Reproducibility of Changes in Behaviour and fMRI Activation Associated with Sleep Deprivation in a Working Memory Task

Julian Lim, BSc; Wei-Chieh Choo, MBBS; Michael W. L. Chee, MBBS

Cognitive Neuroscience Laboratory, Duke-NUS Graduate Medical School, Singapore

Study Objectives: Although the stability of inter-individual differences in vulnerability to sleep deprivation has been shown behaviourally, the neural basis for these differences has yet to be uncovered. In this study, we assessed the reproducibility of fMRI activation and performance on a working memory task before and after 24 hours of sleep deprivation (SD).

Design: All volunteers underwent 2 sessions (pairs of fMRI scans) at rested wakefulness (RW) and after SD.

Participants: 19 healthy, right-handed subjects (mean age = 21.37 ± 1.54 years)

Measurements and Results: Brain activation was highly correlated across sessions in a frontoparietal network previously implicated in working memory function. The magnitude of decline in this activation after SD was preserved in bilateral parietal regions. Among several behavioural metrics investigated, the most robust marker of vulnerability to SD was the change in the intra-individual variability of reaction times. This was shown to be both stable over time and correlated with the drop in left parietal activation from RW to SD in both experimental sessions.

Conclusions: Because of its reproducibility, the modulation of parietal activation may provide a good physiological marker of vulnerability to SD.

Keywords: Sleep deprivation, fMRI, working memory, reproducibility

Citation: Lim J; Choo WC; Chee MWL. Reproducibility of changes in behaviour and fMRI activation associated with sleep deprivation in a working memory task. *SLEEP* 2007;30(1):61-70.

INTRODUCTION

IN THE DOMAIN OF SLEEP RESEARCH, THERE HAS BEEN RECENT INTEREST IN THE LARGE INTERINDIVIDUAL DIFFERENCES IN BEHAVIOUR AND NEURAL activity displayed by human subjects after periods of sleep deprivation.[1,2] Objective and subjective measures obtained from volunteers undergoing multisession, total sleep deprivation (SD) studies indicate that these interindividual differences are both significant and stable across time.[1,3,4] Moreover, vulnerability to SD on objective measures does not appear to covary with vulnerability on subjective measures. Leproult and colleagues[4] demonstrated intersession consistency in subjective, objective, and EEG measures of alertness over 2 separate nights of SD. However, they found no relationship between subjective and objective components within each experimental session, prompting them to propose a multicomponent model of fatigue. Van Dongen et al.[1] also found that interindividual differences in behavioural measures after SD, though trait-like, tended to cluster along 3 independent dimensions: subjective measures, objective measures of sustained attention, and cognitive processing ability.

Several single-session fMRI studies have been able to uncover interindividual variation in neural activity and task performance relating to sleep deprivation. Mu et al.[5] administered the Sternberg working memory task to subjects who had undergone 30 hours of total sleep deprivation. They found both global and task-specific decreases in brain activation after SD, and were also able to show a relationship between parietal lobe activation and reaction times in the sleep-deprived state. Chee et al.[6] reported that SD-resistant (SD-R) and SD-vulnerable (SD-V) groups (as measured by performance decline on a working memory task) showed significantly different levels of parietal activation at rested wakefulness (RW). Recently, Chuah et al.[7] found that SD-V and SD-R subjects showed different patterns of responses across state in the right ventrolateral prefrontal cortex and insula when performing an inhibitory (Go/No-Go) task.

Although the changes in brain activity associated with sleep deprivation have been fairly well documented, to date, no imaging study has been able to provide evidence that these changes are stable across periods of time. Previous reproducibility experiments using fMRI have typically invoked simple sensory[8-11] or motor[8,9,11] paradigms or examined extent of hemispheric lateralization of specialized functions (e.g., language[12]). Further, these experiments have all taken measurements in just a single state. While these studies have shown that intersession correspondence in Blood Oxygenation Level Dependent (BOLD) activation is reasonable at the group level, many of them did not consider the reproducibility of results on a subject-by-subject basis.

More recently, researchers have begun to evaluate the reproducibility of brain activation in more complex paradigms that engage higher-level cognitive functions. Kiehl and colleagues[13] found that hemodynamic responses elicited by novel and target stimuli in an auditory oddball paradigm were reproducible over a 6-week test-retest period. Aron et al.[14] investigated the long-term reliability of fMRI over a year using a classification-learning paradigm, and found that task-related activation in midbrain and frontostriatal regions were extremely reliable, with intraclass correlation coefficients (ICCs) ranging from 0.76 to 0.99.

As far as we are aware, no previous fMRI experiment to date has tested the reproducibility of behaviour and brain activation over both session and state. The first goal of the present study, therefore, was to determine whether measures of performance on the working memory task, as well as self-reported mood and sleepiness, are all trait-like and reproducible.[1]

Disclosure Statement

This was not an industry supported study. Drs. Lim, Choo, and Chee have indicated no financial conflicts of interest.

Submitted for publication July 12, 2006

Accepted for publication September 26, 2006

Address correspondence to: Dr. Michael Chee, Cognitive Neuroscience Laboratory, Duke-NUS Graduate Medical School and Singapore Health Services Pte Ltd, 7 Hospital Drive, #01-11, Block B, SingHealth Research Facilities, Singapore 169611; Tel: (65) 63266915; Fax: (65) 62246386; E-mail: mchee@pacific.net.sg

Second, we wished to investigate whether previously observed frontoparietal brain activation to a Sternberg-like working memory task is stable over time, both during RW and after SD.

Finally, we wanted to evaluate the reproducibility of the decline in task related brain activation following SD. This final aim is closely related to the second. In the context of SD research, demonstrating a low level of within-subject variability in BOLD activation in a particular brain region involves showing that:

(1) Intersession activation is correlated at RW

(2) Intersession activation is correlated at SD, and

(3) The change in activation across state is stable over time.

By uncovering the regions in which these 3 associations hold true, we sought to lay the groundwork for future interventional fMRI studies involving countermeasures against the deleterious cognitive effects of sleep deprivation.

METHODS

Nineteen healthy, right-handed adults, aged between 19 and 24 years (mean age = 21.37 ± 1.54 years) participated in this study after giving informed consent. To recruit these volunteers, we asked subjects who had undergone the same protocol from a previous 35-hour SD study[6] if they were willing to take part in a replicate experiment. Of the 26 volunteers we contacted, 13 gave a positive reply; the rest either could not be contacted or declined to take part. We deemed that more experimental power was needed to yield robust results, and thus recruited 6 additional participants through advertisements on a local university website. Volunteers were prescreened via an online questionnaire to ensure that they:

a. Slept an average of 6.5 - 9 hours per night;
b. Had a score of 22 or less on a Morningness-Eveningness Scale.[15]
c. Had regular sleeping hours (sleep time: before 01:00 and wake time: before 09:00);
d. Had no history of excessive daytime sleepiness or insomnia;
e. Were free from psychiatric illnesses, obstructive sleep apnea, narcolepsy, and periodic leg movements;
f. Had no history of recreational drug use or excessive alcohol consumption;
g. Consumed no more than 2 cups of coffee (or an equivalent amount of caffeine) a day;
h. Had no history of psychoactive drug use for 3 months prior to the study.

