Correspondence of Resting State and Task Brain Networks: Reproducibility and New Insights

Lisa D. Nickerson*[1,2]

[1] Applied Neuroimaging Statistics Lab, McLean Hospital, Belmont, MA, USA
[2] Department of Psychiatry, Harvard Medical School, Harvard University, Boston, MA, USA

Correspondence should be addressed to: LN (lisa_nickerson@hms.harvard.edu)

Abstract: Replication of neuroimaging studies is challenging for many reasons beyond the obvious pressure to publish novel findings. Specialized training is required to analyze neuroimaging data, which contributes to errors in data manipulation and false findings and limits the ability of scientists to reproduce findings. The cost and effort involved with conducting neuroimaging studies places even greater pressure on discovery rather than replication. Fortunately, there is a growing movement to share neuroimaging data for both purposes. Using a different dataset and different methods, a high-value replication of a highly influential neuroimaging study that showed correspondence between the brain networks that are "active" during idle rest and the networks that are engaged during the performance of tasks was conducted. Furthermore, the brain networks that are most strongly activated during task performance are shown for each task studied in the present work, providing new insights into task network function.

In 2009, Smith et al published a paper in the Proceedings of the National Academy of Sciences[1] that showed that the collection of brain networks that are "active" while a person is resting and engaged in idle thought (e.g., resting state networks or RSNs) correspond to the same functional networks used by the brain to perform tasks. This study showed for the first time the extent to which the set of RSNs consistently observable using functional magnetic resonance imaging (fMRI) during rest match the functional networks utilized by the brain during tasks and provided strong supporting evidence to an emerging literature showing that RSNs observable with fMRI were not simply due to non-neural physiological effects. Smith et al has been cited >1,300 times, in the top 1% of highly cited papers in Neuroscience and Behavior (Web of Science), underscoring its importance to the neuroscience and neuroimaging communities. Given the significance of this study, it would be reassuring if the main findings were replicated in an independent study. Replication plays a crucial role in scientific progress, especially given current widespread concerns about reproducibility across disciplines.[2,3] However, neuroimaging studies are difficult to replicate because of the expense involved and pressure to produce novel findings over replication. In the present study, I set out to replicate the main findings of Smith et al. study using new data and different analysis techniques.

The neuroimaging community itself engages in ongoing evaluation of best practices in both data collection and data analysis and has turned up several problems over the last decade, including work by Vul et al. (2009) positing that extremely high correlations between brain activation and personality measures in a large collection of neuroimaging studies arose because of circular analysis practices,[4] work by Bennett et al. (2010) that reported brain activation during a social cognition task in a dead salmon as an illustration of the inadequacy of commonly used corrections for multiple comparisons,[5] and a recent publication by Eklund et al. (2016) that revealed that parametric statistical methods implemented in many common neuroimage analysis software packages are invalid for cluster-wise inference, calling into question findings from a number of fMRI experiments.[6] In parallel, there is growing concern about the reproducibility of published scientific research in general. A recent effort by the Open Science Collaboration (2015) attempted to replicate 100 psychology studies, but succeeded in replicating only 39,[3] and a survey by the journal Nature of 1,576 researchers showed that more than 70% of researchers have tried and failed to reproduce another scientist's experiment and more than half reported failing to reproduce their own experiments.[7] Various strategies have been proposed to improve the reliability and efficiency of scientific research[8], including the use of registered reports adopted by more than 40 journals to date to both enhance and incentivize reproducibility.[9]

Regarding the Smith et al. (2009) study, given the high impact of the paper on the neuroimaging community, it would be reassuring to see the main findings replicated in an independent study. However, the Smith et al. study is challenging to replicate because one needs imaging data collected during the performance of many different tasks covering myriad behavioral domains to

identify the full repertoire of task networks. Smith et al. were able to utilize the BrainMap database[10], the largest database of human brain activation study results obtained using neuroimaging techniques, in combination with resting state fMRI data to investigate the links between RSNs and task activation networks. Using the BrainMap database allowed them to pool together imaging results reported from more than 1600 published studies of brain function over a wide range of experimental paradigms to derive the full repertoire of the brain's functional task networks. Recently, the Washington University-University of Minnesota (Wu-Min) Human Connectome Project (HCP),[11] an initiative to map the human brain connectome, collected neuroimaging data in hundreds of participants and they have made the data publicly available (https://www.humanconnectome.org). The HCP collected task functional MRI (t-fMRI) during seven different tasks chosen to cover multiple domains of function and optimized to activate as many functional nodes, or regions of the brain, as possible.[12] In addition, the HCP collected r-fMRI data in each participant,[13] thus the HCP data present a unique opportunity to conduct a high-quality replication study of the Smith et al. study. In the process of replicating the Smith et al. study, a simple methodological approach was conceived that has great potential for studying network activity during task performance, revealing new insights into brain network activity during the different HCP tasks.

