Published in final edited form as:
Psychophysiology. 2012 July ; 49(7): 873–884. doi:10.1111/j.1469-8986.2012.01382.x.

Test-retest Reliability of an fMRI Paradigm for Studies of Cardiovascular Reactivity

Lei K. Sheu, **J. Richard Jennings**, and **Peter J. Gianaros**
University of Pittsburgh

Abstract

We examined the reliability of measures of fMRI, subjective, and cardiovascular reactions to standardized versions of a Stroop color-word task and a multi-source interference task. A sample of 14 men and 12 women (30–49 years old) completed the tasks on two occasions, separated by a median of 88 days. The reliability of fMRI BOLD signal changes in brain areas engaged by the tasks was moderate, and aggregating fMRI BOLD signal changes across the tasks improved test-retest reliability metrics. These metrics included voxel-wise intraclass correlation coefficients (ICCs) and overlap ratio statistics. Task-aggregated ratings of subjective arousal, valence, and control, as well as cardiovascular reactions evoked by the tasks showed ICCs of 0.57 to 0.87 (*p*s < 0.001), indicating moderate-to-strong reliability. These findings support using these tasks as a battery for fMRI studies of cardiovascular reactivity.

Keywords

fMRI; cardiovascular reactivity; test-retest reliability

Functional neuroimaging studies often seek to characterize patterns of brain activity among individuals to better understand the neurobiological correlates of cognitive processes, affective and social behaviors, personality traits, and putative risk factors for a range of mental and physical health syndromes and disorders. Arguably, interpreting the results of such studies of individual differences rests on a key (and often implicit) assumption that the characteristics of a given individual are moderately stable over time, despite idiosyncratic variation between individuals in dynamic factors that may affect brain function. Accordingly, functional neuroimaging and other psychophysiological methods to characterize the individual should yield stable empirical measures over multiple assessment occasions. Moreover, such occasion-to-occasion measurement stability (or test-retest reliability) is most convincing if assessment occasions are separated by intermediate time periods: days or months, rather than years. This is because longer intervals of assessment engender maturational, experiential, and other life changes that can obscure the reliability of a given individual difference measure. Similarly, closely spaced assessment occasions (e.g., within the same neuroimaging testing session) are likely to overestimate measurement stability. This is because situational influences and transient individual characteristics tend to be more stable over relatively short time periods.

The particular question addressed in the current study was whether individual differences in simultaneously measured changes in cardiovascular, behavioral, and blood-oxygen level-dependent (BOLD) signal activity assessed during a functional magnetic resonance imaging

Address correspondence to Peter Gianaros, Department of Psychology, University of Pittsburgh, 506 Old Engineering Hall, Pittsburgh, PA, 15260. Telephone: 412.624.9578. Fax: 412.624.5407., gianaros@pitt.edu.

(fMRI) protocol were stable over a timeframe of approximately three months. As described below, this question was motivated by recent work on the neural correlates of individual differences in stressor- or behaviorally-evoked cardiovascular reactivity (reviewed in Gianaros & Sheu, 2009), which is a parameter of risk for atherosclerotic coronary heart disease (CHD) and its sequelae (Chida & Steptoe, 2010).

To elaborate, CHD is characterized by the long-term progression of blood vessel remodeling, beginning with endothelial dysfunction and leading to various stages of coronary artery atherosclerosis that vary in severity across individuals (Libby & Theroux, 2005). Given the nature of such a progressive disease process, stable individual differences in biological and behavioral risk factors for CHD are thus most likely to influence trajectories of disease development over the life course (Kamarck & Lovallo, 2003; Krantz & Manuck, 1984; Treiber et al., 2003). In this regard, trait-like dimensions of cardiovascular reactivity to psychological or behavioral challenges have long been thought to be a particularly salient risk factor linked to disease pathophysiology, as these trait-like dimensions emerge early in life and appear to be moderately stable over time and across development – especially when measured in the context of CHD risk (Allen, Matthews, & Sherman, 1997; Kamarck et al., 1992; Llabre, Spitzer, Saab, Ironson, & Schneiderman, 1991; Low, Salomon, & Matthews, 2009; Manuck, 1994; Matthews, Salomon, Brady, & Allen, 2003; Roemmich et al., 2009; Saab et al., 2001; Treiber et al., 1994). More precisely, there is cumulative epidemiological and psychophysiological evidence that the magnitude of an individual's cardiovascular (e.g., blood pressure) reaction to an acute stressor confers risk for CHD (Chida & Steptoe, 2010). And in particular, individuals who appear to exhibit so-called 'exaggerated' or metabolically-excessive reactivity to acute stressors exhibit greater risk for CHD and related outcomes and disease precursors than those exhibiting lesser reactivity, although emerging evidence has begun to suggest that attenuated patterns of stressor-evoked cardiovascular reactivity themselves may also confer risk for other aspects of ill health among some individuals (Lovallo, 2011; Phillips, Hunt, Der, & Carroll, 2011).

Extending such epidemiological and psychophysiological evidence, recent functional neuroimaging work has begun to characterize the neurobiological correlates of individual differences in acute cardiovascular reactions. Hence, individual differences in blood pressure, heart rate, and other forms of cardiovascular reactivity have been linked to concurrent changes in the functionality (e.g., activation and deactivation) of cortical and networked subcortical areas that signal with preautonoimc cell groups to regulate peripheral cardiovascular physiology (Critchley et al., 2003; Gianaros & Sheu, 2009; Lane et al., 2009; Wager, Van Ast, et al., 2009; Wager, Waugh, et al., 2009). A critical question raised by this neuroimaging work remains, however: *Are neural activity changes that presumptively reflect stressor-evoked brain functionality reliable over time*? If so, then it would be more plausible to speculate that such individual differences in functional neural activity correspond to stable or trait-like neurobiological characteristics, perhaps related to CHD risk via downstream mechanisms impacting cardiovascular reactivity.

Importantly, the reliability of fMRI measures in particular has been examined in several prior studies, and these studies have identified several sources of variation that threaten reliability (Bennett & Miller, 2010). Major sources of variation include aspects of the magnetic resonance (MR) scanning environment and MR signal acquisition sequences employed, the demographic, anthropometric, and psychological characteristics of individuals, the task design, and the methods used for data processing and analysis. In a review of 63 fMRI reliability studies, Bennett and Miller (2010) detailed these and other sources of variation, and suggested various means for improving the quality and reliability of fMRI data. These suggestions included increasing (to the extent possible) the signal-to-noise ratio at the time of MRI data acquisition, minimizing individual differences in cognitive and mood states, and increasing sample sizes to improve statistical power. Further, Bennett and Miller suggested several conventional and emerging methods and metrics to assess the reliability of fMRI BOLD signal changes. Among the most common are the cluster overlap metric (Rombouts et al., 1997) and various forms of the intraclass correlation coefficient (ICC) (McGraw & Wong, 1996; Shrout & Fleiss, 1979). Although there are recognized limitations to both kinds of metrics, their interpretational ease has likely led to their preeminence in fMRI reliability studies.

The cluster overlap metric, or so-called overlap ratio (OR) assesses the reproducibility of brain regions engaged by a given task. An OR can be computed simply as the ratio of brain volumes exhibiting significant signal changes in task conditions that are administered in two different sessions to the average of that exhibited in the sessions—providing an estimate of the degree of spatial correspondence in brain regions engaged in different sessions. By contrast, ICC values reflect the *consistency* of signal changes between individuals, as computed by analysis of variance. Such consistency can be computed as the ratio of between-individual variance to the total variance observed. Note here that the OR emphasizes *consistency of spatial agreement*, whereas the ICC emphasizes *consistency of signal changes* in a given location or across a set of locations (e.g., distributed brain regions).

Among the imaging studies reviewed by Bennett and Miller, ORs typically ranged from 0.21 to 0.86, and ICCs of fMRI signal changes ranged from approximately 0.12 to 0.82. Not surprisingly, most studies reviewed reported that within-individual reliability was greater than between-individual reliability. Also, aspects of fMRI measurement reliability depended appreciably on the specific brain regions engaged by the particular experimental tasks studied, as well as the particular task conditions that were contrasted with one another. Hence, regions engaged by so-called ‘lower level’ motor and sensory tasks tended to exhibit greater spatial and signal-change reliability than those engaged by the conditions of ‘higher-level’ or cognitively demanding and complex tasks (Bennett & Miller, 2010).