In order to ensure compliance with regular sleeping patterns (as defined by conditions a. and c. in the prescreening criteria above), we used actigraphy (Mini Mitter Actiwatch, model AWLP) to record subjects' sleep-wake patterns for at least one week prior to each fMRI scan. Subjects also kept a sleep diary in which all sleep episodes were recorded. Examination of the sleep data showed that all subjects had satisfactory sleep patterns and were compliant with the study rules during the periods when these data were recorded.

The 13 volunteers who were recalled from the previous study underwent a near-identical protocol to the one described by Chee et al.[6] The procedure for this repeated experiment, as well as the original protocol, can be seen in Figure 1. Briefly, participants in the original study underwent a total of 3 scans while performing the identical working memory task described in detail below: one in the RW state, one after 24 hours of SD (SD24), and one after 35 hours of SD (SD35).

In the current study, subjects underwent an initial briefing when they were re-trained on the working memory task, following which they returned to the lab for 2 subsequent scans, once during RW and once after SD24. Scans were conducted approximately one week apart, and, to ensure that subjects who underwent the SD scan first had minimal residual sleep debt during the RW scan, were separated by no less than 5 days. Among these 13 subjects, 5 had done their RW scan first in the original experiment; to negate order effects when comparing individual subject data from session to session, we scanned participants in this same RW-SD sequence that they underwent originally.

The 6 remaining volunteers went through 2 sets of RW-SD scans in the same order, with a minimum of 39 days between each set (mean: 44 ± 4.24). To achieve proper counterbalancing, we assigned 4 of these participants to undergo the RW scan first. Although there was substantial variability in the time between pairs of scans across all 19 subjects (standard deviation: 104.27 days), we deemed that this would not affect our experimental results, since variables that show a high intraclass coefficient correlation should remain constant irrespective of the intersession interval. To recapitulate the counterbalancing procedure in brief: in the entire pool of subjects, 9 underwent the RW scan first, and did so for both session 1 and session 2.

Prior to SD scans, subjects came into the lab at 19:00 and were monitored throughout the night to ensure they did not fall asleep. Volunteers were allowed to engage in nonstrenuous activities (reading, homework, watching DVDs, conversing), and were permitted to eat light snacks during the night. Every hour starting from 20:00, subjects performed 10 minutes of the Psychomotor Vigilance Test (PVT).[16] They also rated their subjective sleepiness on the 9-point Karolinska Sleepiness Scale (KSS),[17] and their mood on 6 given dimensions (motivated – unmotivated; elated – depressed; fresh – exhausted; congenial – irritable; relaxed – stressed; calm - anxious) on a 10-point Likert-type scale. Following SD, scanning took place between 06:00 and 08:00 after subjects had completed the KSS and 10-minute PVT for the respective hour. RW scans took place between 08:30 and 10:00 after subjects had completed the KSS and 10-minute PVT. Volunteers were prohibited from smoking and consuming caffeine or other stimulants for a 24-hour period prior to both the RW and the SD fMRI scan.

Because our original study involved 35 hours of continuous SD, and as expectation effects (i.e., subjects anticipating the end of the experiment) are known to modulate behaviour,[18] all volunteers were informed that they would be going through the entire (35-hour) SD protocol, even though we were only interested in reproducing the effects at RW and after SD24. The 6 naïve subjects completed the entire 35-hour SD protocol during their first RW-SD pair of scans, so that their treatment in session 1 would be identical to the other 13 subjects. After the second-session SD24 scan, all subjects were told that the experiment was over and were allowed to leave the lab.

We made one other minor change to the experimental paradigm by replacing the modified Psychomotor Vigilance Test[19] with the test in its original form. Aside from this, all questionnaires and behavioural tests given to subjects were the same as in Chee et al.[6]

