Figure S1

MRI scanner compatibility

This study was a joint venture of the University Medical Center Utrecht (The Netherlands) and the University of Iowa (US). Imaging data were collected using two clinical 3.0 Tesla MRI scanners, both with an 8-channel head coil but from different vendors; Philips Achieva in Utrecht (NL) and Siemens Magneton Trio in Iowa (US). As data were pooled to increase statistical power, the issue of scanner compatibility had to be dealt with, focusing both on scanner hardware and software. The following approach to assess scanner compatibility and to overcome problems resulting from scanner differences was adopted:

First, in the preparation stage of the study, pilot data were collected of three subjects (i.e., researchers GJ and NFR and a research assistant (JZ) on both scanners, acquiring several try-out 3-D anatomical scans and functional time series (using different TE's, TR's and flip angles). Data were tested for homogeneity of signal-to-noise (SNR) and temporal signal-to-noise (tSNR) ratios across sites and vendors. Three pulse sequences were compared: 3D PRESTO, fast 2D-EPI and slow 2D-EPI. The first two scans involved SENSE/GRAPPA (both scanners had the same 8-channel head coil) and yielded different tSNR maps. Moreover, we were not able to make the pulse sequences exactly the same for both scanners. The slow 2D-EPI scan, which uses the 8 channels but does not apply any SENSE or GRAPPA, turned out to be the best in terms of the SNR and tSNR (data not reported) and was selected to proceed with. All scan parameters were the same (TE/TR 35/2000 ms, flip angle 70^{o}, FOV 256 x 256 mm, acquisition matrix 64 x 64, slice thickness 3.6 (plus a 0.4 mm gap), voxelsize 4.0 mm isotropic, 26 slices, scan orientation transaxial for STERN and parallel to the long axis of the hippocampus for PMT). The only difference across scanners was an opposite step-direction of the phase encoding gradient. This caused a difference in the orbitofrontal region, where macro-susceptibility causes some signal loss and signal becomes shifted away from the

brain midline (due to rapid signal decay and a concomitant T2* weighting across K-space), resulting in reduced tSNR locally and thus reduces sensitivity for BOLD-signal change. The step direction could not be made the same for both scanners, so in the lowest slices SNR (and as a consequence to some degree tSNR) is lower in the right hemisphere for Philips and in the left hemisphere for Siemens.

Second, after completion of the study the tSNR maps were quantitatively compared between the two sites to ascertain the expected similarity, by directly comparing tSNR maps for both STERN (working memory) and PMT (associative memory) image data. As tSNR cannot be assumed to be normally distributed (both thermal and physiological noise contribute, which have different distributions), non-parametric statistics were applied, using the SnPM5b toolbox in SPM5 developed by Andrew Holmes and Tom Nichols (see also www.sph.umich.edu/~nichols/**SnPM**/). For each subject's time series, tSNR-images were created based on the preprocessed functional images (i.e. the unwarped, co-registered, normalized and smoothed time series) for each task separately. tSNR-images were entered into a non-parametric two sample T-test in SnPM, using an approximate test of 1000 permutations (default procedure to reduce computation time when the exact test exceeds 5000 permutations), a family-wise-error (FWE) corrected p-value of 0.05, and contrasting tSNR-maps obtained on the US scanner (Siemens) with tSNR-maps obtained on the Dutch scanner (Philips) for both STERN and PMT. The results are shown in Figure A and B, and revealed several areas that showed significant site-related differences in tSNR (as was expected based on the pilot data, due to different step directions), and hence, in activation effect size. Differences are most prominent in the orbito- and ventromedial regions, where the difference in step direction between scanners likely added to susceptibility in these areas to lower signal-to-noise ratios caused by artifacts due to local field distortions in the proximity of the eye sockets and nasal cavities.