For the purposes of the present study, we examined the test-retest reliability characteristics of fMRI BOLD signal changes evoked by two cognitively-demanding tasks that we have used to evoke individual differences in acute cardiovascular reactivity, wherein reactivity is reflected by the change in heart rate and blood pressure from a resting baseline period to a period of task performance. The fMRI tasks consisted of a modified Stroop color-word interference task (Stroop) and multi-source interference task (MSIT) that have been detailed previously (e.g., Gianaros, Onyewuenyi, Sheu, Christie, & Critchley, 2011; Gianaros et al., 2009), but not yet characterized for the reliabilities of their associated neural, behavioral, or cardiovascular outcome variables of interest.

To examine fMRI test-retest reliability characteristics, we first examined the spatial consistency or ‘reproducibility’ of task-evoked BOLD signal changes to each task using the OR described above. Second, we examined the reliability of task-evoked BOLD signal changes *within* individuals by computing ICCs, as derived from a mixed-effects ANOVA model. Third, we examined the consistency of task-evoked BOLD signal changes *between* individuals by ICCs, as also derived from a mixed-effects model. Fourth, based on the assumption that both the Stroop task and MSIT would elicit similar regional patterns of relative BOLD signal ‘activation’ and ‘deactivation’ due to comparable task features (see below), we examined whether aggregating (averaging) BOLD signal changes across the two tasks would improve OR and ICC values over-and-above those for either task alone. This was motivated by prior work showing that the reliability of cardiovascular reactivity measures can be improved in studies of individual differences by following psychometric principles (Kamarck & Lovallo, 2003). And finally, we examined the extent to which subjective ratings and cardiovascular reactivity indices derived from the tasks are stable over time, which is important for making inferences about trait-like or individual difference dimensions of reactivity derived from this battery.

Methods

Participants

Participants were 14 men (*M* age = 39.1 ± 6.3 *SD* years) and 12 women (*M* age = 40.8 ± 5.7 *SD years*), who were tested twice to estimate test-retest reliability. These 26 participants represented a sub-sample of 155 participants who comprise the full sample of a larger and ongoing study of the neurobiological correlates of CHD risk. All participants were recruited by mass mailings from the Department of Epidemiology at the University of Pittsburgh to residents of Allegheny County, Pennsylvania, USA. We have previously reported on the neural correlates of acute changes in baroreflex sensitivity in 96 participants of this larger study (Gianaros, et al., 2011). Manuscripts on the neural (fMRI) correlates of cardiovascular (heart rate and blood pressure) reactivity and subclinical atherosclerosis are in preparation for the full sample of this study, and are thus not reported in this smaller reliability sample of 26 individuals.

In brief, individuals responding to mass mailings were screened to exclude those with (*i*) a history of cardiovascular disease (including treatment for or diagnoses of hypertension, stroke, myocardial infarction, congestive heart failure, and atrial or ventricular arrhythmias); (*ii*) prior cardiovascular surgery (including coronary bypass, carotid artery, or peripheral vascular surgery); (*iii*) chronic kidney or liver conditions, Type I or II diabetes, or any pulmonary or respiratory diseases; (*iv*) reported diagnoses of or treatment for a substance abuse or mood disorder (including alcohol dependence, a somatization disorder, major depression, and panic or other anxiety disorders), as confirmed on interview using the Patient Health Questionnaire (Spitzer, Kroenke, & Williams, 1999), an inventory validated in outpatient (Kroenke, Spitzer, & Williams, 2001; Lowe et al., 2004; Spitzer, et al., 1999) and community samples (Martin, Rief, Klaiberg, & Braehler, 2006) against the Diagnostic and Statistical Manual of Mental Disorders IV (Lowe, Kroenke, Herzog, & Grafe, 2004); (*v*) prior cerebrovascular trauma involving loss of consciousness; (*vi*) prior neurosurgery or any neurological condition; (*vii*) being pregnant (verified by urine test in females); (*viii*) having claustrophobia or metallic implants; or (*ix*) taking psychotropic, lipid lowering, or cardiovascular medications.

All participants provided informed consent after receiving an explanation of study protocols. They were also tested in compliance with the Code of Ethics of the World Medical Association (Declaration of Helsinki), and with the approval of the University of Pittsburgh Institutional Review Board. The 26 participants who completed the reliability protocol described here were tested on two occasions, with the time between occasions ranging from 42 to 210 days (median interval = 88 days). They were asked not to consume caffeine, nicotine, or alcohol in the 12 hours before testing. Testing was completed at approximately the same time of the day on *both* occasions.

fMRI tasks

The Stroop task and the MSIT each lasted 9 min and 20 sec. Both involve processing conflictual information, receiving negative feedback, and making time-pressured responses to unpredictable and uncontrollable stimuli that elicit subjective distress. Briefly, participants completed 4, 52–60 sec blocks of trials in both tasks that defined a congruent condition, which were interleaved with 4, 52–60 sec blocks of trials defining an incongruent condition. Both conditions were preceded by a variable 10–17 sec fixation period.

In the Stroop task, participants identified the color of target words in the center of a screen by selecting 1 of 4 identifier words. Selections were made by pressing 1 of 4 buttons on a response glove, with each button matching an identifier word on the screen (e.g., thumb button 1 = identifier word on the far left, etc.). In congruent Stroop trials: (1) targets were in colors congruent with the target words, and (2) identifiers were in the same colors as targets. In incongruent Stroop trials: (1) targets were in colors incongruent with the targets, and (2) identifiers were in colors incongruent with the colors that the identifiers name.

In the MSIT, which was modified from the original version developed by Bush and Shin (2006), participants selected a number that differed from 2 others by pressing 1 of 3 buttons on the glove, with each button matching a number on the screen (thumb button 1 = number 1, etc.). In congruent MSIT trials, targets were in a position compatible with their position on the glove. In incongruent MSIT trials, targets were in a position incompatible with their glove position.

In incongruent conditions of *both* tasks, accuracy was held to ~60% by adjusting (jittering) inter-trial intervals (ITIs). Thus, accurate performance on 3 consecutive trials in a given incongruent condition prompted shorter ITIs in an incremental (step-wise) fashion, with increments occurring in 300 ms steps and the shortest ITI being 400 ms. Conversely, inaccurate performance on 3 consecutive trials lengthened ITIs in a similar manner, with the longest ITI being 5000 ms. To control for motor response differences between conditions in both tasks, the number of trials in the congruent condition was yoked to the number completed in the incongruent condition. To implement yoking, (1) an incongruent block was administered first and (2) congruent condition trials were presented at the mean ITI of the preceding incongruent block (Figure 1). Both tasks are freely available on request to the corresponding author.

Participants performed the tasks in a counterbalanced order, and the two tasks were separated by an approximate 10–12 min recovery period during which they rested quietly while high-resolution structural brain images were acquired (see below). To assess subjective ratings of valence (1-very unhappy; 9-very happy), arousal (1-very calm; 9-very aroused), and perceived control (1-very little control; 9-very much control), participants completed a modified self-assessment manikin scale (Bradley & Lang, 1994) after an initial baseline period and immediately after each task was completed.

Cardiovascular monitoring and behavioral assessments

During MRI scanning on both testing occasions, blood pressure (BP) and heart rate (HR) were monitored from the participant’s non-dominant arm by an oscillometric device (Multigas 9500, MedRad, Inc., Warrendale, PA). Resting BP and HR were taken every 2 min during a 10 min baseline (quiet rest period) before the tasks started. The last 2 readings were averaged and used as the baseline measure. Task BP and HR were taken every 1 min, concurrent with each condition in each task block. Mean incongruent BP and HR values were calculated by averaging the 4 respective readings obtained over the 4 incongruent blocks (separately for each task). BP and HR reactivity measures were then calculated as the difference of the mean incongruent and baseline measures. Stimulus presentation and behavioral response data acquisition were controlled by E-Prime® software (Psychology Software Tools, Inc., Sharpsburg, PA).

Image acquisition and processing

All functional and structural brain images were acquired by a 3T Trio TIM whole-body scanner (Siemens, Erlangen, Germany). Radio frequency (RF) detection was achieved using a 12-channel phased-array head coil. The blood-oxygen-level-dependent (BOLD) imaging sequence consisted of a T2*-weighted gradient-echo echo planar acquisition routine with the following parameters: echo time (TE) = 28 ms, repetition time (TR) = 2000 ms, flip angle (FA) = 90°, and field of view (FOV) = 205 mm x 205 mm (matrix size 64×64), slice thickness = 3 mm with no gap. A high-resolution, T1-weighted anatomical image was also acquired to register BOLD images to a common space using a magnetization-prepared rapid-gradient echo sequence (MPRAGE, TE = 3.29 ms, TR = 2100 ms, inversion time (TI)= 1100 ms, FA= 8°, FOV = 256 mm x 208 mm (matrix size: 256×208), slice thickness = 1 mm with no gap.

