How can noise alter neurophysiology in order to improve human behaviour? A combined transcranial random noise stimulation and electroencephalography study

James G. Sheffield[a*], Gal Raz[a], Francesco Sella[b], & Roi Cohen Kadosh[a*]

[a]University of Oxford, Oxford, UK .[b]University of Sheffield, Sheffield, UK.

* Corresponding author

Abstract

Transcranial random noise stimulation (tRNS) has been used to improve cognitive performance in patients and healthy individuals in different domains. It is therefore considered as a promising method for cognitive enhancement and rehabilitation. However, the mechanistic underpinnings of tRNS are poorly understood, mainly due to difficulties in separating neural signal from stimulation artefact. Here we suggest a procedure to successfully remove the tRNS artefact in both the time and frequency domains, leading to electroencephalography (EEG) data that are comparable between participants in active and sham (placebo) tRNS conditions. Such a procedure affords us a unique opportunity to examine the neurophysiological mechanisms that are altered by tRNS during complex mental arithmetic. We recorded EEG data in 69 participants who received arithmetic training concurrently with active or sham tRNS above the dlPFC. We successfully found that active tRNS improved arithmetic performance compared to sham tRNS. The underlying event-related potentials revealed that the tRNS effect was associated with increases in components that are associated with domain-general, rather than domain-specific, changes as indicated by alteration in amplitude of attention and preparatory markers. The results indicate that the enhancement effect of tRNS when applied above the dlPFC acts by effecting general attentional mechanisms during cognitive training, which explains the potential improvement seen over a large array of cognitive tasks. In addition, our procedure to remove the tRNS artefact effectively allows future investigation of the mechanisms underlying the effect of tRNS on human behaviour.

1 Introduction

The aim to improve human behaviour in health or disease is shared by scientists in different disciplines, including psychology, neuroscience, medicine, and engineering. One of the main techniques is cognitive training, which encompasses myriad paradigms, regimes, and techniques, and has been applied to improve performance in a range of cognitive domains both in normative populations (Brem et al., 2018; Javadi & Cheng, 2013; Tsai et al., 2018) and in clinical ones (Ahn et al., 2019; Auvichayapat et al., 2013; Sitzer, Twamley, & Jeste, 2006).

Recently, an accumulating number of studies have attempted to improve the benefits of cognitive training using transcranial random noise stimulation (tRNS), a form of transcranial electrical stimulation (tES; Simonsmeier, Grabner, Hein, Krenz, & Schneider, 2018). An increasing body of literature highlights the successful application of tRNS to modulate performance in a variety of training paradigms and cognitive tasks (for reviews see Santarnecchi et al., 2015; Simonsmeier et al., 2018). tRNS has been used to improve perception (Herpich et al., 2019; Pirulli, Fertonani, & Miniussi, 2013; van der Groen, Tang, Wenderoth, & Mattingley, 2018), numerosity discrimination (Cappelletti et al., 2013; Cappelletti, Pikkat, Upstill, Speekenbrink, & Walsh, 2015), executive control and multitasking (Frank, Harty, Kluge, & Cohen Kadosh, 2018)), motor skill learning (Prichard, Weiller, Fritsch, & Reis, 2014; Terney, Chaieb, Moliadze, Antal, & Paulus, 2008), fluid intelligence (Brem et al., 2018), and arithmetic learning (Snowball et al., 2013; Pasqualotto, 2016, Popescu et al., 2016). Given its higher perceptual threshold compared to other forms of tES with higher amplitudes required to detect the application of stimulation, it is easier to create viable sham conditions (Ambrus, Paulus, & Antal, 2010). Furthermore, tRNS has been reported in several studies to yield a stronger effect than anodal transcranial direct current stimulation (Fertonani, Pirulli, & Miniussi, 2011; Inukai et al., 2016; Simonsmeier et al., 2018). Combined, these studies suggest that tRNS is a promising form of neuromodulation for improving complex cognition.

Despite the consistent application of tRNS in cognitive training, we have little understanding about the underlying neurophysiological impact of this stimulation. Magnetic resonance-based methods are less optimal in the context of cognitive training due to the induced noise, limited ecological validity, time-constraints, and cost. Largely, studies that use electroencephalography (EEG) to investigate the changes evoked by tRNS have done so by looking at pre/post changes in brain activity (e.g., Harty & Cohen Kadosh, 2019). To a degree, this reflects difficulties in simultaneously recording the electrical activity from the brain whilst delivering electrical stimulation, which is often done at amplitudes an order of magnitude larger than the physiological signal measured by EEG equipment. However, one cannot be sure that the changes that follow from the end of the stimulation (termed “offline effect”) are reflective of the altered neural response that might occur during stimulation itself (termed “online effect”). An effective procedure to overcome these difficulties could help elucidate possible mechanisms responsible for tRNS-evoked changes.