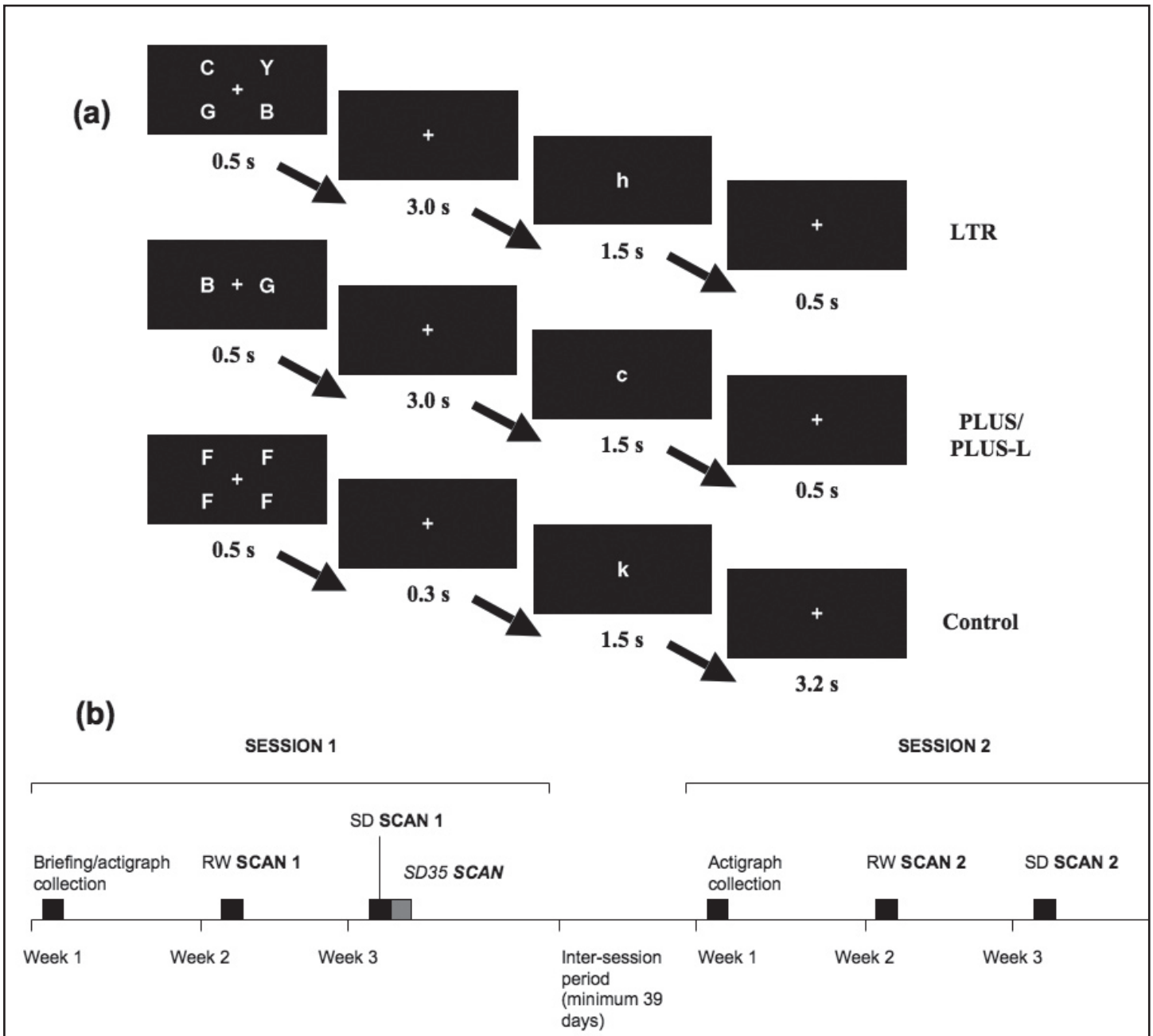

Figure 1—(a) Examples of stimuli used in the in-scanner working memory tasks. The timings indicate the duration of exposure of each stimulus. PLUS and PLUS-L trials differed in the nature of the probe: 50% of the PLUS-L probes were lures (i.e., 1 of the 2 remembered stimulus letters). (b) Schematic showing 1 possible order of the complete experimental protocol. Although all subjects went through an SD35 scan in session 1 (shown in gray), these data were *not* analysed in the present study. Approximately half the subjects did the other order (SD scan in Week 2, RW scan in Week 3). Counterbalancing was done across subjects, but not sessions—a subject undergoing the RW treatment first in session 1 would do the same in session 2. Note that in the current analysis, we took data from 4 scans (excluding the SD35 scan in session 1) which took place in 2 experimental *sessions*.

In-Scanner Experimental Task

For the in-scanner experimental paradigm, we used three working memory tasks in a block design (Figure 1). The tasks were:

1) LTR task: subjects were presented with 4 different uppercase letters for 0.5 seconds. They then viewed a fixation cross for 3.0 s before the appearance of a lowercase probe letter for 1.5 s. Subjects were required to respond to this probe (by indicating a match or a nonmatch to the 4 target letters) while it was still on the screen.
2) PLUS task: subjects viewed 2 different uppercase letters and were required to mentally shift each letter one forward in the alphabet. They then had to indicate whether the probe presented matched either of these new letters. The timing and sequence of presentation were identical to that of the LTR task.
3) PLUS-L task: this was identical to the PLUS task, except that in the nonmatch trials the probe letter was the same as one of the 2 target stimuli. Successful performance in this task was more challenging, as it required subjects to inhibit their prepotent tendency to judge lure trials as matches.

Subjects were instructed to respond to all probes as quickly and as accurately as they could. Each experimental run contained

Table 1—Behavioural means (and standard deviations) from the in-scanner experiment, averaged across all four runs.

	RW		SD	
	S1	S2	S1	S2
Accuracy (%)				
LTR	94.6 (4.83)	90.68 (7.33)	81.47 (13.0)	71.49 (14.9)
PLUS	93.97 (4.70)	91.89 (6.59)	82.46 (12.0)	70.72 (15.2)
PLUS-L	93.75 (6.63)	89.14 (7.99)	79.61 (11.8)	70.83 (13.0)
Average	94.12 (5.39)	90.57 (7.30)	81.18 (12.3)	71.02 (14.3)
Reaction time (ms)				
LTR	761 (101)	777 (80)	809 (84)	837 (120)
PLUS	705 (109)	744 (104)	772 (102)	798 (128)
PLUS-L	747 (104)	771 (107)	809 (95)	829 (107)
Average	738 (105)	764 (97)	797 (94)	822 (118)

6 task blocks alternating with 7 control blocks in a pseudorandomised fashion. Task and control blocks lasted 33 seconds each, and each run lasted 7 minutes 9 seconds in total (excluding an initial fixation period of 12 seconds). We hereafter refer to this task in its entirety as the LTR-PLUS task.

Imaging Procedure

Images were acquired on a 3T Allegra MRI system (Siemens, Erlangen, Germany) using a gradient-echo EPI sequence with the following parameters: TR = 3000 ms, FOV = 192 x 192 mm and 64 x 64 mm pixel matrix. Thirty-six axial slices of thickness 3 mm (0.3 mm gap) approximately parallel to the AC-PC line were collected. To ensure that subjects were positioned as consistently as possible from session to session, we used bitmap screenshots of the T2-localizer images from each subject's first scanning session as a reference for every subsequent scan. In addition, to facilitate appropriate registration of functional and anatomical images, we obtained a high-resolution coplanar T1-weighted anatomical image after the functional runs. A T1-weighted 3D-MPRAGE sequence was also acquired to allow for image display in Talairach space.[20]

Stimuli were projected onto a screen using an LCD projector, and subjects viewed these through a mirror positioned above them on the head coil. Subjects were instructed to press a button with their right index finger to signal a match and another button with their right middle finger to indicate a nonmatch. A bite-bar and firm foam padding were used to minimize head movement. We also eliminated the need for volunteers to talk between runs by having them use their button boxes to answer yes-no questions and to indicate their subjective sleepiness ratings.

Prior to being scanned, subjects performed 1 practice run of the working memory task. To acclimatize them to the scanning environment, they did 1 additional practice run inside the MRI scanner before doing 4 experimental runs. After being removed from the scanner, subjects rated how well they thought they did on the working memory task and rated their mood and subjective sleepiness one final time.

Image Analysis

Functional images were processed using Brain Voyager QX (version 1.52) (Brain Innovation, Maastricht, Netherlands). Motion correction was performed by aligning all images to the first image of the final functional run (i.e., the run directly preceding the acquisition of the coplanar T1-weighted anatomical image).

Table 2—Means (and standard deviations) of the distances (in mm) between peak activated voxels (PAV) in task-related ROIs between S1 and S2.

Region	Distance between PAVs (S1 and S2)
Left parietal (n= 18)*	6.0 (3.7)
Right parietal (n=11)	8.0 (3.3)
Left prefrontal (n=18)	4.6 (2.2)
Anterior cingulate (n=18)	4.8 (3.0)

*The number of values for each ROI varies because certain subjects did not show significant activation in these regions in one or both sessions.

No subject showed more than 2 mm (translational) or 2° (rotational) movement over the course of the 4 functional runs. Linear interpolation was used to correct for interslice timing differences due to acquisition order during each TR. Gaussian filtering was applied in the spatial domain using a smoothing kernel of 4 mm FWHM for individual subject maps and 8 mm FWHM for group-level activation maps. Finally, to eliminate low frequency noise, a temporal high pass filter of 0.007Hz was applied following linear trend removal.