All brain images were processed and analyzed using SPM8 (Wellcome Trust Centre for Neuroimaging, London, UK, www.fil.ion.ucl.ac.uk) and MATLAB® software (The MathWorks, Inc., Natick, MA). For each task, BOLD functional images were realigned to the first image obtained in a given task using a rigid body 6-parameter model. We also applied the SPM Realign and Unwarp algorithm to adjust the geometric distortion due to movement (Andersson, Hutton, Ashburner, Turner, & Friston, 2001). Structural grey matter was extracted from the participant's MPRAGE image using the SPM unified segmentation algorithm (Ashburner & Friston, 2005). The extracted grey matter image was co-reregistered to the mean functional BOLD image from each task for each participant (by rigid-body transformation) and then registered to an SPM MNI grey matter template (ICBM, NIH P-20 project) using nonlinear affine transformation. All functional BOLD images were then spatially normalized to MNI space according to the parameters derived from the grey matter image registrations and re-sliced to voxel size of 2 mm × 2 mm × 2 mm. Finally, the normalized BOLD images were smoothed with a Gaussian kernel of 6mm FWHM. The final BOLD images were inspected for global mean signal and motion-related outliers using Artifact Detection Tools (ART), available for download at http://www.nitrc.org/projects/artifact_detect/. Outliers were defined as a mean BOLD signal intensity value observed in a given volume (for each TR) that was greater than 3 standard deviations distant from the global mean or a movement displacement from the reference (first volume) that was more than 2 mm in translation or 2 degrees in rotation. Across all 280 volumes acquired for each task (imaging run) for each participant on both testing occasions, we observed 5 or fewer outliers per participant according to the above criteria.

The effects of task conditions were estimated by voxel-wise general linear models (GLMs). Specifically, conditions were modeled by a boxcar function that was convolved with the SPM canonical hemodynamic response function (HRF). Further, the six movement parameters obtained from the realignment procedure were mean centered and included in the GLM as covariates to adjust for movement variance. Prior to performing the GLMs, the correlation between conditions and movement parameters were examined for each participant to ensure that there were no confounding task-by-motion correlations (all *r*s < 0.2, *p*s > 0.05). A high-pass filter of 187 seconds was applied to the BOLD temporal signals to remove low-frequency artifacts, e.g., slow BOLD signal drifts or physiological signal aliasing across the imaging run. An autoregressive model (AR[1]) was further used to adjust for voxel-wise time series autocorrelations. The individual contrast images corresponding to the BOLD response to the incongruent condition and the congruent condition of each task were estimated from GLMs using the fixation condition as a reference condition. The contrast of the incongruent vs. congruent condition for each task was also estimated. All of these contrast images were collected for cross-task and cross-session ICC analyses.

ICC analyses

ICC analyses in this study followed the so-called case 3 model, ICC (3,1), as defined by Shrout and Fleiss (1979). The model is a two-way mixed effects ANOVA: $Y_{ij} = \mu + r_i + c_j + e_{ij}$, where r_i is a random effect of factor r at level i; c_j is a fixed effect of factor c at level j; μ is the grand mean; and e_{ij} is a term for random error. We examined three contrast images (or contrast maps), *incongruent vs. fixation*, *congruent vs. fixation*, and *incongruent vs. congruent*, across both tasks and across both sessions. Specifically, when testing the consistency of the contrast values (BOLD signal changes) across the two tasks, Y_{ij} is the signal change of voxel i for the j^{th} task, and wherein r_i is the voxel random effect and c_j is the task fixed effect. When testing the consistency of the measurements across the two sessions, factor c represents session, and factor was either 'participant' or 'voxel' - depending on whether the test was for between- or within-individual consistency. Finally, Y_{ij} is the signal change for participant (or voxel) i in the j^{th} session. ICC, which represents the ratio of between-units (i.e., participants or voxels) variance over the total variance, was calculated as $ICC(3,1)=\frac{BMS-EMS}{BMS+EMS}$, where BMS is the mean square of between-unit variance and EMS is error variance or mean square of within-unit variance taken from the ANOVA.

In this study, we assessed ICCs in regions-of-interest (ROIs) that contained voxels engaged by the tasks (i.e., exhibiting relative patterns of activation or deactivation). These ROIs were determined from a random-effects analysis of a larger sample of 138 out of the full 155 participants described above. This approach for ROI derivation was taken to decrease ROI selection bias by increasing the representativeness and accuracy of the contrast maps derived from the two tasks (as opposed to generating the ROIs from the potentially idiosyncratic changes exhibited by the sub-sample of 26 reliability participants). These 138 subjects were those that had viable (e.g., artifact free) fMRI data out of the full sample of 155. Accordingly, the individual contrasts for each task were estimated as described above for each participant and submitted to a one-sample t-test. Voxels in brain regions identified in these contrast maps passed a whole-brain and family-wise error threshold of $p < 0.05$, which was combined with a simultaneous cluster extent threshold of $k > 20$ voxels.

For uniformity of expression, we refer to regions exhibiting a relative *increase* in BOLD signal activity (from positive contrasts) as 'activated areas'. Conversely, we refer to regions exhibiting a relative *decrease* in BOLD signal activity (from negative contrasts) as 'deactivated areas'. For each task, 6 ROI masks were thus created for the three contrasts described above, and these were used for ICC analysis to examine: 1) the cross-task ICC to assess the consistency of relative activation and deactivation evoked by the Stroop task and the MSIT in the ROIs; 2) the within-individual, cross-session ICC to assess the consistency of activation and deactivation patterns within the ROIs for each participant; and 3) the between-individual, cross-session ICC to assess the consistency of activation and deactivation patterns between individuals within the ROIs.

Analysis of task-related subjective and physiological responses

We examined changes in self-report ratings of subjective arousal, valence, and control for each task using paired t-tests, with baseline ratings serving as references for each test. Cardiovascular responses to both tasks were also examined by paired t-tests, with the average of the baseline readings serving as referents. Ratings data and cardiovascular reactions across testing sessions were then evaluated for the degree of consistency between individuals by ICC analyses (ICC (3,1) model). Statistical and ICC analyses were executed in SPSS (version 19, IBM SPSS Statistics, Chicago, IL).

Analyses of spatial reproducibility of brain regions engaged

The activation and deactivation patterns exhibited by brain regions revealed by each contrast in each session were estimated for the 26 participants by a random-effects analysis, as described above. To evaluate the spatial reproducibility of the regions engaged by the tasks, we first calculated the OR as $OR=\frac{2V_{12}}{V_1+V_2}$, wherein V_{12} is the number of significant voxels that pass a given F-test threshold in both sessions, and wherein V_i is the number of significant voxels for the i^{th} session (Rombouts, et al., 1997). The OR was calculated for the three contrasts described above with a whole-brain, false-detection rate threshold of 0.05 combined with a cluster extent threshold of $k = 20$ voxels.

In addition to the OR, spatial reproducibility can also be evaluated by the ICCs of the voxel-wise t-maps generated by group-level analyses across sessions (Raemaekers et al., 2007). An arguable advantage of this approach is that it is not limited by arbitrary statistical or extent threshold choices. Accordingly, we examined the ICCs of the t-maps across sessions within the ROIs derived from each task and derived from task aggregation to compare with the results of the OR analyses.

Reliability assessments of task-related BOLD signal changes

The consistency of BOLD signal changes within- and between-individuals observed in the ROIs was assessed by the ICC(3,1) model, and calculated using the intra-voxel ICC and intra-subject reliability for cluster analysis routines in an ICC toolbox (Caceres, Hall, Zelaya, Williams, & Mehta, 2009). To examine within-individual reliability, we calculated ICCs in the ROIs for each participant and estimated within-individual reliabilities in each ROI by the median ICC among participants. To examine between-individual reliability, we generated ICC maps for each ROI from all 26 participants, and then estimated the between-individual reliability in each ROI by the median ICCs in the ROIs. All of the median ICCs and their standard errors (SE) were estimated by bootstrapping with 1000 iterations.