Results

Smith et al. applied independent component analysis (ICA) to the BrainMap data and to resting state fMRI data to derive 20 independent components from each dataset representing both large-scale brain networks and artifact-related effects. He then identified task networks that corresponded to commonly observed RSNs[14,15] using spatial cross-correlation (Smith et al. Figure 1). In the present study, these Smith et al. main findings were highly reproducible using the HCP task and resting state data. The spatial cross correlations between the HCP task networks and Smith's BrainMap networks ranged from r = 0.26-0.74, with the minimum r=0.26 being highly significant ($p=4x10^{-4}$, corrected). The spatial cross correlations between the HCP task and HCP RSNs ranged from 0.44-0.81, with the minimum r=0.44 being highly significant ($p<4x10^{-4}$, corrected). The correspondence between the HCP task and HCP RSNs is much higher than the correspondence between the BrainMap networks and the RSNs reported in Smith et al. ($r=0.26$, $p<1x10^{-5}$, corrected), although both are highly significant. This is due to the different nature of the task data that was used for each study (BrainMap pseudo-activation maps versus actual task activation Z-stat maps in the present study).

Figure 1 shows all four sets of network maps, the HCP RSN and HCP task networks from the present study and the RSN and BrainMap networks from Smith et al. study (their Figure 1). The Smith et al. maps displayed alongside the HCP maps were constructed from their data files available at http://fsl.fmrib.ox.ac.uk/analysis/brainmap+rsns/. For Figure 1, all ICA Z-statistic maps were thresholded with Z=3.0 (the same as in Smith et al. Figure 1) and red-yellow and blue-light blue colors indicate the network (positive and negative values, respectively), overlaid onto the MNI standard brain image. The Z=3.0 threshold used in Smith et al. is based on an alternative hypothesis testing approach which applies a Gaussian-Gamma mixture model to the independent component spatial maps to determine the threshold for each map.[16] In this case, a threshold of p=0.5 will achieve an equal probability of obtaining a false positive or a false negative (e.g., of a given voxel being in the background signal or the IC signal). Mixture modeling to threshold ICA maps is used to address the fact that spatial maps derived from the fixed-point iteration ICA algorithm in FSL MELODIC (and from Infomax or other similar algorithms) are optimized for maximal non-Gaussianity of the distribution of spatial intensities. In this case, simple transformation of ICA maps to Z scores and subsequent thresholding will not provide control of the false-positive rate. Using mixture modeling allows for such control and Z = 3 is approximately the average Z value that one obtains from thresholding a typical group ICA

spatial map at p=0.5 when 20 components have been estimated (e.g., if more components are estimated, the threshold will increase due to reduced residuals). Applying a mixture model to the ICA maps, with p=0.5, in the present study gave an average value of Z ~ 2.5 so the threshold used to create the figure corresponds to a slightly greater probability of a voxel being signal rather than noise for the HCP-derived RSN and task network maps.

Correspondence between HCP and Smith et al. networks were similarly reproducible for the 70-component analysis, which results in a finer parcellation of the brain into sub-networks as compared with 20 estimated components (Smith et al. Figure 3; eight occipital and two sensorimotor networks). Spatial cross correlations between the RSNs from Smith et al. and the HCP RSNs ranged from 0.36 – 0.69 (results not shown).

It is not possible to reproduce how strongly each component in Figure 1 relates to the 66 possible behavioral domains as shown in Smith et al. (Figure 2) due to the limited number of tasks in the present study. However, it is possible to determine the magnitude of network activation during each task component. This was not possible in Smith et al. because of the constraints on the BrainMap data, that only the location in standard space of an activation blob is encoded in the BrainMap database with no spatial extent or magnitude information. To address the spatial extent issue, Smith created pseudo-activation maps by placing Gaussian blobs at each activation point. In addition, Smith et al. estimated a measure related to how often a spatial pattern was observed during tasks for a given behavioral domain. While this metric is somewhat informative of which networks are engaged during which tasks, there is no information about how strong a network may be activated during a particular task. In the present study, the strength of activation of each network for each task component, covering 29 different behavioral domains, was computed as follows. The set of spatial maps from the ICA of the HCP task Z-stat maps (contrast maps of brain activation during different task conditions) were used in a multiple spatial regression against each set of task Z-stat maps to extract a set of "subject courses" for that particular task component. These subject courses reflect the strength of the particular network activation in each subject, and can be averaged over all subjects for a particular task to compute the average magnitude of activation of each network during a task. This approach is different than the approach used to compute the matrix values shown in Smith et al. Figure 2. By using actual task Z-stat maps and the multiple spatial regression approach in the present study, the magnitude of the activation is preserved giving results that are more easily interpretable than the measure of strength derived by Smith et al.

