1+6s, then begin to taper off from t2 to around t2+6 seconds. Thus, we must convert all dummy variables into variables that ramp up and ramp down with the hemodynamic response function. We do this by convolving our timing with a gamma function.

For example, if we have a screen that is up from 3 seconds to 6 seconds we have a dummy variable that looks like Figure A1. A gamma function with a mean of 6 seconds and a standard deviation of 3 seconds looks like Figure A2. When we convolve the two we get a signal that looks like Figure A3.

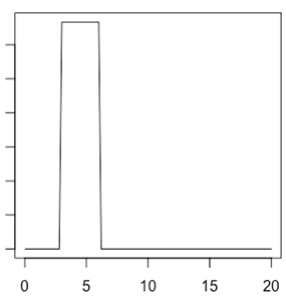


Figure A1. Dummy Variable of Screen Shown from 3–6 Seconds

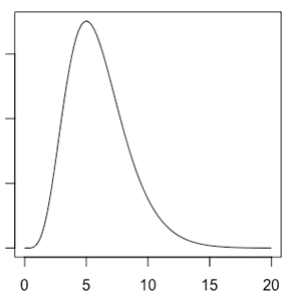


Figure A2. Gamma Function with Mean of 6 Seconds and Standard Deviation of 3 Seconds

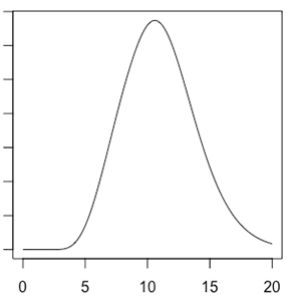


Figure A3. Convolved Function from Representations in Figure A1 and Figure A2

Henceforth we will refer to our timings as dummy variables, but technically they are dummy variables convolved with gamma distributions to account for the hemodynamic response function.

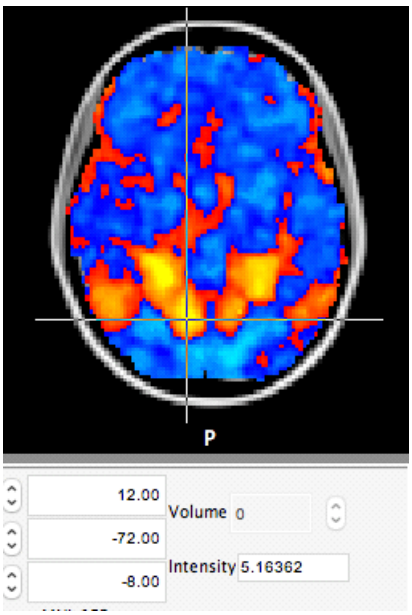
For each voxel, we have an observation of oxygenated blood levels every 2.5 seconds. We take these numeric observations and fit a regression of the form $Y = \mathbf{\beta X} + \text{error}$. Y is the level of activation as measured by the scanner. Note, activation is the numerical value the scanner produces and is proportional to oxygenated blood levels, but it is not oxygenated blood level;it is signal strength from nuclear decay. **X** is a set of vectors of dummy variables measuring which screen the subjects were observing at a specific time (convolved with gamma functions). **β** is a vector of parameters to be estimated. The error term is corrected for autocorrelation using prewhitening (Woolrich et al. 2001).

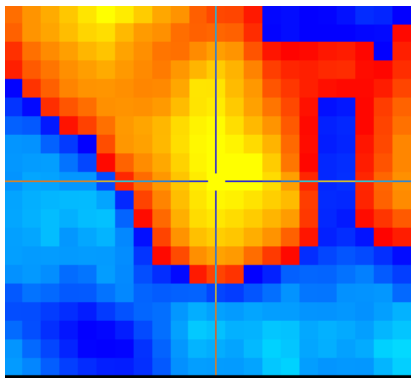
The main model we report has the form

$$\text{Activation}_t = \beta 1 \text{Bubble\_Extraction}_t + \beta 2 \text{Bubble\_Integration}_t + \beta 3 \text{Surface\_Extraction}_t + \beta 4 \text{Surface\_Integration}_t + \beta 5 \text{Fixation}_t$$

In this case Bubble_Extraction means that the subject is viewing the screen where a bubble graph is being paired with an extraction question. The other dummy variables are obvious, except the fixation dummy, which refers to the time between the other screens. Note that this regression is run for each voxel, where there are approximately 50,000 $3mm^3$ voxels in a typical brain.