Reliability assessments of task-aggregated BOLD signal changes

Finally, we examined the ICCs of BOLD signal changes in brain areas engaged by both the Stroop task and the MSIT to determine if functional neural responses to these tasks can be aggregated (averaged across the two tasks) to reduce measurement error and improve test-retest reliability characteristics following psychometric work done on individual differences in cardiovascular reactivity (Kamarck, et al., 1992; Kamarck & Lovallo, 2003). This was done across participants and within the voxels where we observed BOLD signal changes that met whole-brain corrected thresholds to control our false positive detection rate. Accordingly, an average contrast map of the two tasks was calculated for each participant to derive a task-aggregated measure of the change in functional neural activity. We then examined if task-aggregation improved test-retest reliability metrics over-and-above those associated with a single task. Accordingly, the reliability of the aggregated BOLD signal changes was assessed using the same methods described above for each task, and was compared with the single-task estimates of reliability.

Results

Behavioral task performance, subjective ratings, and cardiovascular reactivity

Performance and self-report measures of subjective feeling states showed expected task-related changes for both testing sessions. Hence, percent accuracy during the incongruent condition was appropriately titrated to be approximately 60% (M = 59% ± 6.2% SD for the Stroop task in the first session, and M = 61% ± 7.6% SD in the second session; M = 58% ± 3.6% SD for the MSIT in the first session, and M = 56% ± 3.0% SD in the second session).

Self-report ratings showed that the tasks elicited subjective arousal and decreased emotional valence in both sessions, with all *t*'s > 3.50, *p*'s < 0.02. Participants also reported a decrease in their sense of control in the first session for both tasks (*t*'s > 3.5, *p*'s < 0.002), but not significantly so in the second session, $t = 1.8$, $p = 0.07$ for the Stroop task and $t = 1.2$, $p = 0.22$ for the MSIT). And as expected and detailed in Table 1, significant cardiovascular reactions to the tasks were observed in both sessions, with all *t*'s > 2.4, *p*'s < 0.02.

Subjective ratings and cardiovascular reactions to the tasks showed strong consistency across sessions among participants (see Table 1). The ICCs of the subjective ratings ranged from 0.36 to 0.77 (all *p*'s < 0.04), and those for cardiovascular reactivity ranged from 0.75 to 0.85 (all *p*'s < 0.001). The ICCs of cardiovascular reactivity measures also showed strong consistency across the two tasks: ICCs ranged from 0.79 to 0.96, with all *p*'s < 0.02. The ICCs of task-aggregated reactivity measures in the two sessions were comparable to those of the single task estimates for SBP (ICC = 0.84, p<0.01) and better for HR (ICC = 0.87, p<0.01).

Brain regions engaged by the Stroop and MSIT tasks

The Stroop task and the MSIT engaged a comparable set of brain areas in the 138 participants from the full sample of 155, as revealed by a random-effects analysis and in replication of our prior studies using these two tasks (Gianaros, et al., 2011; Gianaros & Sheu, 2009; Gianaros et al., 2008). Specifically, both tasks evoked increases in BOLD signal activity ('activation') within brain areas that are reliably engaged by effortful cognitive control tasks involving the processing of conflict and response inhibition, including areas within the anterior cingulate cortex, anterior insula, parietal cortex, basal ganglia, thalamus, and cerebellum (Figure 2; Supplemental Tables 1–2). In addition and also replicating our prior work with these tasks, decreases in BOLD signal activity ('deactivation') during both tasks were observed in brain areas that are thought to comprise the so-called 'default-mode network', including areas within the ventromedial prefrontal cortex, perigenual anterior cingulate cortex, posterior cingulate cortex, and adjacent precuneus (Figure 2; Supplemental Tables 1–2). These areas that exhibited relative patterns of activation and deactivation as revealed by each task contrast were saved as composite ROI masks, and then applied in the ICC analyses below.

Among the sub-sample of 26 participants who completed the MRI protocol twice, we found that the BOLD signal changes revealed by the incongruent vs. fixation contrast showed good consistency across tasks in what we henceforth refer to as the 'activated' and 'deactivated' ROIs: the median cross-task ICC was 0.70 (SE = 0.03) in the activated ROIs and 0.47 (SE = 0.05) in the deactivated ROIs. For the incongruent vs. congruent contrast, the ICC across tasks was lower: the median ICC = 0.31 (SE = 0.06) for the activated ROIs, and the median ICC = 0.31 (SE = 0.03) for the deactivated ROIs.

Spatial reliability of activated and deactivated brain regions across the two sessions

The group analysis of BOLD signal changes evoked by the Stroop task and MSIT across two sessions yielded findings that were comparable for both tasks. Hence, the spatial reliabilities measured by the OR for the incongruent vs. fixation contrast map were 0.76 for the Stroop task, 0.70 for the MSIT, and 0.77 for the task-aggregated map. For the incongruent vs. congruent contrast map, the ORs were 0.66, 0.70, and 0.71 for the Stroop task, the MSIT, and the task-aggregated map, respectively. (We varied significance thresholds in ancillary analyses of the ORs, but did not detect noteworthy differences across more lenient and conservative thresholds.)

The above results were comparable to those in which we assessed the ICCs of the t-maps generated from the two MRI sessions. Hence, the median ICC of the t-statistics for the incongruent vs. fixation contrast maps across sessions in the activated ROIs was 0.82 for the Stroop task, 0.74 for the MSIT, and 0.81 for the aggregated map; for deactivated areas revealed by this contrast, the respective median ICCs of the t-statistics were 0.70, 0.67, and 0.76. The ICCs for the t-maps are summarized in Table 2.

Reliability of BOLD signal changes within individuals

The boxplots in Figure 3 illustrate the ICCs of BOLD signal changes for each contrast in all ROIs for all 26 participants. Median ICCs among participants for the contrasts of each task are also in Table 2. In general, the ICCs in the activated ROIs were stronger and more consistent (i.e., less variable) among participants than the deactivated ROIs. The ICCs for the incongruent vs. congruent contrast in particular were notably lower than those for the other contrasts (median ICC < 0.5). And in the main, the ICCs for the task-aggregated contrasts were superior to those for any single task alone. In particular, the ICC for the task-aggregated contrast map had a median of 0.82 (SE = 0.01) for the incongruent vs. fixation contrast in the activated ROIs, and of 0.56 (SE = 0.04) in the deactivated ROIs. These compare with 0.77 (SE = 0.01) and 0.47 (SE = 0.04) for the Stroop task and 0.72 (SE = 0.03) and 0.46 (SE = 0.04) for the MSIT. And even for the incongruent vs. congruent contrast, which was relatively weak for each task as noted above, the task-aggregated contrast maps exhibited more acceptable ICCs, with a median = 0.52 (SE = 0.04) in the activated ROIs and 0.42 (SE = 0.08) in deactivated ROIs for this particular contrast (see Table 2).

Reliability of BOLD signal changes between individuals

Measures of between-individual reliability in the ROIs are illustrated by the voxel-wise ICC maps in Figure 4 (panels A–C), accompanied by the ICC distributions in the activated and deactivated ROIs (panels D–F). Further, the median ICCs in the ROIs for each contrast and task are shown in Figure 5 (also summarized in Table 2). In all, activated regions generally exhibited stronger ICCs than did deactivated regions (see Figure 5). An examination of the ICC distributions in Figure 4 particularly suggests that task-aggregation increases the power or sensitivity to detect ‘significant’ voxels, as compared with either task alone. Further, the incongruent vs. fixation contrast maps for both tasks and for the task-aggregated contrasts exhibited stronger ICCs, as compared with the other contrast maps. And, the incongruent vs. congruent contrast map exhibited relatively weak ICCs for both tasks (< 0.4, except for the Stroop incongruent vs. fixation contrast in the deactivated ROIs). Finally, the task-aggregated maps outperformed the single task ICCs for all contrast maps (median ICCs ranged from 0.46 to 0.64).