Network activation magnitudes are displayed in Figure 2. No inference was done on this matrix as it is meant to parallel directly the qualitative results shown in Smith et al. Figure 2. However, we consider the N-back working memory task network activation strengths to demonstrate the utility of the approach. The executive control network, LFPN, RFPN, and cerebellum all showed a statistically significant activation during the 2 Back relative to 0 Back condition (p = 0.0001, taking into account family structure, corrected for number of networks and contrasts (e.g., 2-sided tests, with differences in reaction times and accuracy between 2 Back and 0 Back included as covariates of interest in the general linear model)). All other networks showed statistically significant suppression during 2 Back relative to 0 Back (p<0.02, corrected). Moreover, increasing network activation magnitude was associated with increases in reaction times during 2 Back relative to 0 Back (p=0.0001, corrected), for the executive control network, RFPN, and LFPN, suggesting that as the working memory load increases, these networks increase their activity. The executive control network is comprised of dorsal anterior cingulate cortex (dACC), medial superior frontal cortex (msFC), and bilateral anterior insula/frontal operculum. This network has also been referred to as the cingulo-opercular network and has been shown to be a core system for the implementation of task sets that provides stable "set-maintenance" over entire task epochs over a variety of tasks.[17,18] In the present study, this network shows increased

activity during nearly all of the tasks (relative to their lower level control conditions), which is consistent with being the core system as described by Dosenbach et al. (2006). The two lateralized left and right fronto-parietal networks, LFPN and RFPN, respectively are posited by Dosenbach et al. (2008) to be control networks (or a single network in their study, but split in our study and in other ICA-based fMRI studies) that potentially initiate and adjust control on a trial-to-trial basis and respond to events that carry performance feedback information. The prefrontal and parietal regions in all three of these networks have been implicated in previous studies of working memory.[19] Figure 2 results also suggest that the DMN is suppressed during most tasks, occipital networks are activated during tasks in which there are visual stimuli, but not when the tasks involved auditory stimuli (math and story blocks of the language task) or the motor tasks. Instead, the story blocks of the language task activated the auditory network while the motor tasks activated the sensorimotor network. However, these activation magnitudes were not assessed for statistical significance.

Discussion

The main findings of correspondence between the large-scale brain networks observed at rest and the brain networks that are actively utilized by the brain during task performance that were presented in the seminal study by Smith et al. were found to be highly replicable using new data and different analysis techniques. Thus, the present study lends additional critical support to the interpretation of RSNs measured using fMRI as reflecting neural activity, not just hemodynamics and other physiological processes, and as corresponding to task activation networks. Furthermore, the simple approach for deriving the task activation strength of each network by regressing ICA network maps against task activation maps can be used to gain a clearer understanding of brain network activity during task performance.

It was shown that the executive control network, RFPN, and LFPN showed increased activation with increasing reaction times (suggesting greater cognitive effort) during the working memory 2 Back relative to 0 Back condition, which is consistent with previous studies showing activation of regions in prefrontal and parietal cortex with increasing working memory load. Also, in Figure 2, default mode network (DMN) activity is suppressed during most of the tasks, with stronger suppression for those tasks with presumably greater cognitive load. Several studies have shown that the magnitude of BOLD activity within the DMN is related to task-load during brain activation.[20,21] The executive control network also showed increased activation during most of the task conditions, supporting the role of this network as a core system for task set maintenance. Thus, the present findings are reasonable and consistent with current knowledge of DMN and executive control network function. However, there is a much more subtle point about Figure 2. Namely, all of the networks (represented in the IC maps) that are engaged during a particular task are identified with this analysis approach, and the activity of a specific task network during each task component is determined with all other network activity "partialled out" by virtue of the multiple spatial regression of all network maps against the subject activation maps. This means that the activity in each network can be separately assessed – and that the specific set of networks that are engaged during a given task can be determined. This is in contrast to the activation maps from a standard voxel-wise general linear model analysis. E.g., activation maps resulting from a first-level voxel-wise GLM analysis shows all of the brain regions that are activated during the task aggregated together into a single map, and the map can include regions that constitute different networks, thus the specific networks are not differentiated. For example, consider again the working memory 2 Back vs 0 Back contrast. In Barch et al. (2013), the group activation map obtained using a multi-level general linear model shows deactivations in medial prefrontal cortex and auditory regions (Figure 3 in Barch et al.). It is unknown whether the activated regions represent one or more networks in the aggregate activation map that results from the voxel-wise GLM analysis. Using the analysis approach in the present study, network

activity can be disentangled, and we conclude that executive control, LFPN, and RFPN are all activated during the 2 Back relative to 0 Back. The auditory network is also strongly suppressed during this condition (and during many other conditions, possibly due to a need to block out distracting scanner noise). Thus, using the analysis approach developed in the present study allows for each network that may be activated, or suppressed, to be studied separately from other networks and represents a new approach for probing brain network performance during tasks.