To test the hypothesis that bubble graphs yield greater responses than three-dimensional surface graphs in a particular voxel, we test the hypothesis $\beta 1 + \beta 2 > \beta 3 + \beta 4$. This yields 50,000 t-tests. We convert the t-values to z-values (i.e., the normal distribution) so that the maps are comparable without having to account for degrees of freedom. It is not reasonable to display the outcomes of each of these tests. Moreover, spatial location is important. Rather than produce a table of coordinates and z-values, we spatially project the z-values onto an image of the brain and use color to denote the z-values. This will give us the image shown in Figure A4, where at voxel 12, -72, -8 (the location of the cross hairs) the t-value is 5.16362. As the color moves from red to yellow, the z-value increases along the positive axis; as the color moves from blue to white, the z-value decreases along the negative axis. Roughly speaking, yellow voxels are significantly positive and light blue voxels are significantly negative.





Full brain z-score map	Zoomed to show individual voxels

Figure A4. Example of z-values Mapped onto Brain Image

There are a few problems with this kind of map. First, it provides too much information, showing both significant and non significant voxels. Second, it obscures the contours of the brain, so it makes identifying structures problematic.

To address these issues, we extract only the significant voxels by specifying some threshold. We use a z-score of 2.3 as the threshold and only display voxels that have values higher than 2.3 (a 1% one-sided p-value). For other voxels, we display the brain contours. This yields an image shown in Figure A5.

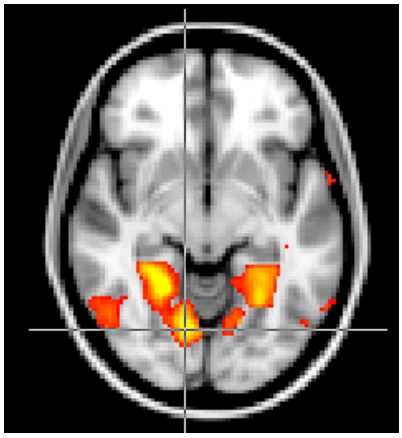


Figure A5. Significant Voxels Using a z-score Threshold of 2.3

There is also a statistical problem of multiple comparisons. With 50,000 tests, by definition we would expect to see 500 significant activations even if there were no effect (assuming an alpha value of 0.01). There are a variety of ways to do this,which we discuss below, but for now we will illustrate the one we use in the paper.

First, we note that the brain is not divided into voxels, the scanner is. Brain structures can and do extend beyond a single voxel, so we expect that genuine activation will occur in contiguous voxels. We can use random field theory to calculate the probability that a contiguous cluster of size N would occur by chance. We then only accept clusters that have a p-value of less than 0.05. We only accept clusters of activation that are so large they are unlikely to have occurred by random chance. This gives us a picture like the one shown in Figure A6.

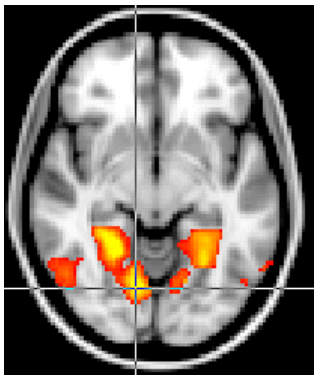


Figure A6. Example of Significant Clusters

Notice that the small clusters fell out. The ones that seem small are the bottoms of clusters that extend upward into the brain. Notice also that this is a scale linking the colors to their associate z-scores. Pure red corresponds to a z-score of 2.3 and pure yellow corresponds to a z-score of 5.0 (or greater). Thus, this image represents the outcomes of two statistical tests, and it shows only the areas of significant activation, in clusters of significant size.

In addition to the image we report a table where the first row looks like Table A1.

Table A1. Example Statistical Report

Cluster	# of Voxels	P-value Cluster	Largest Z-score	MNI152 Coordinates of Largest Z	Region
4	3985	2.82E-14	6.09	24, -46, -8	Right Temporal Occipital Fusiform Cortex

This tells us there is a cluster of 3,985 contiguous voxels. The probability of such a cluster forming by random chance is 2.82E-14. It also details the single voxel where the highest z-score occurs: voxel 24, -46, -8. The z-score in that voxel is 6.09 and is located in the right temporal occipital fusiform cortex.