Discussion

In this assessment of the test-retest reliability characteristics of fMRI BOLD signal changes to a Stroop task and a multi-source interference task designed to evoke cardiovascular reactivity, we found that (1) the brain regions engaged by these tasks are consistent across the two tasks, (2) the fMRI BOLD signal changes within these regions are (for the most part) moderately reliable over a timeframe of approximately three months, and (3) aggregating BOLD signal changes across the two tasks improves test-retest reliability metrics, as compared with those associated with either task alone. Our evaluation of the reliability of the cardiovascular reactions to the tasks among participants across the two sessions yielded ICCs ranging from 0.75 to 0.85, values comparable to those previously reported in the reliability literature on cardiovascular reactivity (Kamarck et al., 2003; Jennings et al., 1998). In parallel, the behavioral reliability of subjective reports of arousal, valence, and control exhibited moderate to strong levels of reliability as well, with ICCs ranging from 0.36 to 0.77 for the Stroop and multi-source interference tasks. In all, the present results would seem to suggest that BOLD signal changes to the tasks used here may be interpreted as reflecting relatively stable characteristics of the individual, with reliabilities on par with those of peripheral physiological (cardiovascular) responses and even subjective responses reported in the psychophysiological reactivity literature.

It is important to note here, however, that the reliability of brain activity changes would not be expect to be perfect over a three-month period, and one would not expect such changes to appreciably exceed the reliability of concurrent physiological or subjective responses to the tasks. Moreover, there is at present no consensus on what represents an acceptable ICC value in functional neuroimaging studies. Accordingly, we tentatively infer that the incongruent contrast maps of BOLD signal changes to both the Stroop task and MSIT showed what could be interpreted as reasonable consistency among participants in general (the median ICCs ranged from 0.45 to 0.56 in activated and deactivated brain regions). The incongruent vs. congruent contrasts, however showed lower temporal consistency, in which median ICCs fell below 0.4 in most brain regions of interest. This pattern of results would seem to suggest that the congruent condition of both tasks might serve as more than just a comparative ‘baseline’ or ‘referent’ period for contrast analyses. In extension and perhaps more importantly to consider for future studies, the incongruent vs. fixation contrast may be more informative, robust, and reliable than the incongruent vs. congruent contrast for the present reactivity task paradigm.

Another key finding from the present study was that aggregating (or averaging) BOLD signal changes across the Stroop task and MSIT improved all reliability metrics over-and-above those for any single task (the improvement was 15% on average for the incongruent vs. fixation contrast and 31% for the incongruent vs. congruent contrast). Also, aggregating BOLD signal changes across the tasks improved ICC values for the otherwise poorer performing (less reliable) incongruent vs. congruent contrast, with a median ICC of 0.46 for activated regions and of 0.48 for deactivated regions. In view of these findings, we conclude that the task-aggregated incongruent vs. fixation contrast map exhibits the best ICC, with median ICCs of 0.64 and 0.56 for activated and deactivated regions, respectively.

The present test-retest fMRI results also provide an impetus for considering ways to improve the reliability characteristics of measures derived from functional neuroimaging paradigms for the study of individual differences in cardiovascular reactivity. In the past two decades, the reliability of fMRI measures has been assessed using many different methods, and no standard reliability parameters or criteria have yet emerged. Although the difficulty of determining fMRI reliability may in part be due to idiosyncratic features of the scanning measurement environment, it seems likely that the test-retest variability of neuroimaging measures is due largely to the dynamic neural process under study, as well the methods used for quantifying these processes. Moreover, an optimal fMRI data analysis strategy of reliable or stable neural processes depends on being able to detect robust statistical changes in fMRI BOLD signal changes. In this study, we found that aggregating BOLD signal changes over multiple tasks improved the sensitivity to detect patterns of relative activation and deactivation in brain areas than evaluating individual tasks alone. Also, task aggregation resulted in improved test-retest reliability over time. As evidence, aggregating BOLD signal changes over tasks resulted in ORs ranging from 0.66 to 0.77, which were appreciably higher than the ORs from either task alone. We note here though that OR values are likely to be influenced by the registration method used during preprocessing, and OR values may be improved further by the use of more sensitive or accurate algorithms that are argued to better co-register functional activity maps onto individual anatomical images prior to normalization routines (Klein et al., 2009; Lacadie, Fulbright, Rajeevan, Constable, & Papademetris, 2008).

It is also mentionable here that procedures that may increase the statistical power in group-level functional imaging analyses involving the present task paradigm might also be considered in light of the present findings. To elaborate, the overall within-individual ICC for the voxel-wise reliability of BOLD signal changes evoked by the Stroop task and MIST appeared to be relatively reasonable (e.g., the incongruent vs. fixation contrast for both tasks had median ICCs > 0.72 in activated regions and > 0.46 in deactivated regions, whereas the task-aggregated contrast had median ICCs of 0.82 and 0.55 in activated and deactivated regions, respectively). Striking individual differences, however, were still notable. In the context of group-level analyses, such individual differences in repeatability may not only represent individual (e.g., trait) characteristics, but could also impact statistical significance and inference for particular task contrasts (e.g., with contrasts of weaker repeatability possibly engendering weaker statistical power at the group level). Finally, several approaches not used here could be considered to improve ICC measures. One involves the more 'robust' estimation of BOLD signal changes. Currently, popular GLM analyses are sensitive to outliers, yet robust regression methods, which are designed to reduce such sensitivity, may be considered (Wager, Keller, Lacey, & Jonides, 2005). Another involves accounting for the between-region variability in derived reliability metrics. As one example, voxel-wise ICC analyses (see Figure 4) demonstrated that ICC values within the dACC were typically higher than those within the amygdala, for reasons that may be attributable to response habituation (e.g., Breiter et al., 1996), related variation over time in learning, error, and conflict resolution processes, or even factors that differentially impact measurement error across different brain regions by these tasks. Therefore, it is possible that different requirements or methods may be needed to reach the same statistical power to detect BOLD signal changes in different regions of interest engaged by the present task paradigm.

Conclusion

In this assessment of the test-retest reliability of fMRI BOLD signal changes engaged by a standardized cardiovascular reactivity task battery, we found an arguably acceptable level of measurement reliability over a timeframe of approximately three months. Such stability over an approximately three-month timeframe would appear to support the assumption that this reactivity battery may be useful for the study of stable or trait-like neurobiological characteristics of the individual - and consequently the neural correlates of individual differences in cardiovascular reactivity (and perhaps other neurocognitive and psychophysiological response processes). We note, however, that the reliability of the measures studied here can and should be improved through different measurement, fMRI preprocessing, and statistical strategies. And perhaps most importantly, the present study showed for the first time to our knowledge that aggregating (averaging) measurements of functional changes in regional brain activity across tasks that share stimulus and response features yield more robust and reliable outcome variables than those associated with individual tasks. However, whether such aggregation would improve the temporal stability characteristics of functional neural responses evoked within the context of other task paradigms or even other functional brain imaging metrics (e.g., functional or effective connectivity) are open and important questions.

In view of the above, the present findings can be taken to support further work examining the neural correlates of individual differences in cardiovascular reactivity evoked by the Stroop task and MSIT used here, as well as testing whether such correlates associate with indicators of CHD risk that have themselves been linked to cardiovascular reactivity. In this way, such work could deepen our understanding of the neurobiological pathways and functional neural processes linking individual differences in cardiovascular reactivity to cardiovascular health and disease states.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was funded by National Institutes of Health Grants HL089850 and HL101421 to Peter J. Gianaros. The authors thank Sara Snyder for her assistance in data collection.

References

Allen MT, Matthews KA, Sherman FS. Cardiovascular reactivity to stress and left ventricular mass in youth. Hypertension. 1997; 30(4):782–787.10.1161/01.HYP.30.4.782 [PubMed: 9336373]

Andersson JL, Hutton C, Ashburner J, Turner R, Friston K. Modeling geometric deformations in EPI time series. Neuroimage. 2001; 13(5):903–919. S1053-8119(01)90746-3 [pii]. 10.1006/nimg.2001.0746 [PubMed: 11304086]

Ashburner J, Friston KJ. Unified segmentation. Neuroimage. 2005; 26(3):839–851. S1053-8119(05)00110-2 [pii]. 10.1016/j.neuroimage.2005.02.018 [PubMed: 15955494]

Bennett CM, Miller MB. How reliable are the results from functional magnetic resonance imaging? Annals of the New York Academy of Sciences. 2010; 1191:133–155. NYAS5446 [pii]. 10.1111/j.1749-6632.2010.05446.x [PubMed: 20392279]

Bradley MM, Lang PJ. Measuring emotion: the Self-Assessment Manikin and the Semantic Differential. Journal of Behavior Therapy and Experimental Psychiatry. 1994; 25(1):49–59. [PubMed: 7962581]

Breiter HC, Etcoff NL, Whalen PJ, Kennedy WA, Rauch SL, Buckner RL, Rosen BR. Response and habituation of the human amygdala during visual processing of facial expression. Neuron. 1996; 17(5):875–887.10.1016/S0896-6273(00)80219-6 [PubMed: 8938120]