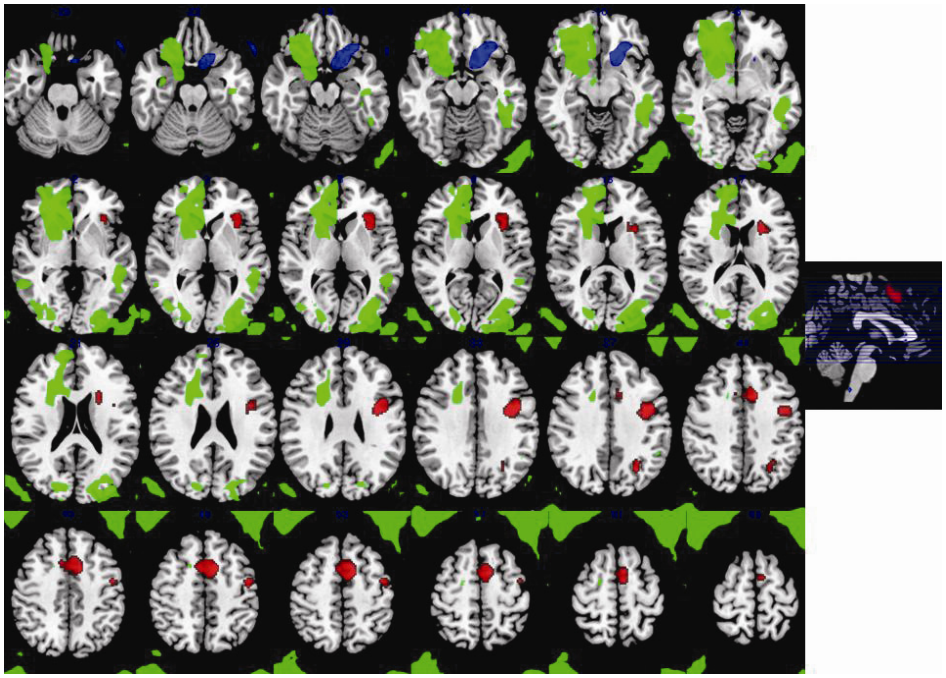

Figure A: Graphic presentation of the results of a nonparametric two sample T-test for the working memory task (STERN) contrasting signal-to-noise (SNR) images from the US scanner (n=23) and NL-scanner (n=21), family-wise-error (FWE)-corrected p-value = 0.05 (corresponding threshold value of T = 5.11). In green, the areas that show significantly higher SNR in US-scans compared to NL-scans. In blue, the areas of the opposite contrast, i.e. significant higher SNR in NL-scans compared to US-scans. In red (superimposed) the regions-of-interest (ROI) for STERN, used in the ROI-analysis. Images are in radiological orientation, i.e. left = right.