Recently, Rufener, Ruhnau, Heinze, & Zaehle (2017) used frequency band-pass filters to remove tRNS-induced artefacts from EEG data analysed in the time domain,

revealing reduced N1 latency in the tRNS group compared to the sham during a pitch discrimination task. Whilst this method allows demonstration of changes in the time domain, EEG data are multidimensional; we can characterise it not just in the time domain, but also in various facets of the frequency domain, including power, phase, and frequency (Cohen, 2014). Additionally, as the process of creating an event-related potential (ERP) acts as a low-pass filter, it mitigates tRNS noise in higher frequency bands, and though tRNS is usually applied at higher frequency ranges (e.g., 100–640Hz), this process does not remove low-frequency artefacts.

If stimulation noise can be removed from ongoing EEG activity, such progress will allow us to investigate neurophysiological changes that are associated with the effect of tRNS on behaviour, including the causal mechanisms in cognitive training. A prevalent hypothesis is that the effect of tRNS on behaviour is due to stochastic resonance (Fertonani et al., 2011; van der Groen & Wenderoth, 2016a), a phenomenon in which introducing an appropriate level of noise can enhance the output of subthreshold signals. Some studies lend support to the stochastic resonance explanation by showing that the effect of tRNS is found in participants with poorer baseline ability (Evans, Banissy, & Charlton, 2018; Frank, Harty, Kluge, & Cohen Kadosh, 2018; Harty & Cohen Kadosh, 2019) or in when participants are presented with more difficult tasks (Popescu, Krause, Terhune, Twose, Page, Humphreys, Cohen Kadosh, et al., 2016; van der Groen & Wenderoth, 2016a). However, the support for the stochastic resonance explanation of tRNS is limited to the behavioural level, and, at least in some cases, might arise from a ceiling effect due to high performance in easy tasks or high-performing participants.

Here, we present the results of a short mathematical training study in which we recorded EEG during active or sham tRNS while participants solved complex multiplication problems. The motivation to use mathematical training was threefold: 1) math is a strong predictor of academic performance in childhood and adolescence (Parsons & Bynner, 2006), individual outcome, and job performance (Butterworth, Varma, & Laurillard, 2011); 2) most of the existing tRNS studies have focused on improving mathematical performance, while the mechanisms for such enhancement have remained unclear (Simonsmeier et al., 2018); 3) math represents a more realistic mode of complex cognition due to its requirement on cognitive processes such as executive functions (Cragg & Gilmore, 2014), emotional regulation (Dowker, Sarkar, & Looi, 2016), spatial abilities (Sella, Sader, Lolliot, & Cohen Kadosh, 2016), and algorithmic and memory processes (Rickard, 1997). We tested whether we could successfully remove the artefact induced by tRNS from the ongoing EEG data in both the time and frequency domains. We then aimed to use the tRNS removed data to investigate the neurophysiological mechanisms that are altered by tRNS enhancement and their effect on behaviour.

2 Methods

2.1 Participants

Eighty-one participants (44 females, mean age = 23.7 years, SD = 5.03) were asked to complete a two-day arithmetic learning paradigm in a study investigating the effects

of tRNS stimulation on learning. Participants were excluded if they had a previous history of neurological disorders, a family history of seizure disorders, metal implanted in their head, had previously undergone a neurosurgical procedure, or had received any other form of brain stimulation in the past week. Participants were also requested to refrain from alcohol (for the past 24 hours) or caffeine (in the hour leading up to the study), and to have gotten at least 6 hours of sleep the night before their testing sessions. Four participants were removed during the collection process: 1 due to software failure, and 3 due to not returning for a second day of testing. During analysis, 7 participants were removed because they did not provide enough clean EEG data for analysis, and 1 participant had a corrupted EEG file. Final comparisons were carried out on 69 participants.

2.2 Arithmetic learning task

In each trial, participants were presented with a fixation cross (1000ms) before seeing a multiplication problem, which was left up until the participants responded. Participants responded via a microphone in front of them to measure reaction times (RTs), and then via a keyboard to evaluate correctness. If participants realised their answer was wrong following their verbal response but before their typed response, they were asked to type in their verbal response anyway. Following their typed response, they received feedback for 2000ms (**Figure 1**).

Each multiplication was a two-digit number multiplied by a one-digit number resulting in a two-digit number (see Table S1). Neither of the operands included a 0 or a 5, nor were there operands with identical digits. Participants completed 18 blocks, and each block was comprised of the same 8 multiplication problems presented in random order. After each block participants reported how much they used retrieval vs. calculation in that block. At the end of the first day, participants were asked to report on whether they received active or sham stimulation. Though not analysed here, as it is beyond the scope of the present study, participants returned for a second day of testing.

Analysis was carried out on the accuracy and median reaction time for correct responses, and their report of the chosen strategy (retrieval vs. calculation).