Bush G, Shin LM. The Multi-Source Interference Task: an fMRI task that reliably activates the cingulo-frontal-parietal cognitive/attention network. Nature Protocols. 2006; 1(1):308–313.10.1038/nprot.2006.48

Caceres A, Hall DL, Zelaya FO, Williams SC, Mehta MA. Measuring fMRI reliability with the intra-class correlation coefficient. NeuroImage. 2009; 45(3):758–768. doi:http://dx.doi.org/10.1016/j.neuroimage.2008.12.035. [PubMed: 19166942]

Chida Y, Steptoe A. Greater cardiovascular responses to laboratory mental stress are associated with poor subsequent cardiovascular risk status: a meta-analysis of prospective evidence. Hypertension. 2010; 55(4):1026–1032.10.1161/HYPERTENSIONAHA.109.146621 [PubMed: 20194301]

Critchley HD, Mathias CJ, Josephs O, O'Doherty J, Zanini S, Dewar BK, Dolan RJ. Human cingulate cortex and autonomic control: converging neuroimaging and clinical evidence. Brain. 2003; 126(10):2139–2152.10.1093/brain/awg216 [PubMed: 12821513]

Gianaros PJ, Onyewuenyi IC, Sheu LK, Christie IC, Critchley HD. Brain systems for baroreflex suppression during stress in humans. Human Brain Mapping. 201110.1002/hbm.21315

Gianaros PJ, Sheu LK. A review of neuroimaing studies of stressor-evoked blood pressure reactivity: emerging evidence for a brain-body pathway to coronary heart disease risk. Neuroimage. 2009; 47(3):922–936. http://dx.doi.org/10.1016/j.neuroimage.2009.04.073. [PubMed: 19410652]

Gianaros PJ, Sheu LK, Matthews KA, Jennings JR, Manuck SB, Hariri AR. Individual differences in stressor-evoked blood pressure reactivity vary with activation, volume, and functional connectivity of the amygdala. Journal of Neuroscience. 2008; 28(4):990–999.10.1523/JNEUROSCI.3606-07.2008 [PubMed: 18216206]

Gianaros PJ, Sheu LK, Remo AM, Christie IC, Crtichley HD, Wang J. Heightened resting neural activity predicts exaggerated stressor-evoked blood pressure reactivity. Hypertension. 2009; 53(5):819–825.10.1161/HYPERTENSIONAHA.108.126227 [PubMed: 19273741]

Kamarck TW, Jennings JR, Debski TT, Glickman-Weiss E, Johnson PS, Eddy MJ, Manuck SB. Reliable measures of behaviorally-evoked cardiovascular reactivity from a PC-based test battery: results from student and community samples. Psychophysiology. 1992; 29(1):17–28. [PubMed: 1609024]

Kamarck TW, Lovallo WR. Cardiovascular reactivity to psychological challenge: conceptual and measurement considerations. Psychosomatic Medicine. 2003; 65(1):9–21. [PubMed: 12554812]

Klein A, Andersson J, Ardekani BA, Ashburner J, Avants B, Chiang MC, Parsey RV. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. Neuroimage. 2009; 46(3):786–802.10.1016/j.neuroimage.2008.12.037 [PubMed: 19195496]

Krantz DS, Manuck SB. Acute psychophysiologic reactivity and risk of cardiovascular disease: a review and methodologic critique. Psychological Bulletin. 1984; 96(3):435–464. [PubMed: 6393178]

Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. Journal of General Internal Medicine. 2001; 16(9):606–613. [PubMed: 11556941]

Lacadie CM, Fulbright RK, Rajeevan N, Constable RT, Papademetris X. More accurate Talairach coordinates for neuroimaging using non-linear registration. Neuroimage. 2008; 42(2):717–725.10.1016/j.neuroimage.2008.04.240 [PubMed: 18572418]

Lane RD, Waldstein SR, Chesney MA, Jennings JR, Lovallo WR, Kozel PJ, Cameron OG. The rebirth of neuroscience in psychosomatic medicine, part I: historical context, methods and relevant basic science. Psychosomatic Medicine. 2009; 71(2):117–134.10.1097/PSY.0b013e31819783be [PubMed: 19196808]

Libby P, Theroux P. Pathophysiology of coronary artery disease. Circulation. 2005; 111(25):3481–3488.10.1161/CIRCULATIONAHA.105.537878 [PubMed: 15983262]

Llabre MM, Spitzer SB, Saab PG, Ironson GH, Schneiderman N. The reliability and specificity of delta versus residualized change as measures of cardiovascular reactivity to behavioral challenges. Psychophysiology. 1991; 28(6):701–711. [PubMed: 1816598]

Lovallo WR. Do low levels of stress reactivity signal poor states of health? Biological Psychology. 2011; 86(2):121–128.10.1016/j.biopsycho.2010.01.006 [PubMed: 20079397]

Low CA, Salomon K, Matthews KA. Chronic life stress, cardiovascular reactivity, and subclinical cardiovascular disease in adolescents. Psychosomatic Medicine. 2009; 71(9):927–931. [PubMed: 19737856]

Lowe B, Grafe K, Zipfel S, Witte S, Loerch B, Herzog W. Diagnosing ICD-10 depressive episodes: superior criterion validity of the Patient Health Questionnaire. Psychotherapy and Psychosomatics. 2004; 73(6):386–390.10.1159/000080393 [PubMed: 15479995]

Lowe B, Kroenke K, Herzog W, Grafe K. Measuring depression outcome with a brief self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). Journal of Affective Disorders. 2004; 81(1):61–66.10.1016/S0165-0327(03)00198-8 [PubMed: 15183601]

Manuck SB. Cardiovascular reactivity in cardiovascular disease: Once more unto the breach. International Journal of Behavioral Medicine. 1994; 1(1):4–31.10.1207/s15327558ijbm0101_2 [PubMed: 16250803]

Martin A, Rief W, Klaiberg A, Braehler E. Validity of the Brief Patient Health Questionnaire Mood Scale (PHQ-9) in the general population. General Hospital Psychiatry. 2006; 28(1):71–77.10.1016/j.genhosppsych.2005.07.003 [PubMed: 16377369]

Matthews KA, Salomon K, Brady SS, Allen MT. Cardiovascular reactivity to stress predicts future blood pressure in adolescence. Psychosomatic Medicine. 2003; 65(3):410–415. [PubMed: 12764214]

McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychological Methods. 1996; 1(1):30–46.10.1037/1082-989X.1.4.390

Phillips AC, Hunt K, Der G, Carroll D. Blunted cardiac reactions to acute psychological stress predict symptoms of depression five years later: evidence from a large community study. Psychophysiology. 2011; 48(1):142–148.10.1111/j.1469-8986.2010.01045.x [PubMed: 20536905]

Raemaekers M, Vink M, Zandbelt B, van Wezel RJ, Kahn RS, Ramsey NF. Test-retest reliability of fMRI activation during prosaccades and antisaccades. Neuroimage. 2007; 36(6):532–542.10.1016/j.neuroimage.2007.03.061 [PubMed: 17499525]

Roemmich JN, Lobarinas CL, Joseph PN, Lambiase MJ, Archer FD III, Dorn J. Cardiovascular reactivity to psychological stress and carotid intima-media thickness in children. Psychophysiology. 2009; 46(2):293–299. PSYP776 [pii]. 10.1111/j.1469-8986.2008.00776.x [PubMed: 19207200]

Rombouts SA, Barkhof F, Hoogenraad FG, Sprenger M, Valk J, Scheltens P. Test-retest analysis with functional MR of the activated area in the human visual cortex. AJNR American Journal of Neuroradiology. 1997; 18(7):1317–1322. [PubMed: 9282862]

Saab PG, Llabre MM, Ma M, DiLillo V, McCalla JR, Fernander-Scott A, Schneiderman N. Cardiovascular responsivity to stress in adolescents with and without persistently elevated blood pressure. Journal of Hypertension. 2001; 19(1):21–27. [PubMed: 11204300]

Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin. 1979; 2(2):420–428. [PubMed: 18839484]

Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: The PHQ primary care study. Journal of the American Medical Association. 1999; 282(18):1737–1744. [PubMed: 10568646]

Treiber FA, Kamarck TW, Schneiderman N, Sheffield D, Kapuku G, Taylor T. Cardiovascular reactivity and development of preclinical and clinical disease states. Psychosomatic Medicine. 2003; 65(1):46–62. [PubMed: 12554815]