The fMRI study designs used for the HCP tasks employed a "subtraction paradigm" that relies on using two task conditions that differ along a single dimension. The difference between the two conditions, e.g., subtraction of activation maps from each task, isolates activation related to the behavioral process of interest while removing activity related to any common processes. Inspection of Figure 2 shows that for many of the tasks, brain networks with similar activity during two task conditions show no activity for the subtraction. In addition, some networks show greater suppression of activity in one condition versus the other. For example, RFPN activity is suppressed for Math (versus baseline), but more so for Story (versus baseline), such that the contrast Math-Story shows a positive network activity difference, even though the network is actually suppressed to different degrees during each task component. The RFPN is associated with cognitive processing, including reasoning, attention and memory.[22] Moreover, the results show that the networks that are more strongly activated during Math versus Story are the executive control network and the LFPN, which is implicated in cognition/language paradigms.[22] In this case, RFPN activity is suppressed in both conditions, but more so for Story than Math, while there is greater activation in ECN and LFPN during Math than Story.

In conclusion, the present study shows that the highly-cited Smith et al. (2009) neuroimaging study is replicable and demonstrates that multivariate spatial regression of network template maps, such as those from the Smith et al. study or from the present study (contact the author to obtain the maps from this study), against the set of subject parameter estimate maps can be used to investigate brain network activation during task performance. By including all network maps in the multivariate regression against the 4D data, the resulting subject loadings for a given network will have other network effects "partialled out", thus reflecting network activity disentangled from other network activity during a task.

Methods

I. Human Connectome Project

The second major release of the HCP data collected in 500 healthy adults (aged 22-35) was used for the current study. Individuals with severe neurodevelopmental disorders, neuropsychicatric disorders, or neurologic disorders, or with illnesses such as diabetes and high blood pressure, were excluded from the HCP study. MRI scanning was done using a customized 3T Siemens Connectome Skyra using a standard 32-channel Siemens receive head coil and a body transmission coil. T1-weighted high resolution structural images acquired using a 3D MPRAGE sequence with 0.7 mm isotropic resolution (FOV = 224 mm, matrix = 320, 256 sagittal slices, TR = 2400 ms, TE = 2.14 ms, TI = 1000 ms, FA = 8°) were used in the HCP minimal pre-processing pipelines to register functional MRI data to a standard brain space. Resting state fMRI data were collected using gradient-echo echo-planar imaging (EPI) with 2.0 mm isotropic resolution (FOV=208x180 mm, matrix =104x90, 72 slices, TR = 720 ms, TE = 33.1 ms, FA = 52°, multi-band factor = 8, 1200 frames, ~15 min/run). Task fMRI data were collected using the same scanning sequence as the resting state fMRI data, although the number of frames per run (with 2 runs/task) varied from task to task. Runs with left-right and right-left phase encoding were done for both resting state and task fMRI to correct for EPI distortions.

II. Identification of Task Networks from HCP Task fMRI Data

Task fMRI data were utilized from seven different tasks: emotion processing, incentive processing/gambling, language, motor, relational processing, social cognition, and working memory. I used volumetric outcomes from the minimal pre-processing pipelines developed by the HCP[23]. For the HCP minimal pre-processing pipeline, the task fMRI data for each subject underwent corrections for gradient distortions, subject motion, and echo-planar imaging (EPI) distortions, and were also registered to the subject's high-resolution T1-weighted MRI. All corrections and the transformation of the fMRI data to MNI standard space (via non-linear transformation of the subject's T1-weighted structural MRI into MNI standard space) were implemented in a single resampling step using the transforms for each registration step (fMRI to T1 and T1 to MNI) and the distortion corrections.

First-level statistical modeling was also implemented by the HCP. The pre-processed fMRI timeseries at each voxel (or spatial location) in the task fMRI data was fit with a general linear model (GLM). Regressors that modeled the brain's fMRI signal in response to the task conditions were included in the model. 3D spatial maps (e.g., one value per voxel in the brain) of contrasts of the parameter estimates (COPEs) were computed corresponding to the average activation during each task component and to differences in activation between different task components (e.g., subtraction images). Notably, these COPE maps reflect the magnitude of the brain activation between two task conditions.

Twenty-five COPEs were selected from across tasks to be fed into a group independent component analysis (ICA) to identify task networks. Table 1 describes each task COPE, with 3-5 COPEs per task.

The sample size for each task is as follows:

- Emotion Processing. 452 subjects, 3 COPES, 1,356 total images
- Incentive Processing. 449 subjects, 3 COPES, 1,347 total images
- Language. 433 subjects, 3 COPES, 1,299 total images
- Motor. 415 subjects, 5 COPES, 2,075 total images
- Relational Processing. 435 subjects, 3 COPES, 1,305 total images
- Social Cognition. 452 subjects, 3 COPES, 1,356 total images
- Working Memory. 411 subjects, 5 COPES, 2,055 total images

The number of subjects varies for each task because participants with greater than 2 mm of motion (maximum absolute root mean square) in any task run led to exclusion of their COPE map from the analysis. E.g., two runs were done for each task (with different phase encoding directions to correct for EPI distortions), with average COPE maps being calculated using a second-level GLM. These average COPE maps are used in the present analysis.