We analysed the functional imaging data using a general linear model with 12 predictors of interest (all combinations of the following -- 2 SESSION conditions (session being defined as a pair of RW-SD scans): S1, S2; 2 STATE conditions: RW, SD; and 3 TASK conditions: LTR, PLUS, and PLUS-L). For this model, functional imaging data from the RW and SD24 states collected in our original dataset (Chee et al, 2006) were combined with the data obtained in the retest (in the case of the 6 new subjects, data from all 4 scans were newly collected). The fMRI data collected after 35 hours of SD (in our original experiment) were not included in this analysis.

In order to perform intersession comparisons, we used the 5 task-related regions-of-interest (ROIs) from our previous analysis[6] to obtain parameter estimates of activation in all conditions. These regions were selected based on a 3-way conjunction map (across 26 subjects) of all conditions (LTR, PLUS and PLUS-L) at the RW state in the original experiment, thresholded at $P < 0.001$ (Bonferroni corrected), and were: left parietal (LPT) (-27, -64, 37) left prefrontal (LPFC) (-39, 7, 31), right parietal (RPT) (30, -55, 34), left thalamus (LTHAL) (-15, -22, 13), and anterior cingulate (ACC) (-5, -1, 55). Although there was some degree of spatial separation between these ROIs and regions conjointly activated by the 19 subjects in the 2 sessions of this experiment, the approach used is a more unbiased way to extract measures of intersession reliability than picking separate ROIs for S1 and S2. Parameter estimates were compared between S1 and S2 using bivariate correlations. In addition, we calculated between-subject ICCs for the 4 areas showing consistency of activation from S1 to S2, since it has been suggested that computing the ICC of signal change is a more powerful method of assessing test-retest reliability than the comparison of thresholded activation maps.[14] The ICC is a way of expressing the proportion of variance in the data accounted for by interindividual variability. Here, ICC was computed using a 2-way random effects model as $(MS_b - MS_{res})/(MS_b + MS_{res}) + (2[MS_{sess} - MS_{res}]/n)$, where MS_b is the between subject mean square error, MS_{res} is the residual error, MS_{sess} is the mean square error for session, and n is the number of observations (19); this value was calculated for intrastate as well as intrasession pa-

Table 3—Bivariate correlations of parameter estimates of activation across sessions in functional ROIs.

Region	RW		SD		RW - SD	
	r value	Sig. level	r value	Sig. level	r value	Sig. level
Left parietal	0.74	P < 0.001	0.77	P < 0.001	0.50	P = 0.03
Right parietal	0.61	P = 0.005	0.80	P < 0.001	0.50	P = 0.03
Left prefrontal	0.74	P < 0.001	0.65	P = 0.002	0.23	N.S.
Anterior cingulate	0.65	P = 0.002	0.60	P = 0.007	0.28	N.S.
Left thalamus	0.07	N.S.	0.08	N.S.	-0.16	N.S.

Table 4—Intraclass correlation coefficient (ICC) values of parameter estimates of activation in four ROIs (for individual states and changes across state). ICCs were computed as $(MS_b - MS_{res})/(MS_b + MS_{res}) + (2(MS_{sess} - MS_{res})/n)$, where MS_b is the between-subject mean square error, MSres is the residual error, MS_{sess} is the mean square error for session, and n is the number of observations (19).

Region	ICC		
	RW	SD	RW-SD
Left parietal	0.69	0.68	0.49
Right parietal	0.58	0.62	0.46
Left prefrontal	0.67	0.58	*
Anterior cingulate	0.58	0.59	*

* ICCs were not computed as Wald-Z tests for between-subject variance were nonsignificant in these regions.

rameter estimates. Finally, to assess the spatial similarities in activation, we computed mean Euclidean distance between the peaks of activation (PAVs) in the 5 ROIs using individual subjects' activation maps. All statistical computation was performed using Statistical Package for the Social Sciences (SPSS) 13.0 (SPSS Inc., Chicago, IL).

RESULTS

Behaviour

We performed cross-correlations on the time courses of both the KSS and self-reported mood ratings from 20:00 to 06:00 on the nights that subjects underwent SD. Cross correlations were at a maximum for both scales when there was no time lag (KSS: r = 0.96, P < 0.001; mood: r = 0.99, P < 0.001), indicating that both subjective sleepiness and mood declined in parallel fashion during S1 and S2. Correspondingly, performance on the PVT, as measured by the number of lapses (response > 500 ms after stimulus onset) and the difference between the 90th and 10th percentile RT, worsened throughout the night in both experimental sessions. As we had changed the version of the PVT used from S1 to S2, we could not quantitatively compare these two sets of data (the trendlines were similar).

Raw behavioural scores of accuracy and reaction time to the task are reported in Table 1. In both sessions, in-scanner task performance worsened significantly from RW to SD, whether this was measured by accuracy decline or reaction time increase (RT_{S1}: t(18) = -3.89, P = 0.001; RT_{S2}: t(18) = -2.92, P = 0.009; $accuracy_{S1}$: t(18) = 4.66, P <0.001; $accuracy_{S2}$: t(18) = 5.80, P < 0.001). There was no significant difference in accuracy across subjects between the LTR, PLUS, and PLUS-L tasks ($F_{2,17}$ = 1.04, P = 0.38) Accuracy differed greatly between sessions, with no correlation in either intra- or inter-state comparisons (RW: r = 0.17, P = 0.49; SD: r = 0.40, P = 0.09; change over state: r = 0.11, P = 0.66). Because differences in accuracy could be attributed to 2 sources (lapses—when subjects failed to respond while the probe letter was on the screen—and mistakes), we did a further analysis of these to determine which variable was a greater contributor to intersession variance. Overall, mistakes accounted for slightly more than half (55.3% ± 6.7) of the error rate. There were no significant correlations between mistakes or lapses in either state across sessions.

We next looked at the intersession correspondence of reaction times to each of the target stimuli. After removing all lapse trials, we conducted a repeated-measures ANOVA of raw reaction times using the same model as for activations. Main effects were significant for SESSION ($F_{1,18}$ = 5.45, P = 0.03), STATE ($F_{1,18}$ = 24.55, P < 0.001), and TASK ($F_{2,17}$ = 20.43, P < 0.001). There were no significant interaction effects. Reaction times for each subject were then averaged across tasks and correlated across session; these were found to be consistent, but only within each state (RW: r = 0.79, P < 0.001; SD: r = 0.68, P = 0.001).