Treiber FA, Murphy JK, Davis H, Raunikar RA, Pflieger K, Strong WB. Pressor reactivity, ethnicity, and 24-hour ambulatory monitoring in children from hypertensive families. Behavioral Medicine. 1994; 20(3):133–142.10.1080/08964289.1994.9934628 [PubMed: 7865933]

Wager TD, Keller MC, Lacey SC, Jonides J. Increased sensitivity in neuroimaging analyses using robust regression. Neuroimage. 2005; 26(1):99–113.10.1016/j.neuroimage.2005.01.011 [PubMed: 15862210]

Wager TD, Van Ast V, Hughes B, Davidson M, Lindquist MA, Ochsner KN. Brain mediators of cardiovascular responses to social threat, Part II: prefrontal-subcortical pathways and relationship with anxiety. Neuroimage. 2009; 47(3):836–851.10.1016/j.neuroimage.2009.05.044 [PubMed: 19465135]

Wager TD, Waugh CE, Lindquist MA, Noll DC, Fredrickson BL, Taylor SF. Brain mediators of cardiovascular responses to social threat, Part I: reciprocal dorsal and ventral sub-regions of the medial prefrontal cortex and heart-rate reactivity. Neuroimage. 2009; 47(3):821–835.10.1016/j.neuroimage.2009.05.043 [PubMed: 19465137]

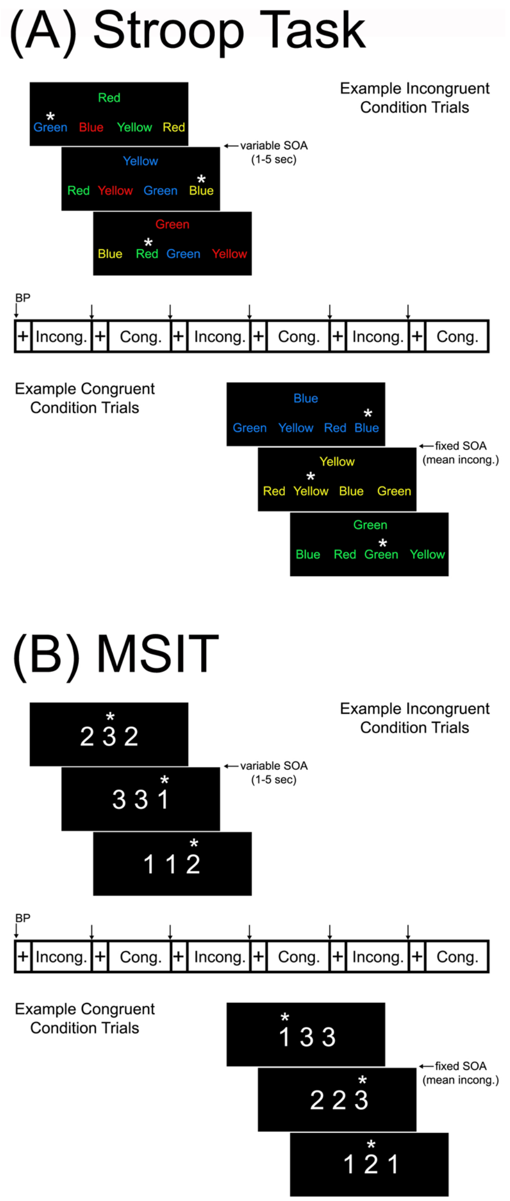

Figure 1.

An illustration of the (A) Stroop task and (B) multi-source interference task (MSIT) paradigms. Shown for both tasks are sample trials from the incongruent (top) and congruent (bottom) task conditions, as administered in a blocked fMRI paradigm. Participants completed four blocks of the two conditions in each task, with each block lasting 54–58 sec and beginning with a 12–16 sec fixation cue. For the incongruent condition of each task, trials were presented with a variable stimulus onset asynchrony (SOA) to maintain task accuracy across participants to control for individual differences in task performance. For the congruent condition of both tasks, trials were presented at a fixed SOA, defined by the mean SOA of the preceding incongruent condition. A brachial blood pressure cuff was inflated during each fixation period, and readings of oscillometric blood pressure and heart rate were obtained by the end of each congruent and incongruent condition in both tasks.

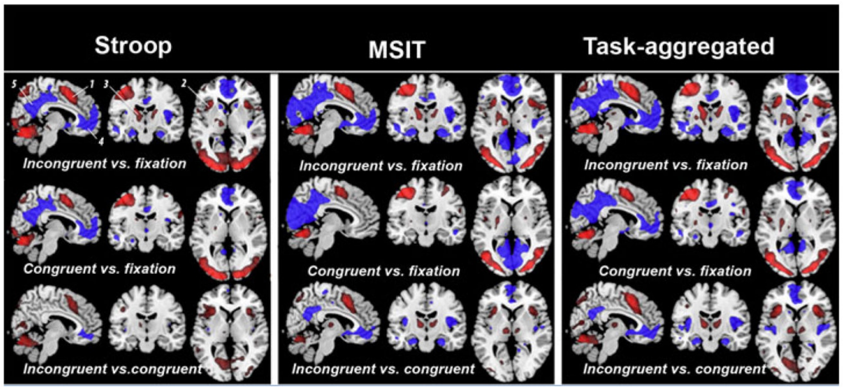

Figure 2.

Regions of interest (ROIs) for ICC analyses of each task and contrast. Shown are anatomical slices in the coordinate space of the Montreal Neuroimaging Institute: x = 6, y = 13, and z = 4. The ROIs represent brain areas engaged by the task, which were determined in a larger sample of 138 participants to decrease selection bias. The activated (red) and deactivated (blue) regions were identified for each task and contrast by a random-effects analysis employing a corrected whole-brain and family-wise error rate threshold of $p < 0.05$, combined with a cluster extent threshold of $k > 20$ voxels. The Stroop task and the MSIT exhibited comparable or spatially overlapping ‘activated’ and ‘deactivated’ regions for the incongruent vs. fixation contrast. These regions included the: (1) dorsal anterior cingulate cortex (dACC, BA32), (2) anterior insula (BA13), (3) thalamus, (4) perigenual anterior cingulate cortex (pACC, BA32), and (5) posterior cingulate cortex (pCC, BA31). More detailed labeling of these and other regions (peak MNI coordinates, t-values, and cluster sizes are in Tables S1–S2).

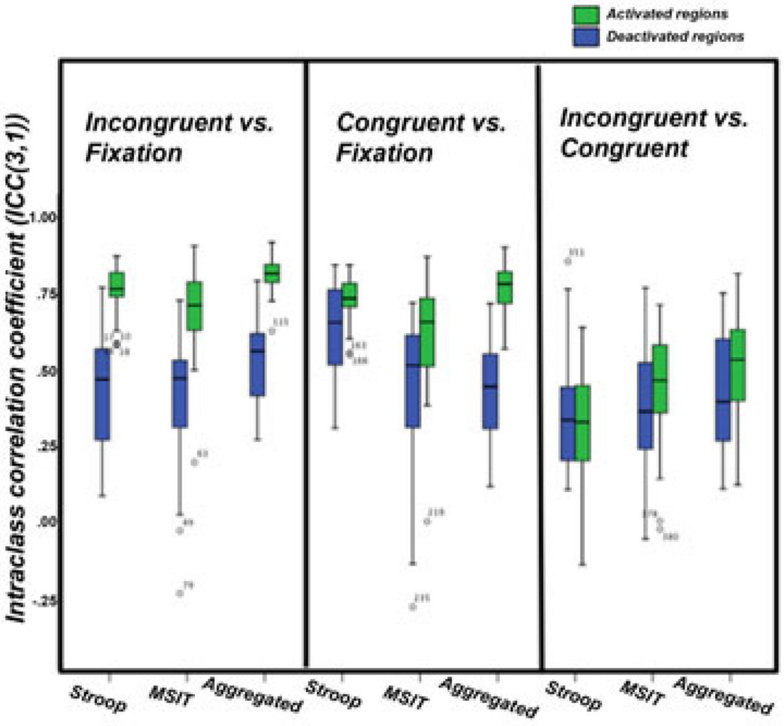

Figure 3.

Boxplots of the distributions of within-individual ICCs among participants for incongruent vs. fixation, congruent vs. fixation, and incongruent vs. congruent contrasts in the activated and deactivated regions for the Stroop task, MSIT, and the aggregated (or average) contrasts across both tasks. Note that ICCs can range from −1 to 1, where negative ICCs can be attributed to more variation within than between voxels.