We also note that this is an average of all subjects' brains. It is worthwhile to read Dimoka (2012) to understand how the brains are mapped onto one another (this is called registration). We plot it on a standard template called the Montreal Neurological Institute (MNI) 152, which is the average of 152 brains scanned at the Montreal Neurological Institute.

Technical Details

The scans were done using a 3-tesla Siemens Skyra scanner. Functional images had a repetition time of 2500 ms, echo time of 30 ms, and flip angle of 90 degrees. 44 transverse slices were acquired in an interleaved fashion. Voxels were $3mm^3$. The number of volumes taken depended on how fast subjects answered but was limited to 35 minutes. High-resolution T1 images with 192 slices in the sagittal plane were also taken with a voxel resolution of .9375 mm × .9375 mm and a slice thickness of .9000 mm.

The functional images were registered to the T1 images using FSL's FLIRT (Jenkinson et al. 2002; Jenkinson and Smith 2001) BBR algorithm (Greve and Fischl 2009). The high resolution images were then registered to the standard 2mm Montreal Neurological Institute 152 brain using

FSL's FNIRT (Andersson et al. 2007a, 2007b) using a warp resolution of 10 mm and a 12 degree of freedom search. Motion correction was done using FSL's MCFLIRT (Jenkinson et al. 2002). B0 unwarping was carried out using an effective echo spacing of 0.58 ms. a change in echo time of 2.46 ms, and a signal loss threshold of 10% (Jenkinson 2003, 2004). Images were spatially smoothed using a Gaussian kernel with a 5mm width. High pass filtering was applied using a least-squares straight-line fitting, with a normal kernel with a sigma of 50 seconds. Non-brain matter was extracted using BET (Smith 2002).

To analyze the fMRI data, a least squares model was fit to the individual level data. Five dummy variables were created for the time periods that corresponded to (1) fixation times, (2) answering in-sample surface graph questions, (3) answering out-of-sample surface graph questions, (4) answering in-sample bubble graph questions, and (5) answering out-of-sample bubble graph questions. In the paper, we also include a continuous variable measuring the visual complexity of each image. For each of the screens subjects viewed, we calculated the visual complexity score (Rosenholtz et al. 2007).[1] We then centered the score and used this as an additional regressor to explain neural activation. Each variable was convolved with a gamma function with a lag of 6 seconds and a standard deviation of 3 seconds to approximate the hemodynamic response function (Woolrich and Jenkinson 2004). We added temporal derivatives to compensate for minor differences in the hemodynamic response function. Statistical parametric maps generated at the first level were all converted into standardized (2mm MNI152) space, and a higher-level random effects analysis was run on the standardized data using FSL's FLAME 1 + 2 (Woolrich Behrens 2004). Images were thresholded at a z-value of 2.3 and then random field theory was used to retain only clusters of activation that were less than 5% likely to have occurred by chance (Worsley 2001). This corrects for multiple comparisons.

Robustness Tests

There are a variety of ways to control for false positives in whole brain activation maps. We use cluster-based thresholding based on FSL's Flame (Woolrich Behrens 2004). FSL's Flame1 was found to have a false positive rate under the null hypothesis close to the predicted false positive rate (Eklund et al. 2016). We use Flame1+2, which runs Flame1 then takes all the values that are close to the margin and preforms Markov chain Monte Carlo simulation to get an enhanced measure on those voxels. Flame1 is relatively fast, but Flame1+2 is computationally intensive, which is presumably why it was not tested. In principle, Flame1+2 should be at least as good as Flame1, but that is currently unknown.

While Flame1 showed good statistical properties, other cluster-based thresholding methods showed inflated false positive rates, and thus alternative methods have been recently proposed (Eklund et al. 2016). The first method is permutation testing (Winkler et al. 2014). Unfortunately, true permutation testing only works for a multi-sample t-test. For a one sample t-test, we use a sign flipping procedure, for which the statistical properties are not consistent (Eklund et al. 2016).