Previous studies have suggested that, when analyzing reaction time data, the intra-individual coefficient of variation (ICV) of reaction times (calculated as the standard deviation of RT over the mean) is a more valid measure of intra-individual performance than simple mean RT.[7,21,22] After removing the trials on which subjects lapsed, we calculated ICVs for each experimental session (across all tasks) and tested for intersession reproducibility of this measure. As with mean reaction times, intrastate correspondence of ICVs across session was good (RW: r = 0.60, P = 0.007; SD: r = 0.82, P < 0.001). In addition, the normalized change in ICV across state, calculated as $(ICV_{RW} - ICV_{SD})/ ICV_{RW}$, was significantly correlated from S1 to S2 (R = 0.67, P = 0.002), and had an ICC (across sessions) of 0.48.

Functional Activation

The 2 (SESSION) x 2 (STATE) x 3 (TASK) repeated measures ANOVA performed on activation values in the left parietal ROI showed main effects of TASK ($F_{2,17}$ = 12.78, P < 0.001), STATE, ($F_{1,18}$ = 11.23, P = 0.001), and SESSION ($F_{1,18}$ = 15.30, P = 0.004). Similar results were found in the right parietal lobe and left prefrontal cortex. There were no interactions among any of the conditions in any of these ROIs; thus, for simplicity and to improve power, activation values were collapsed across tasks for all subsequent analysis.

To measure the reproducibility of BOLD response from session to session, we interrogated the 5 task-related ROIs in individual subject maps to find the peak activated voxels (PAV) in each session for all individuals. Mean Euclidean distances between peaks in S1 and S2 are summarized in Table 2. PAV distances were all less than 1 cm. All ROIs were robustly activated in most individual activation maps (thresholded at P < 0.001) with the exception of the left thalamus, which failed to activate significantly in

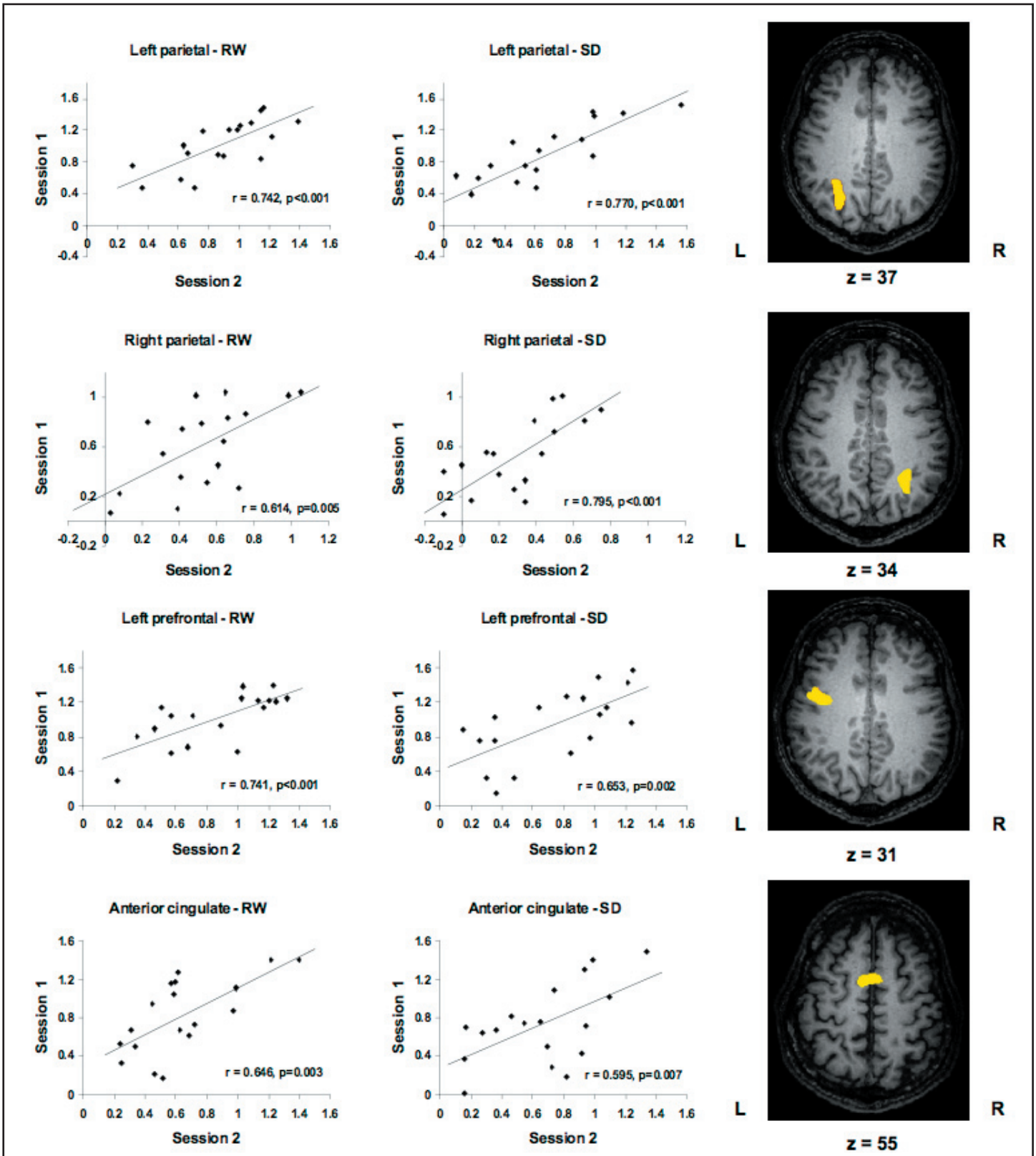

Figure 2—Correlations of parameter estimates of activation in task-related ROIs for session 1 and session 2

one or both sessions in all but 2 subjects, and was thus excluded from the table.

When compared across sessions, activations within each state were significantly correlated in 4 out of the 5 selected ROIs, the only exception being the left thalamus (Figure 2; Table 3). In addition, the magnitude of the change in activation across state (RW – SD) was preserved bilaterally in parietal ROIs; these were the only areas in which the change was robustly reproducible across Session 1 and 2 (Table 3).