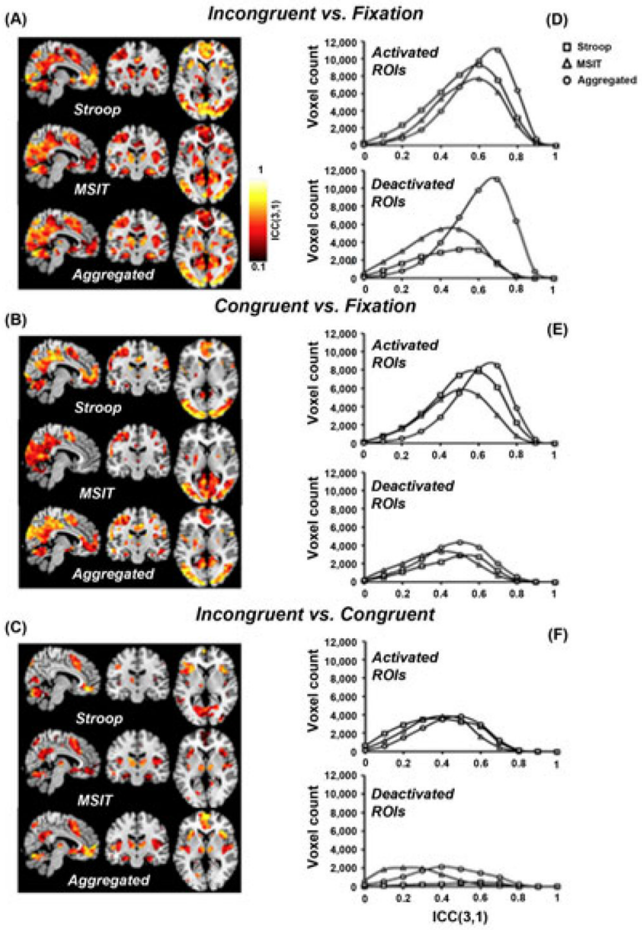

Figure 4.

ICC maps corresponding to between-individual reliability (consistency) in the ROIs for (A) the incongruent vs. fixation contrast, (B) the congruent vs. fixation contrast, and (C) the incongruent vs. congruent contrast. Shown in the figure are MNI slices x = −6, y = −13, and z = 5. The distributions of ICCs for each contrast in ROIs generated from the Stroop task, MSIT and the average of the two tasks are plotted in (D)–(F).

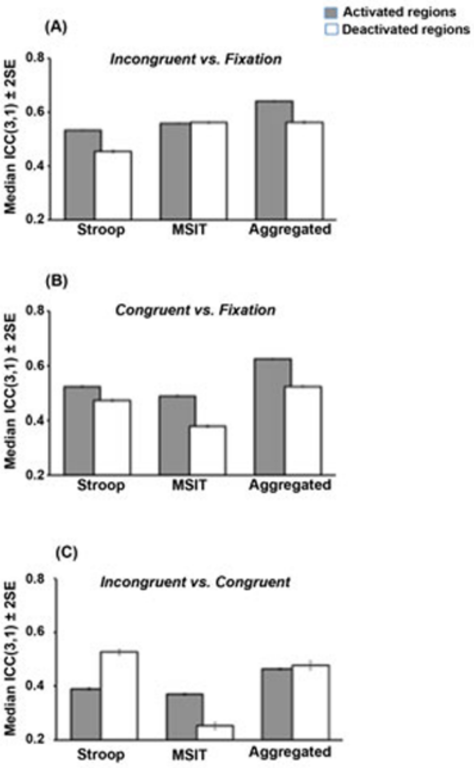

Figure 5.

A comparison of the median ICCs corresponding to between-individual reliability (consistency) estimates in areas showing relative activation and deactivation to the Stroop task, MSIT, and the average of the two tasks for the contrasts of (A) incongruent vs. fixation, (B) congruent vs. fixation, and (C) incongruent vs. congruent.

Table 1

Summary of participants’ subjective rankings of arousal, control, and valence across experimental periods, as well as cardiovascular responses to the incongruent conditions of the Stroop task and multi-source interference task (MSIT). The intraclass correlation coefficients (ICCs, model 3,1) reflect the consistency of the measures among participants across two sessions.

	1st Session		2nd Session		ICC (3,1)
	Mean (SE)	Mean Changes (SE)	Mean (SE)	Mean Changes (SE)	
Arousal (1–9)					
Baseline	2.62 (0.36)	--	2.21 (0.37)	--	--
Stroop	5.72 (0.32)	3.04 (0.43)	5.17 (0.33)	2.96 (0.41)	0.46 *
MSIT	5.19 (0.34)	2.58 (0.41)	4.63 (0.40)	2.42 (0.35)	0.63 **
Aggregated	5.44 (1.48)	2.83 (0.38)	4.90 (1.59)	2.69 (0.34)	0.57 **
Control (1–9)					
Baseline	6.08 (0.40)	--	6.17 (0.53)	--	--
Stroop	3.84 (0.34)	−2.12 (0.45)	5.25 (0.39)	−0.92 (0.52)	0.68 **
MSIT	4.35 (0.36)	−1.73 (0.49)	5.58 (0.39)	−0.58 (0.47)	0.77 **
Aggregated	4.15 (1.65)	−1.92 (0.44)	5.42 (1.82)	−0.75 (0.48)	0.82 **
Valence (1–9)					
Baseline	6.65 (0.25)	--	7.21 (0.31)	--	--
Stroop	4.88 (0.27)	−1.68 (0.28)	5.58 (0.33)	−1.63 (0.33)	0.36 *
MSIT	5.27 (0.29)	−1.38 (0.32)	5.75 (0.34)	−1.46 (0.38)	0.65 *
Aggregated	5.08 (1.19)	−1.58 (0.26)	5.67 (1.58)	−1.54 (0.35)	0.60 **
SBP (mmHg)					
Baseline	126.42 (2.68)	--	122.69 (2.14)	--	--
Stroop	131.03 (2.68)	4.61 (1.45)	126.44 (2.24)	3.50 (0.85)	0.85 **
MSIT	132.25 (3.02)	5.83 (1.35)	124.98 (2.20)	2.14 (0.85)	0.75 **
Aggregated	131.64 (2.70)	5.22 (1.33)	125.71 (2.19)	2.82 (0.77)	0.84 **
HR (bpm)					
Baseline	66.15 (2.68)	--	64.77 (2.21)	--	--
Stroop	73.11 (2.68)	9.31 (1.33)	73.11 (2.32)	8.34 (0.97)	0.85 **

	1st Session		2nd Session		ICC (3,1)
	Mean (SE)	Mean Changes (SE)	Mean (SE)	Mean Changes (SE)	
MSIT	73.71 (3.02)	7.56 (0.85)	69.57 (2.19)	4.80 (0.94)	0.82 **
Aggregated	74.59 (2.70)	8.43 (1.09)	71.34 (2.21)	6.57 (0.85)	0.87 **

* *p*<.05;

** *p*<.01

Table 2

Summary of intraclass correlation coefficient (ICC) analysis for the incongruent vs. fixation, congruent vs. fixation, and incongruent vs. congruent contrasts of the Stroop task, the multi-source interference task (MSIT) and the task-aggregated measures in activated and deactivated regions (ROIs) across two sessions

Contrast	Task	Intra-voxel ICC for group t-map		Median intra-voxel ICC for contrast map		Median intra-subject ICC for contrast map	
		Activated ROI	Deactivated ROI	Activated ROI	Deactivated ROI	Activated ROI	Deactivated ROI
Incongurent vs. Fixation	Stroop	0.82	0.70	0.77	0.47	0.53	0.45
	MSIT	0.74	0.67	0.72	0.46	0.56	0.56
	Aggregated	0.81	0.76	0.82	0.56	0.64	0.56
Congruent vs. Fixation	Stroop	0.81	0.91	0.74	0.66	0.52	0.47
	MSIT	0.77	0.51	0.64	0.49	0.49	0.38
	Aggregated	0.83	0.80	0.78	0.46	0.62	0.52
Incongruent vs. Congruent	Stroop	0.52	0.43	0.33	0.34	0.39	0.53
	MSIT	0.39	0.39	0.47	0.37	0.37	0.25
	Aggregated	0.39	0.75	0.53	0.41	0.46	0.48