A second method is to use voxel-wise thresholding instead of cluster-based thresholding (Poline et al. 1997). Voxel-wise inference is known to be overly conservative, but it does eliminate the false positive problem. Actually, it causes the false positive problem to manifest in the opposite direction—too few false positives—but it is worth examining.

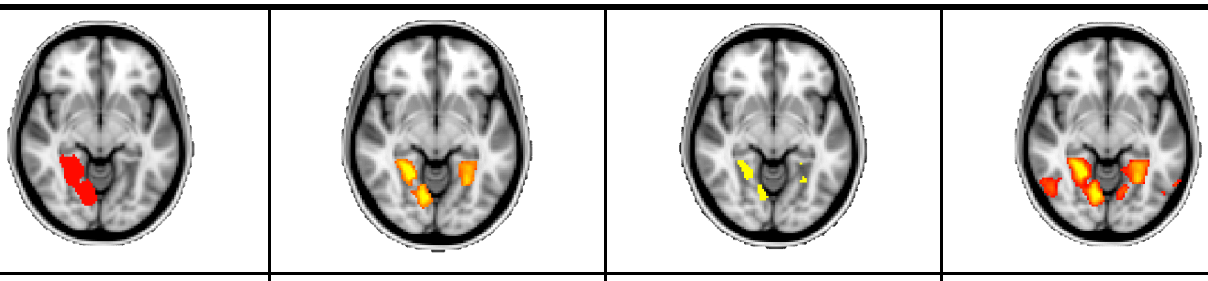
The third method is to increase the cluster defining threshold from 2.3 (one-sided p-value of 0.01) to 3.1 (one-sided p-value of 0.001). It is not clear why this helps, but it is conjectured that the assumptions underlying cluster-based thresholding are more realistic for small clusters, and a higher cluster defining threshold generates smaller clusters (Eklund et al. 2016).

While none of these methods are a panacea for the issues surrounding one sample t-tests, they all produce similar results in this sample, which suggests that our results are robust.

For our permutation test, we used FSL's Randomise (Winkler et al. 2014) to perform 100,000 permutations and a cluster-defining threshold of 2.3. Following Eklund et al. (2016) we use cluster mass thresholding.

The results are shown in Figure A7. Note that Randomise reports the whole cluster p-value rather than individual voxel p-values. Thus, the cluster is a uniform color.

[1]We thank the associate editor for this suggestion.

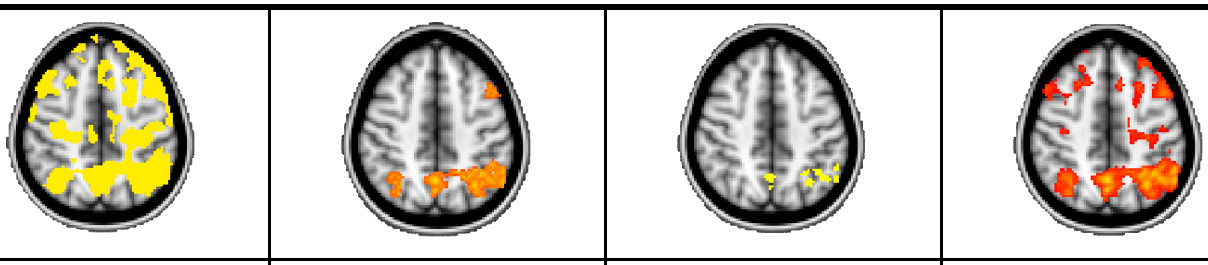


Permutation test	Cluster-defining threshold = 3.1	Voxel-based thresholding	Cluster-defining threshold = 2.3

Z score 2.3 5.0. Left is on the left. All images shown at Z = -8 in MNI 152 coordinates.

Figure A7. Areas of Greater Activation for 3D Surface Graphs than Bubble Graphs

The results are what would be expected. Increasing the cluster defining threshold from 2.3 to 3.1 decreased the size of the clusters, and rendered one of the smaller ones non-significant. Voxel-based thresholding, which is known to be too conservative, removed most of the activation, although some persisted. The sign flipping permutation test rendered all but the right cluster insignificant. Taken together, these results suggest that there is more activation in the ventral stream when viewing three-dimensional surface graphs than when viewing bubble graphs. This activation is stronger on the right side. The results are shown in Figure A8.