The intersession correspondence in activation suggests that there is a trait-like component of brain function associated with performing the experimental working memory task following a period of sleep deprivation. ICC values were used to quantify this. Wald tests for between-subject variance components were significant at the 0.05 level for all ROIs, and ICC values ranged from 0.58 to 0.69 (Table 4). The ICCs for state related change in

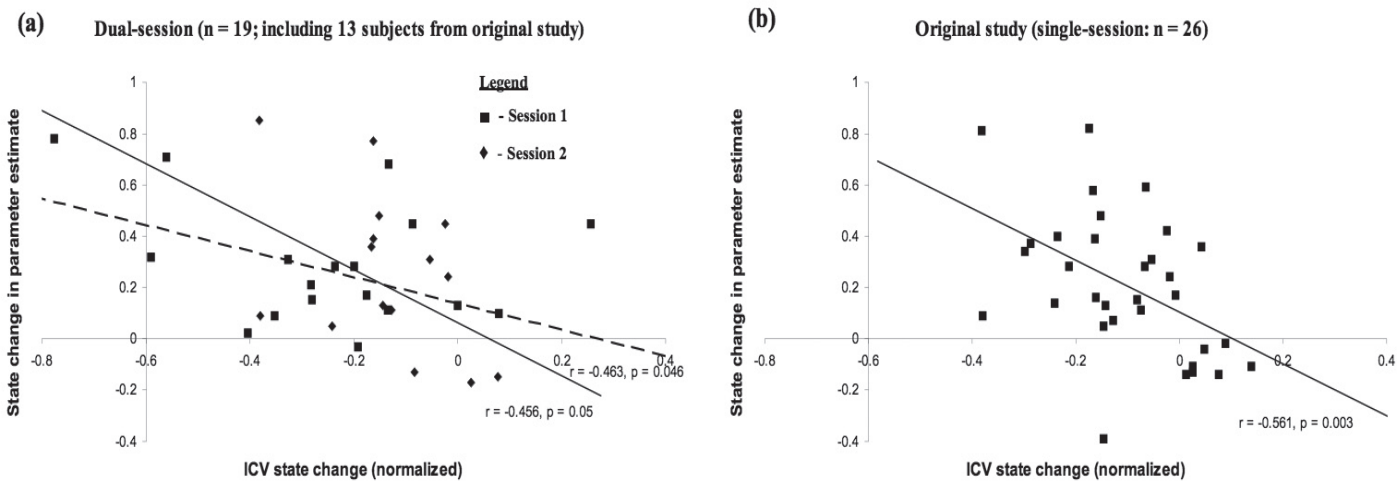

Figure 3—ICV state change is negatively correlated with the decrease in magnitude of activation in the left parietal lobe from RW to SD across subjects (a) in both sessions of the present experiment, and (b) in our original work (Chee et al., 2006). The solid and dotted lines in (a) are the linear trends for session 1 and session 2 data respectively.

activation within left and right parietal ROIs were found to be 0.49 and 0.46 respectively.

Behaviour-Activation Correlations

Because of the differences in accuracy from S1 to S2, activation in left parietal and prefrontal regions at RW were no longer significantly correlated with the drop in accuracy over state in either experimental session (S1: r = 0.40, P= 0.09; S2: r = 0.14, P = 0.56). As RT metrics appeared to show greater reliability from S1 to S2, we next chose to investigate whether these showed any relationship with changes in BOLD signal.

Within state, there was no association between frontoparietal activation and either ICV or raw averaged reaction time. However, across state, normalized ICV change was significantly correlated with the drop in activation in left parietal regions in both S1 and S2 (S1: r = -0.46, P = 0.05; S2: r = -0.46, P = 0.05) (Figure 3). We reanalyzed the data from all 26 volunteers of our previous study to see if the result was obtainable in the original data set, and found this to be the case (r = -0.56, P = 0.003)

With this result, we wish to highlight the fact that our current data demonstrate three interconnected relationships: reliability of (1) a behavioural metric (ICV change), (2) BOLD activation, and (3) an association between (1) and (2).

DISCUSSION

The results of this study support the claim that there exist both behavioural and physiological markers of vulnerability to sleep deprivation that are trait-like and reproducible over time. Time courses of self-rated sleepiness and mood were highly correlated across sessions, as were measures of response time to the LTR-PLUS task. Additionally, individual differences in the magnitude of BOLD activation across subjects, as well as the foci of activation in each of our task-related ROIs, were highly preserved from session to session. These findings complement the growing body of behavioral evidence suggesting that there is differential vulnerability to the effects of sleep deprivation between individuals,[1,23] and also point to a neural basis for these differences.

As we have noted in previous work,[6] the areas that are most consistently reported to show a decline in activation during a working memory task following SD are the intraparietal sulcus and surrounding regions of the superior parietal lobe.[5,24,25] In agreement with this corpus of research, we found that the left parietal ROI activated by the LTR-PLUS task showed the highest intersession ICCs among all the regions we investigated. Additionally, the difference in activation across states was preserved exclusively in bilateral parietal regions.

The ICC values for intersession BOLD activation we obtained in this study were lower than those found by Aron et al.[14] on a classification learning task, even when compared only at RW. Behavioral ICCs, as reported by Van Dongen et al.,[1] were also substantially higher than those seen in this experiment. However, the ICCs reported by Van Dongen were obtained by comparing averaged values taken from tests over the course of one night of sleep deprivation, whereas our comparisons were made across 2 single points in time. The reproducibility of neuroimaging findings after SD is certainly robust; however, further work is needed to assess its relative sensitivity (when compared to behavioral indices) in picking out differences in vulnerability to SD between individuals.

We found session differences in activation, with subjects in the second session showing significantly less activation than in the first. Contrast maps revealed that these differences were manifest only in task-related ROIs, indicating that the effect was not due to a systematic, global difference in signal intensities between scanning sessions. It is tempting to conclude that these changes were not due to practice effects—subjects were highly practiced on the task before they entered the scanner, and there was no significant improvement in task performance from S1 to S2. However, it is still possible that activation decreases may be associated with neural consolidation of performance[26,27] that persists even after significant gains in performance measures can be observed.

The functional significance of parietal activation in this study merits discussion. Although the regions surrounding the intraparietal sulcus are known to be involved in visual working memory task performance,[27,28] success on the 3 tasks used here (as measured by accuracy) was not predicted by the magnitude of parietal lobe activation at RW; that is, we did not replicate our original