All 10,793 COPE maps were fed into a group ICA conducted using FSL MELODIC[16]. Two group ICAs were conducted, one with twenty and one with seventy components estimated from the group ICA, the same as in the Smith et al. study, to identify large-scale brain networks and to do a finer parcellation, respectively. The spatial independent component maps were thresholded using a Gaussian-Gamma mixture model with p=0.5 such that equal weight was given to obtaining either a false positive or a false negative in the spatial map. Note that Smith et al., constructed 7,342 activation-peak images (pseudo-brain activation maps constructed by filling an empty image with points corresponding to reported standard space coordinates of statistically significant local maxima in the activation maps from the original study that are archived in the BrainMap database, then convolving these points with a Gaussian kernel to mimic spatial

extent), which were submitted to a group ICA. Thus, the number of maps used in the present study is the same order of magnitude as the number of maps used in the Smith et al. study.

Task activation networks were identified corresponding to those from Smith et al. (e.g., from the ICA of the BrainMap data) by visual inspection and spatial cross-correlation. Significance of cross-correlations was determined as follows. The corrected p-value was computed based on a Bonferroni correction for the number of possible paired comparisons (400 for HCP task vs Smith task; 400 for HCP task vs HCP rest) and a correction for the spatial degrees of freedom using Gaussian random field theory and an empirical smoothness estimation (average number of resels = 322 for HCP task spatial maps, which was lower than both the average for the BrainMap task maps (2143 resels) and the average for the HCP RSNs (375)). For example, the correlation probability for r = 0.26 with 322 degrees of freedom is $p=1x10^{-6}$ (one-sided), multiplying by this value by 400 gives $p=4x10^{-4}$ corrected. See Smith et al. for more details.

The activation magnitudes for each network during each task component were computed as follows. Multiple spatial regression of the HCP dimensionality 20 ICA maps (e.g., all 20 maps together) against the 4D file of COPE maps, e.g., with one COPE map for each subject concatenated across all N subjects, was done for each task component. For each COPE file, the resulting regression parameters are a "subject-series" of loadings (one per subject) that were averaged together to give the average network activity during that component of the task. The regression parameters are Z-statistics since the COPE Z-stat maps from the HCP first-level analysis were used for the analysis. The values in Figure 2 were computed as the average of the subject-series of loadings for each COPE. To assess the activated networks during the 2 Back (2B) vs 0 Back (0B) contrast and the relationships with change in accuracy (0B-2B) and change in reaction time (2B-0B), a general linear model was implemented with change in %accuracy and change in reaction times as covariates using PALM (Permutation Analysis of Linear Models)[24] to take into account the family structure of the HCP data, and to correct for the number of networks and contrasts. Exchangeability blocks were determined that captured family structure to determine acceptable permutations, and 10,000 permutations were done.

III. Identification of Resting State Networks from HCP Resting State fMRI Data

RSNs were identified from the HCP data using outcomes from the minimal pre-processing pipeline of the resting state fMRI data that were provided by the HCP.[13] Minimal pre-processing of resting state fMRI data included corrections for spatial distortions caused by gradient nonlinearities, head motion, B_0 distortion, denoising using FSL FIX[25] and registration to the T1-weighted structural image. All transforms were concatenated together with the T1 to MNI standard space transformation and applied to the resting state fMRI data in a single resampling step to register the corrected fMRI data to MNI standard space. fMRI data were also temporally filtered with a high pass filter and then each subject's fMRI data was analyzed using spatial ICA, with the MELODIC algorithm estimating the number of components. In the HCP framework, these ICA maps are used to denoise the fMRI data prior to any subsequent resting state analyses. In the present analysis, these single-subject ICA maps, from 20 participants, were fed together into a group ICA using FSL MELODIC to identify the collection of RSNs that were common to the group of subjects. The single-subject ICA maps were used for the group ICA instead of the original minimally-preprocessed resting state fMRI data simply to reduce computational load, which is much greater with HCP data due to the extremely high spatial and temporal resolution of the fMRI data (2x2x2 mm^3 with 0.75 second sampling intervals, 15 minutes/run, 1200 volumes/run). As an aside, this is the same order of magnitude of participants in the original study by Smith et al., which utilized resting state fMRI data from 36 participants.

Two ICAs were done, with the number of components fixed to twenty and seventy (as in Smith et al.) and RSNs that corresponded to the RSNs in Smith et al. were identified by visual inspection and spatial cross-correlation. For the 20-component ICA, the spatial cross-correlations between the HCP RSNs and the Smith RSNs (for 10 networks shown in Smith Figure 1) ranged from 0.43-0.74 (except for the cerebellum network which was cut off in Smith's data, but fully covered in the HCP data, resulting in a spatial cross correlation of 0.33). The spatial cross-correlations between HCP RSNs and HCP task networks (Figure 1) ranged from 0.44-0.81. The minimum spatial correlation of 0.44 is even greater than the minimum reported in Smith et al., $r=0.25$, and so is even more highly significant than $p<5x10^{-6}$ (although I did not estimate the actual p value since it's not necessary given the high level of significance).