Permutation yest	Cluster-defining threshold = 3.1	Voxel-based thresholding	Cluster-defining threshold = 2.3

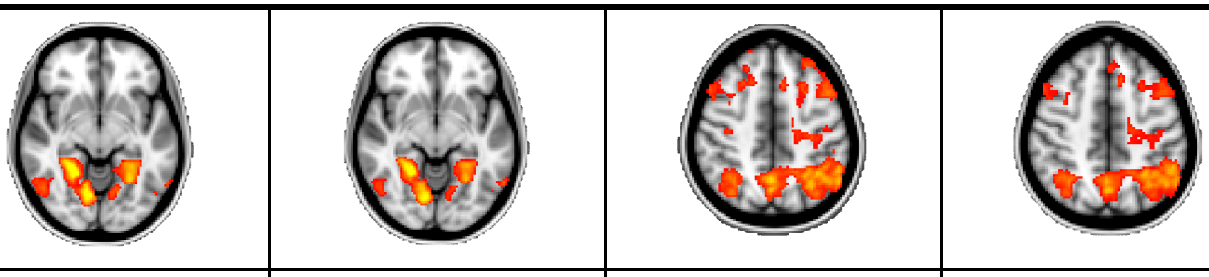
Z scores 2.3 5.0. Left is on the left. All images shown at Z = 46 in MNI 152 coordinates.

Figure A8. Areas of Greater Activation for Bubble Graphs than 3-D Surface Graphs

Again, the increase in the cluster-defining threshold decreased the size of the clusters and rendered some insignificant. Voxel-based thresholding removed all but the most significant clusters. Permutation testing seems to have created one huge cluster spread all over the brain. Overall, these results demonstrate a much wider range of areas activated when viewing bubble graphs than when viewing three-dimensional surface graphs.

Other Tests

Early neuroimaging studies excluded left-handed individuals from studies because of potential lateralization issues. However, this has created problems for scientific learning in an era of large-scale databases of subjects. Moreover, there are ethical issues with excluding any demographic based on inborn differences which are merely correlated with problems, particularly when subjects are compensated. Thus, following developing scientific standards for inclusiveness (Willems et al. 2014), we include left-handed individuals. While we do not propose lateralization hypotheses, for robustness we statistically test the differences between right-handed and left-handed individuals in our study. There were no significant differences between left-handed and right-handed individuals in this study. This could be due to the power of the test. However, a comparison of the right-handed individuals to the whole sample shows similar patterns across the two samples and are shown in Figure A9.

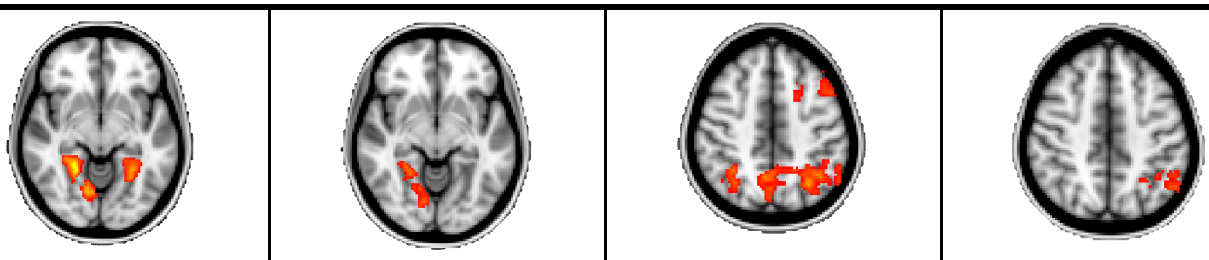


Surface > Bubble full sample	Surface > Bubble right-handed sample	Bubble > Surface full sample	Bubble > Surface right-handed sample

Z score 2.3 5.0. Left is on the left. All images shown at Z = -8 and Z = 46 in MNI 152 coordinates.

Figure A9. Handedness Comparison

In the past, females were also often excluded, which is particularly odd considering females make up more of the human population than males. However, that is typically not done today. But sometimes there are differences between the sexes, so we test for it and find no significant differences between the two groups for any contrast. The group analyses look slightly different, probably because both are underpowered. Even at such low power,both show the same result with activation in the ventral stream for three-dimensional surface graphs and higher in the parietal lobe for bubble graphs. The results are displayed in Figure A10.