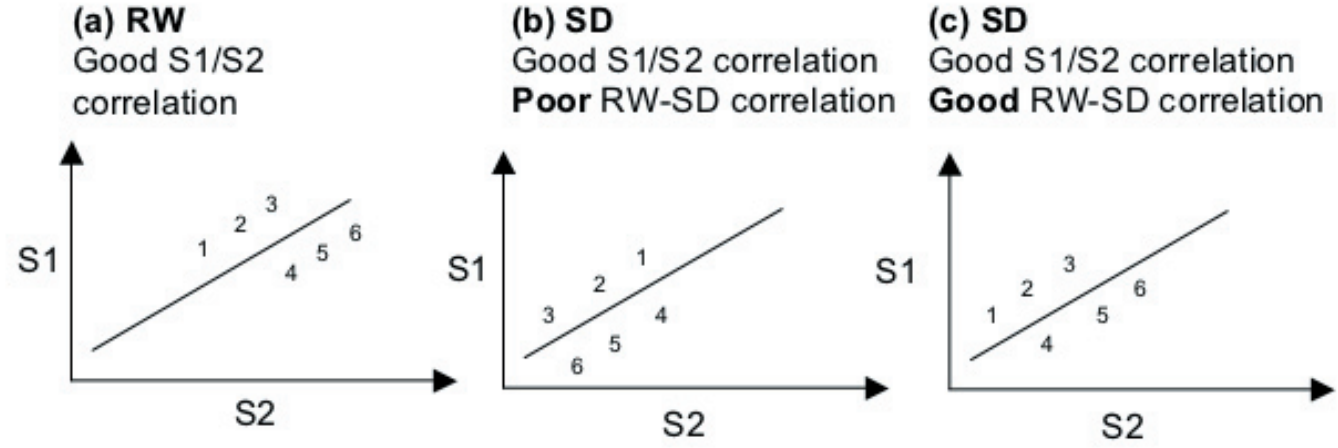

Figure 4—The graphs above illustrate how activation values can be well-correlated from session to session within states but not across them. The numbers represent individual subjects. All graphs show good intersession correlations. However, the pairing of results (a) and (b) demonstrates a poor (RW - SD) correlation (e.g. what we observe in prefrontal/anterior cingulate areas), while the pairing of (a) and (c) results in a good (RW - SD) correlation (e.g., what we observe in parietal areas).

finding that activation in a left frontoparietal network at RW was negatively correlated with the decline in performance accuracy after a period of sleep deprivation.[6] The difference in findings between the present study and the previous one was largely due to the variability of subject accuracy from S1 to S2 in the SD session and not to the neuroimaging findings between sessions. Although this difference could have been caused by experimental asymmetries in S1 and S2, this is not likely, as we were extremely careful in the way we recreated the experimental protocol.

The lack of significant correlation between RW activation and accuracy decline in S2 raises the possibility that while frontoparietal regions index working memory processing in both states, the decline in activation in these areas from RW to SD might also track some other aspect of performance that is not usually considered in studies performed only during RW.

In support of this hypothesis, we found that in both experimental sessions the change in the ICV of RTs was negatively correlated with the change in magnitude of left parietal activation from RW to SD. This was also true of the reanalysis of data from our original pool of 26 volunteers.

It may be reasonably inferred from the correlation between ICVs and parietal activation decline that, in this particular task, activity in the IPS might reflect the engagement of sustained attention—this is in addition to the working memory function previously implicated. The ability to sustain attention is a necessary prerequisite for most higher-level cognitive functioning, particularly working memory,[29] and its neural correlates are generally agreed to be in frontoparietal areas similar to those engaged by the LTR-PLUS task.[30,31] The decline in sustained attention after sleep deprivation is well documented,[16,32,33] and would manifest itself most directly in an increase in reaction time variability.

Thus, in response to the LTR-PLUS task, parietal activation may track 2 separate and dissociable processes: the ability to sustain attention as well as the engagement of working memory. By the nature of our task design, (long run length, block design), successfully engaging working memory processes following SD is heavily contingent on being able to sustain attention over the entire course of each task run. In other words, a person who is resilient to the effects of SD when it comes to memory processes but vulnerable to its effects on sustained attention may perform almost as poorly on accuracy measures as a person who is vulnerable along both dimensions. Additionally, difficulty with sustaining attention may affect subject responses in a stochastic manner, since, post-SD, attention is likely to randomly wax and wane over the course of the fMRI scan.[34]

We posit that accuracy in the LTR-PLUS task was not reproducible because it might be affected by 2 interrelated sources of variance. In contrast, mean reaction time and ICV are relatively more stable metrics because they measure a more fundamental process: sustained attention. Only if we assume that the SD-related sustained attention deficits in a given pool of subjects are equal can we unmask the correlations between our working memory measure (accuracy) and parietal lobe function. This was possibly a feature of our subjects in our original work.[6]

Regardless of which is the dominant mechanism, the present data suggest that the decrease in parietal lobe activation is a reproducible neurophysiological marker for individual differences in response to sleep deprivation. A single imaging session appears to provide comparably reliable information regarding sleep deprivation performance as behavior sampled over several points in time in the sleep deprivation period.

The present results further highlight the difficulty in uncovering reproducible relationships between brain activation and behavior across states. Figure 4 shows conceptually how activation and behavior can be correlated within states without necessarily showing a correspondence across them. Hence, in the context of multisession fMRI studies involving subjects in different states—one cannot assume that good intrastate correlations necessarily point towards good interstate reliability. Our study fills a methodological lacuna by demonstrating that regional brain activation can be consistent both within and across states. However, researchers doing multi-state fMRI studies should consider the region-specificity of these effects; in the current study, across-state (RW – SD) correlations were seen only in parietal areas, even though we observed good within-state correlations in prefrontal and cingulate areas as well. It is important that all within- and across-state measures prove consistent before activation in a brain region can be considered completely reproducible.

The results of this study suggest that fMRI may be a viable tool for use in assessing the efficacy of interventions against the

detrimental effects of SD. It is known that psychoactive substances such as caffeine,[35-37] amphetamines,[37,38] and modafinil[39,40] can temporarily reverse SD-associated cognitive deficits. Being able to track in vivo brain regions where these agents modulate cognitive function is a potentially important step towards evaluating their relative effectiveness. Although work needs to be done on improving task selection and minimizing intersession differences, the repeatability and reliability of fMRI is a hopeful sign for its usefulness in future studies of this nature.

ACKNOWLEDGMENTS

Financial support for this research was granted by DSO National Laboratories, Singapore (No. DSOCL05141) and SingHealth Foundation. The authors would like to thank Dr. David Dinges for the loan of the two PVT-192 monitors. We acknowledge the work of Shen Li-Juan and Joanna Sze in collecting and pre-processing the session 1 data, and thank Diana Teo for her help during session 2.