Acknowledgements

This work was supported by the National Institutes of Health (LDN: DA037265, AA024565).

Data were provided by the Human Connectome Project WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil, 1U54MH091657). The Human Connectome Project was supported by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

Table 1

Task	HCP Cope #	Behavioral Domain
Emotion Processing	1	*Angry/Fearful Faces* (which of two faces at the bottom of the screen match face at top of screen; faces are angry or fearful)
	2	*Shapes* (which of two shapes at the bottom of the screen match shape at top of screen)
	3	*Angry/Fearful Faces – Shapes*
Incentive Processing/Gambling	1	*Punishment* (lose money when guessing the number on a mystery card)
	2	*Reward* (win money when guessing the number on a mystery card)
	3	*Punishment – Reward*
Language	1	*Math Problems* (auditory presentation of math problems, requires subjects to complete addition/subtraction problems)
	2	*Stories* (auditory presentation of stories, participants answer questions about the story)
	3	*Math Problems – Stories*
Motor	2	*Squeeze Left Toes*
	3	*Tap Left Fingers*
	4	*Squeeze Right Toes*
	5	*Tap Right Fingers*
	6	*Move Tongue*
Relational Processing	1	*Match Condition* (does single object at bottom of screen match either of two objects at top of screen on either shape or texture?)
	2	*Relational Condition* (does bottom pair of objects vary on the same dimension (shape or texture) as top pair of objects?)
	4	*Relational – Match*
Social Cognition (Theory of Mind)	1	*Random* (participants judge video clips of objects to be either interacting in some way or moving randomly; in this condition, objects are moving randomly)
	2	*Theory of Mind:ToM* (participants judge video clips of objects to be either interacting in some way or moving randomly; in this condition, objects are interacting)
	6	*ToM – Random*
Working Memory*	11	*2 Back – 0 Back* (blocks of trials consisting of pictures, faces, places, tools, and body parts, with ½ of blocks using 2-back working memory task and ½ blocks using 0-back working memory task being contrasted)
	5	*Body Parts (0 Back blocks) vs. Baseline*
	6	*Faces (0 Back blocks) vs. Baseline*
	7	*Places (0 Back blocks) vs. Baseline*
	8	*Tools (0 Back blocks) vs. Baseline*
	1	*Body Parts (2 Back blocks) vs. Baseline*
	2	*Faces (2 Back blocks) vs. Baseline*
	3	*Places (2 Back blocks) vs. Baseline*
	4	*Tools (2 Back blocks) vs. Baseline*

**For the ICA, working memory contrasts that averaged 0 Back and 2 Back blocks together for each stimulus type were used (COPEs 15-18), but the COPEs for 0 Back and 2 Back blocks relative to baseline were used to analyze network activation to facilitate interpretability.*