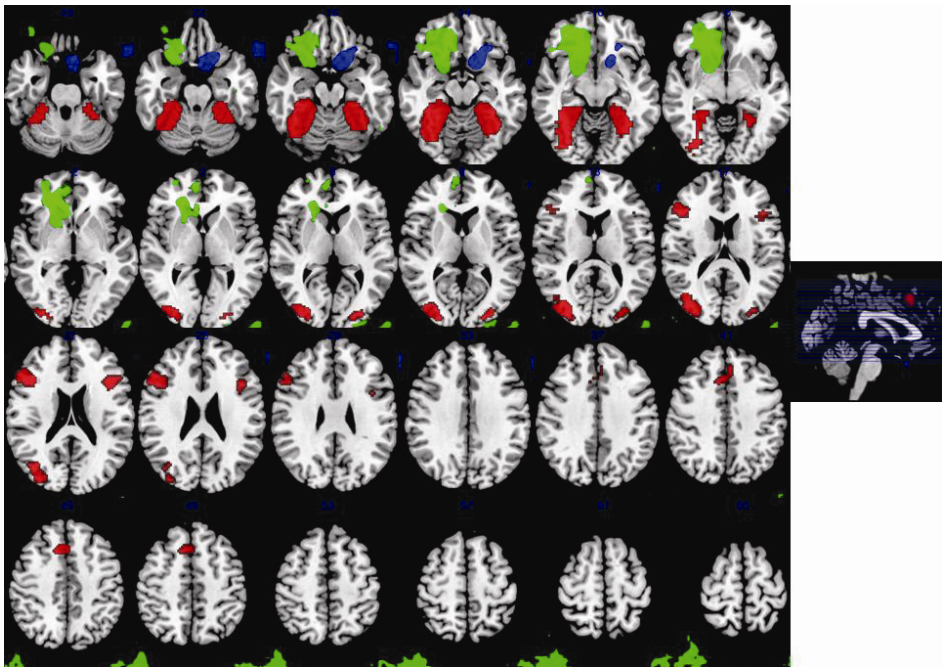

Figure B: Graphic presentation of the results of a nonparametric two sample T-test for the associative memory task (PMT) contrasting signal-to-noise (SNR) images from the US scanner (n=24) and NL-scanner (n=21), family-wise-error (FWE) corrected p-value = 0.05 (corresponding threshold value of T = 5.18). In green, the areas that show significantly higher SNR in US-scans compared to NL-scans. In blue, the areas of the opposite contrast, i.e. significant higher SNR in NL-scans compared to US-scans. In red (superimposed) the regions-of-interest (ROI) for PMT, used in the ROI-analysis. Images are in radiological orientation, i.e. left = right.

Finally, for data analysis, two procedures were adopted to minimize systematic effects of scanner differences on the results: 1) Preprocessing included a smoothing step of 8mm FWHM (i.e. two-times the voxelsize). This is not uncommon in fMRI analyses because of the assumptions of Gaussian Random Field Theory needed for some algorithms, but important in the context of scanner differences is that smoothness equalization (the procedure of smoothing image data from different scanners with scanner-related variability in 'raw' smoothness to a constant FWHM) markedly reduces any possible activation effect size differences between scanners[1]. Second, country (US / NL) was entered as a covariate in the group-wise comparisons on brain activity, as any site-related differences in tSNR would result in a systematic difference in the magnitude of activation (i.e. the contrast-to-noise ratio).

Preprocessing of fMRI data

Imaging data were analyzed using SPM5 (http://www.fil.ion.ucl.ac.uk/spm). Pre-processing included realignment (motion correction) and unwarping, co-registration, normalization and smoothing with an 8 mm (FWHM) Gaussian kernel. For realignment (motion correction) all images were aligned to the first functional image using a rigid body transformation procedure (default option SPM). EPI-images are sensitive to distortions due to magnetic field inhomogeneities, caused by magnetic susceptibility differences in neighboring tissues within the head, resulting in geometrical distortion and signal loss. SPM5 offers a default procedure (unwarping), using algorithms to calculate the geometric distortion with a field mapping sequence, and then compensates for these artifacts by geometrically unwarping the EPI images and by applying cost-function masking in registrations to ignore areas of signal loss. Co-registration involved moving all source images (the anatomical scan and the unwarped realigned EPI images) to a reference image (mean unwarped EPI image) using interpolation methods and affine transformation. An indirect normalization approach was used, co-

registering EPI images from each individual subject with his high resolution T1 anatomical scan (default settings in SPM5), and normalizing this anatomical scan to stereotaxic space (i.e. the MNI305-template). Parameters from this step were used to transfer the EPI images to stereotaxic space. No re-sampling was done (i.e. voxel size used was the size of acquisition (i.e. 4.0 isotropic)). 1st level fMRI time series model specifications included a vector representing the block designs of the tasks, and modeling the four task conditions (i.e. onset and durations (in scans) for CT, PT, NT, and instruction frames for STERN, and AL, SC, RE and instruction frames for PMT) along with a basic set of cosine functions that high-pass filtered the data (cutoff value of 249.6 sec for STERN, and 324 sec for PMT). Cutoff values for high pass filtering were calculated using in-house developed software to determine the optimal high-pass filter taking into account the R-square between the factors included in the model (i.e. minimizing multicollinearity).

References:

1. Friedman L, Glover GH, Krenz D, Magnotta V; FIRST BIRN. Reducing inter-scanner variability of activation in a multicenter fMRI study: role of smoothness equalization. *Neuroimage* 2006;32:1656-1668.