REFERENCES

1. Van Dongen HP, Baynard MD, Maislin G, Dinges DF. Systematic interindividual differences in neurobehavioral impairment from sleep loss: evidence of trait-like differential vulnerability. Sleep 2004;27(3):423-33.
2. Van Dongen HP, Vitellaro KM, Dinges DF. Individual differences in adult human sleep and wakefulness: Leitmotif for a research agenda. Sleep 2005;28(4):479-96.
3. Frey DJ, Badia P, Wright KPJ. Inter- and intra-individual variability in performance near the circadian nadir during sleep deprivation. J. Sleep Res. 2004;13(4):305-15.
4. Leproult R, Colecchia EF, Berardi AM, Stickgold R, Kosslyn SM, Van Cauter E. Individual differences in subjective and objective alertness during sleep deprivation are stable and unrelated. Am J Physiol Regul Integr Comp Physiol 2003;284(2):R280-90.
5. Mu Q, Nahas Z, Johnson KA, et al. Decreased cortical response to verbal working memory following sleep deprivation. Sleep 2005;28(1):55-67.
6. Chee MW, Chuah LY, Venkatraman V, Chan WY, Philip P, Dinges DF. Functional imaging of working memory following normal sleep and after 24 and 35 h of sleep deprivation: Correlations of fronto-parietal activation with performance. Neuroimage 2006;31(1):419-28.
7. Chuah LY, Venkatraman V, Dinges DF, Chee MW. The neural basis of inter-individual variability in inhibitory efficiency following sleep deprivation. J. Neurosci 2006;26(27):7156-62.
8. Cohen MS, DuBois RM. Stability, repeatability, and the expression of signal magnitude in functional magnetic resonance imaging. J Magn Reson Imaging 1999;10(1):33-40.
9. McGonigle DJ, Howseman AM, Athwal BS, Friston KJ, Frackowiak RS, Holmes AP. Variability in fMRI: an examination of intersession differences. Neuroimage 2000;11(6 Pt 1):708-34.
10. Miki A, Raz J, van Erp TG, Liu CS, Haselgrove JC, Liu GT. Reproducibility of visual activation in functional MR imaging and effects of postprocessing. AJNR Am J Neuroradiol 2000;21(5):910-5.
11. Yetkin FZ, McAuliffe TL, Cox R, Haughton VM. Test-retest precision of functional MR in sensory and motor task activation. AJNR Am J Neuroradiol 1996;17(1):95-8.
12. Rutten GJ, Ramsey NF, van Rijen PC, van Veelen CW. Reproducibility of fMRI-determined language lateralization in individual subjects. Brain Lang 2002;80(3):421-37.
13. Kiehl KA, Liddle PF. Reproducibility of the hemodynamic response to auditory oddball stimuli: a six-week test-retest study. Hum Brain Mapp 2003;18(1):42-52.
14. Aron AR, Gluck MA, Poldrack RA. Long-term test-retest reliability of functional MRI in a classification learning task. Neuroimage 2006;29(3):1000-6.
15. Horne JA, Ostberg OA. A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. Int J Chronobiol 1976;4:97-110.
16. Dinges DF, Pack F, Williams K, et al. Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4-5 hours per night. Sleep 1997;20(4):267-77.
17. Akerstedt T, Gillberg M. Subjective and objective sleepiness in the active individual. Int J Neurosci 1990;52(1-2):29-37.
18. Haslam DR. The incentive effect and sleep deprivation. Sleep 1983;6(4):362-8.
19. Dorrian J, Rogers NL, Dinges DF. Psychomotor vigilance performance: a neurocognitive assay sensitive to sleep loss. In: Kushida C, ed. Sleep deprivation: clinical issues, pharmacology and sleep loss effects. New York, NY: Marcel Dekker Inc, 2005:39-70.
20. Talairach J, Tournoux P. Co-planar stereotaxic atlas of the human brain. New York: Thieme Medical Publishers, 1988.
21. Castellanos FX, Sonuga-Barke EJ, Scheres A, Di Martino A, Hyde C, Walters JR. Varieties of attention-deficit/hyperactivity disorder-related intra-individual variability. Biol Psychiatry 2005;57(11):1416-23.
22. Stuss DT, Murphy KJ, Binns MA, Alexander MP. Staying on the job: the frontal lobes control individual performance variability. Brain 2003;126(Pt 11):2363-80.
23. Mallis MM, Mejdal S, Nguyen TT, Dinges DF. Summary of the key features of seven biomathematical models of human fatigue and performance. Aviat Space Environ Med 2004;75(3 Suppl):A4-14.
24. Drummond SP, Brown GG, Salamat JS, Gillin JC. Increasing task difficulty facilitates the cerebral compensatory response to total sleep deprivation. Sleep 2004;27(3):445-51.
25. Chee MW, Choo WC. Functional imaging of working memory after 24 hr of total sleep deprivation. J Neurosci 2004;24(19):4560-7.
26. Hempel A, Giesel FL, Garcia Caraballo NM, et al. Plasticity of cortical activation related to working memory during training. Am J Psychiatry 2004;161(4):745-7.
27. Olesen PJ, Westerberg H, Klingberg T. Increased prefrontal and parietal activity after training of working memory. Nat Neurosci 2004;7(1):75-9.
28. Pessoa L, Gutierrez E, Bandettini P, Ungerleider L. Neural correlates of visual working memory: fMRI amplitude predicts task performance. Neuron 2002;35(5):975-87.
29. Baddeley AD. Is working memory still working? Am Psychol 2001;56(11):851-64.
30. Coull JT, Frith CD, Frackowiak RS, Grasby PM. A fronto-parietal network for rapid visual information processing: a PET study of sustained attention and working memory. Neuropsychologia 1996;34(11):1085-95.
31. Sarter M, Givens B, Bruno JP. The cognitive neuroscience of sustained attention: where top-down meets bottom-up. Brain Res Brain Res Rev 2001;35(2):146-60.
32. Graw P, Krauchi K, Knoblauch V, Wirz-Justice A, Cajochen C. Circadian and wake-dependent modulation of fastest and slowest reaction times during the psychomotor vigilance task. Physiol Behav 2004;80(5):695-701.
33. Philip P, Taillard J, Sagaspe P, et al. Age, performance and sleep deprivation. J Sleep Res 2004;13:105-10.
34. Drummond SP, Bischoff-Grethe A, Dinges DF, Ayalon L, Mednick SC, Meloy MJ. The neural basis of the psychomotor vigilance task. Sleep 2005;28(9):1059-68.
35. Beaumont M, Batejat D, Pierard C, et al. Slow release caffeine and prolonged (64-h) continuous wakefulness: effects on vigilance and cognitive performance. J Sleep Res 2001;10(4):265-76.
36. Lieberman HR, Tharion WJ, Shukitt-Hale B, Speckman KL, Tulley R. Effects of caffeine, sleep loss, and stress on cognitive perfor-

mance and mood during U.S. Navy SEAL training. Sea-Air-Land. Psychopharmacology (Berl) 2002;164(3):250-61.
37. Magill RA, Waters WF, Bray GA, et al. Effects of tyrosine, phentermine, caffeine D-amphetamine, and placebo on cognitive and motor performance deficits during sleep deprivation. Nutr Neurosci 2003;6(4):237-46.
38. Caldwell JA, Caldwell JL, Darlington KK. Utility of dextroamphetamine for attenuating the impact of sleep deprivation in pilots. Aviat Space Environ Med 2003;74(11):1125-34.
39. Caldwell JA, Jr., Caldwell JL, Smythe NK, 3rd, Hall KK. A double-blind, placebo-controlled investigation of the efficacy of modafinil for sustaining the alertness and performance of aviators: a helicopter simulator study. Psychopharmacology (Berl) 2000;150(3):272-82.
40. Wesensten NJ, Belenky G, Thorne DR, Kautz MA, Balkin TJ. Modafinil vs. caffeine: effects on fatigue during sleep deprivation. Aviat Space Environ Med 2004;75(6):520-5.