Figures

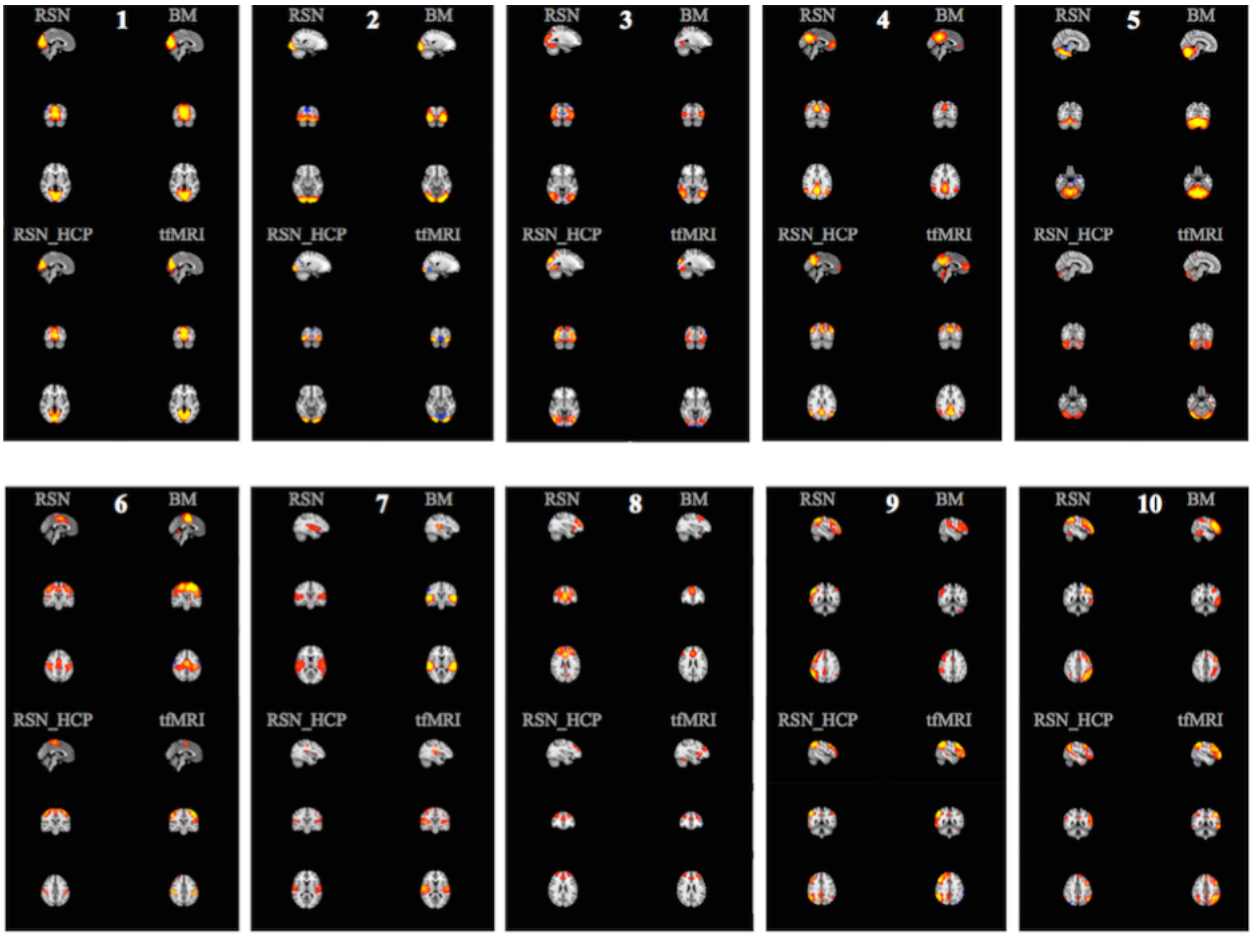

Figure 1. RSNs and BrainMap-derived task networks (top left and right of each panel, respectively) for ten networks reported in Smith et al. (2009) and corresponding RSNs and task networks derived using HCP data (bottom left and right of each panel, respectively). Networks are shown in red-yellow/blue-light blue, thresholded Z>3. Networks are: 1) medial visual, 2) occipital pole, 3) lateral occipital, 4) default mode, 5) cerebellum, 6) sensorimotor, 7) auditory, 8) executive control, 9) right frontoparietal, 10) left frontoparietal.