Surface > Bubble Male	Surface > Bubble Female	Bubble > Surface Male	Bubble > Surface Female

Z score 2.3 5.0. Left is on the left. All images shown at Z = -8 and Z = 46 in MNI 152 coordinates.

Figure A10. Gender Comparison

For experiment 2 we completed similar analysis and the results were again robust to the different tests.

References

Andersson, J. L., Jenkinson, M., and Smith, S. 2007a. "Non-linear Optimisation," FMRIB Tchnical Report TR07JA1, FMRIB Centre, University of Oxford, Oxford, UK.

Andersson, J. L., Jenkinson, M., and Smith, S. 2007b. "Non-linear Registration, aka Spatial Normalisation," FMRIB Technical Report TR07JA2, FMRIB Centre, of the University of Oxford, Oxford, UK.

Dimoka, A. 2012. "How to Conduct a Functional Magnetic Resonance (fMRI) Study in Social Science Research," *MIS Quarterly* (36:3), pp. 811-840.

Eklund, A., Nichols, T. E., and Knutsson, H. 2016. "Cluster Failure: Why fMRI Inferences for Spatial Extent Have Inflated False-Positive Rates," *Proceedings of the National Academy of Sciences* (113:28), pp. 7900-7905.

Greve, D. N., and Fischl, B. 2009. "Accurate and Robust Brain Image Alignment Using Boundary-Based Registration," *Neuroimage* (48:1), pp. 63-72.

Jenkinson, M. 2003. "Fast, Automated, N-Dimensional Phase-Unwrapping Algorithm," *Magnetic Resonance in Medicine* (49:1), pp 193-197.
Jenkinson, M. 2004. "Improving the Registration of B0-Disorted EPI Images Using Calculated Cost Function Weights," in Proceedings of the 10th International Conference on Human Brain Mapping, Budapest, Hungary.
Jenkinson, M., Bannister, P., Brady, M., and Smith, S. 2002. "Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images," *Neuroimage* (17:2), pp. 825-841.
Jenkinson, M., and Smith, S. 2001. "A Global Optimisation Method for Robust Affine Registration of Brain Images," *Medical Image Analysis* (5:2), pp. 143-156.
Poline, J. B., Worsley, K. J., Evans, A. C., and Friston, K. J. 1997. "Combining Spatial Extent and Peak Intensity to Test for Activations in Functional Imaging," *Neuroimage* (5:2), pp. 83-96.
Rosenholtz, R., Li, Y., and Nakano, L. 2007. "Measuring Visual Clutter," *Journal of Vision* (7:2), pp 1-22.
Smith, S. M. 2002. "Fast Robust Automated Brain Extraction," *Human Brain Mapping* (17:3), pp 143-155.
Willems, R. M., Van der Haegen, L., Fisher, S. E., and Francks, C. 2014. "On the Other Hand: Including Left-Handers in Cognitive Neuroscience and Neurogenetics," *Nature Reviews Neuroscience* (15:3), pp. 193-201.
Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., and Nichols, T. E. 2014. "Permutation Inference for the General Linear Model," *Neuroimage* (92), pp. 381-397.
Woolrich, M. W., Behrens, T. E., Beckmann, C. F., Jenkinson, M., and Smith, S. M. 2004. "Multilevel Linear Modelling for FMRI Group Analysis Using Bayesian Inference," *Neuroimage* (21:4), pp. 1732-1747.
Woolrich, M. W., Jenkinson, M., Brady, J. M., and Smith, S. M. 2004. "Fully Bayesian Spatio-Temporal Modeling of fMRI Data," *IEEE Transactions on Medical Imaging* (23:2), pp. 213-231.
Woolrich, M. W., Ripley, B. D., Brady, M., and Smith, S. M. 2001. "Temporal Autocorrelation in Univariate Linear Modeling of fMRI Data," *Neuroimage* (14:6), pp. 1370-1386.
Worsley, K. 2001. "Statistical Analysis of Activation Images," in *Functional Magnetic Resonance Imaging: An Introduction to Methods,* P. Jezzard, P. M. Matthews, and S. M. Smith (eds.), Oxford, UK: Oxford Scholarship Online., pp. 251-270.