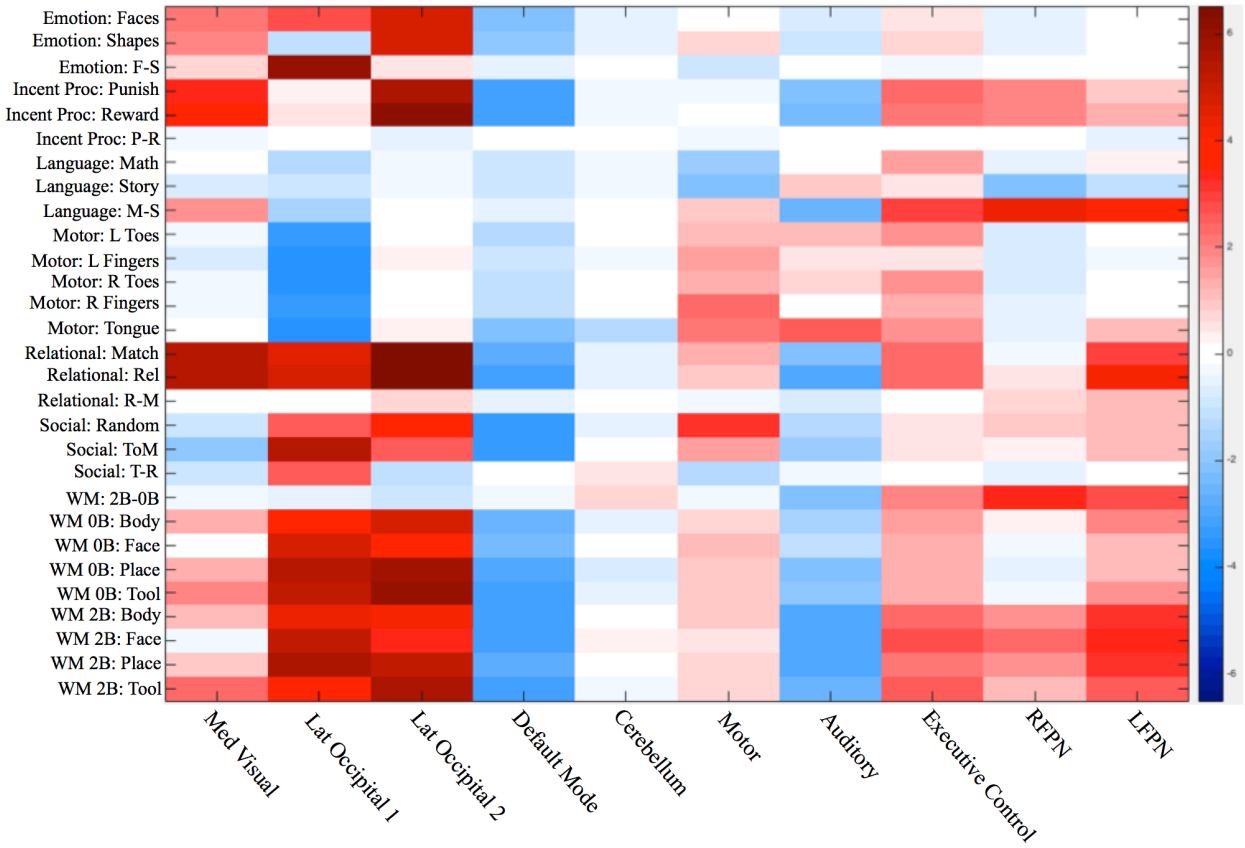

Figure 2: Magnitude of activation (Z-value) of each network during each task component.

References

1. Smith, S. M. *et al.* Correspondence of the brain ' s functional architecture during activation and rest. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 13040–13045 (2009).
2. Ioannidis, J. P. A. Why most published research findings are false. *PLoS Med.* **2,** 0696–0701 (2005).
3. Open Science Collaboration. Estimating the reproducibility of psychological science. *Science (80-.).* **349,** aac4716-aac4716 (2015).
4. Vul, E., Harris, C., Winkielman, P. & Pashler, H. Puzzlingly High Correlations in fMRI Studies of Emotion, Personality, and Social Cognition. *Psychol. Sci.* **4,** 274–290 (2009).
5. Bennett, C. M., Baird, A., Miller, M. B. & Wolfrod, G. L. Neural correlates of interspieces perspective taking in the post-mortem Atlantic Salmon: An argument for multiple comparisons correction. *Hum. Brain Mapp.* **1,** 1995 (2009).
6. Eklund, A., Nichols, T. E. & Knutsson, H. Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proc. Natl. Acad. Sci.* 201602413 (2016). doi:10.1073/pnas.1602413113
7. Baker, M. First results from psychology's largest reproducibility test. *Nature* 1–4 (2015). doi:10.1038/nature.2015.17433
8. Munafò, M. R. *et al.* PERSPECTIVE A manifesto for reproducible science. *Nat. Publ. Gr.* **1,** 1–9 (2017).
9. Promoting reproducibility with registered reports. *Nat. Hum. Behav.* **1,** 34 (2017).
10. Fox, P. T. *et al.* BrainMap taxonomy of experimental design: Description and evaluation. *Hum. Brain Mapp.* **25,** 185–198 (2005).
11. Van Essen, D. C. *et al.* The WU-Minn Human Connectome Project: An overview. *Neuroimage* **80,** 62–79 (2013).
12. Barch, D. M. *et al.* Function in the human connectome: Task-fMRI and individual differences in behavior. *Neuroimage* **80,** 169–189 (2013).
13. Smith, S. M. *et al.* Resting-state fMRI in the Human Connectome Project. *Neuroimage* **80,** 144–168 (2013).
14. Beckmann, C. F., DeLuca, M., Devlin, J. T. & Smith, S. M. Investigations into resting-state connectivity using independent component analysis. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **360,** 1001–13 (2005).
15. Damoiseaux, J. S. *et al.* Consistent resting-state networks. (2006).
16. Beckmann, C. F. & Smith, S. M. Probabilistic ICA for fMRI. **23,** 137–152 (2004).
17. Dosenbach, N. U. F. *et al.* A Core System for the Implementation of Task Sets. *Neuron* **50,** 799–812 (2006).
18. Dosenbach, N. U. F., Fair, D. a., Cohen, A. L., Schlaggar, B. L. & Petersen, S. E. A dual-networks architecture of top-down control. *Trends Cogn. Sci.* **12,** 99–105 (2008).
19. Owen, A. M., McMillan, K. M., Laird, A. R. & Bullmore, E. N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Hum. Brain Mapp.* **25,** 46–59 (2005).
20. Singh, K. D. & Fawcett, I. P. Transient and linearly graded deactivation of the human default-mode network by a visual detection task. *Neuroimage* **41,** 100–112 (2008).
21. Mckiernan, K. A., Kaufman, J. N., Kucera-thompson, J. & Binder, J. R. A Parametric Manipulation of Factors Affecting Task-induced Deactivation in Functional Neuroimaging. *J. Cogn. Neurosci.* **15,** 394–408 (2003).
22. Laird, A. R. *et al.* Behavioral interpretations of intrinsic connectivity networks. *J. Cogn. Neurosci.* **23,** 4022–37 (2011).
23. Glasser, M. F. *et al.* The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage* **80,** 105–124 (2013).
24. Winkler, A. M., Webster, M. A., Vidaurre, D., Nichols, T. E. & Smith, S. M. Multi-level block permutation. *Neuroimage* (2015). doi:10.1016/j.neuroimage.2015.05.092

25. Salimi-Khorshidi, G. *et al.* Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers. *Neuroimage* **90,** 449–468 (2014).