2.3 tRNS

Participants were assigned to a stimulation condition using a variance minimisation algorithm based on their baseline arithmetic abilities (Sella, Raz, & Cohen Kadosh, *submitted*), which aims to allow better matching between groups in comparison to random allocation (Treasure & MacRae, 1998). Half the participants (N=39) received active high-frequency tRNS (100–500Hz) on the first day of the experiment (**Figure 2**). Stimulation was generated using a StarStim R32 (Neuroelectrics, Barcelona, Spain) and delivered via Ag/AgCl electrodes with 3.14cm^2 surface area and carried to the scalp by conductive gel. tRNS was delivered over F3 and F4 according to the 10–20 system, with an amplitude of 1mA peak-to-peak for a duration of 20 minutes, beginning with the start of the arithmetic learning task. The tRNS had a ramp period at the beginning and end of the stimulation of 30 seconds. The remaining participants (n=37) received sham stimulation. Participants in the sham condition received 30 seconds of stimulation

at the beginning of the task: 15 seconds of ramp up and 15 seconds of ramp down. To ensure experimenter blinding, the stimulation groups were given pseudonyms in the stimulation software, and the allocation algorithm provided only these pseudonyms, and not the real stimulation conditions.

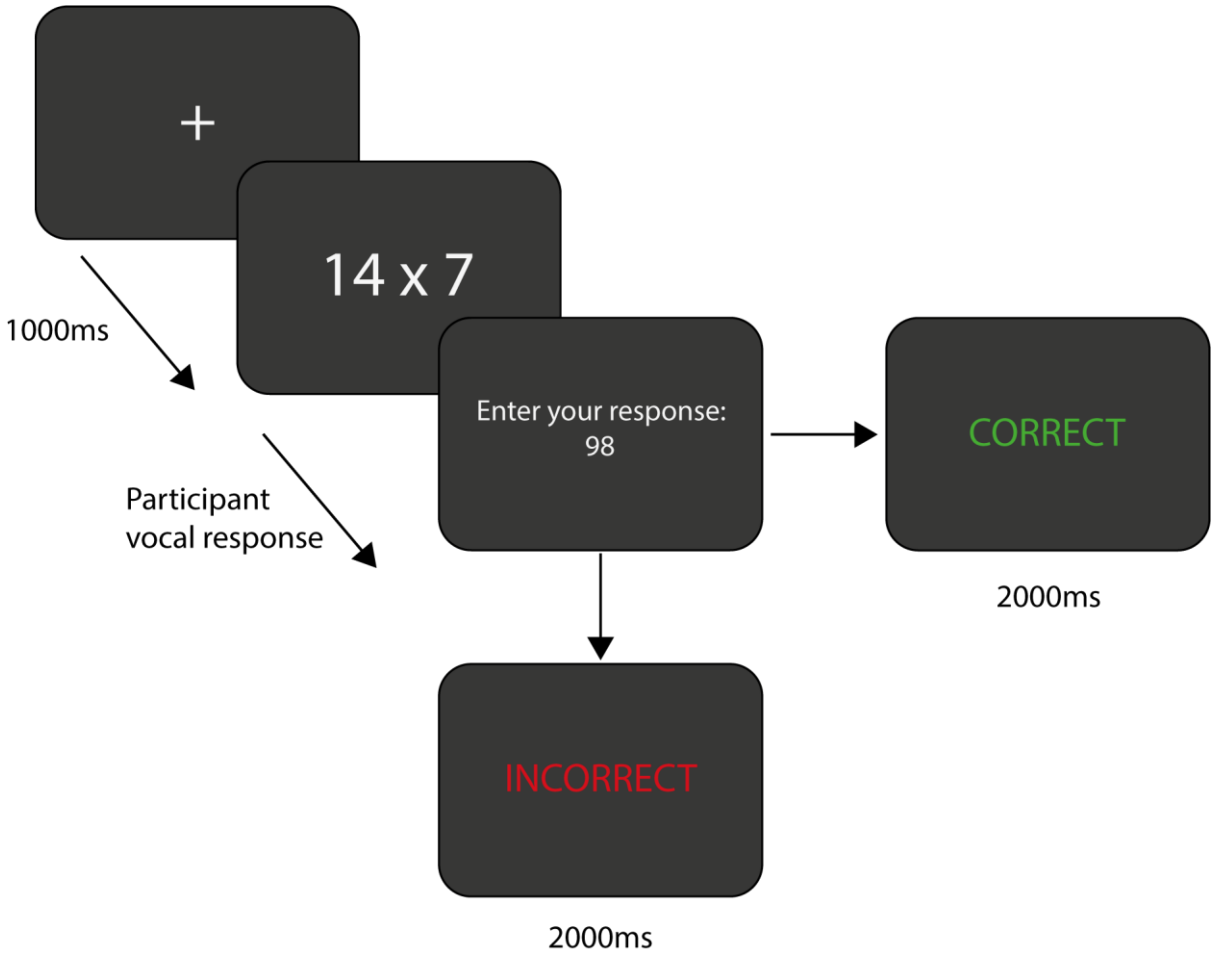


Figure 1: An illustration of a trial from the arithmetic training task.

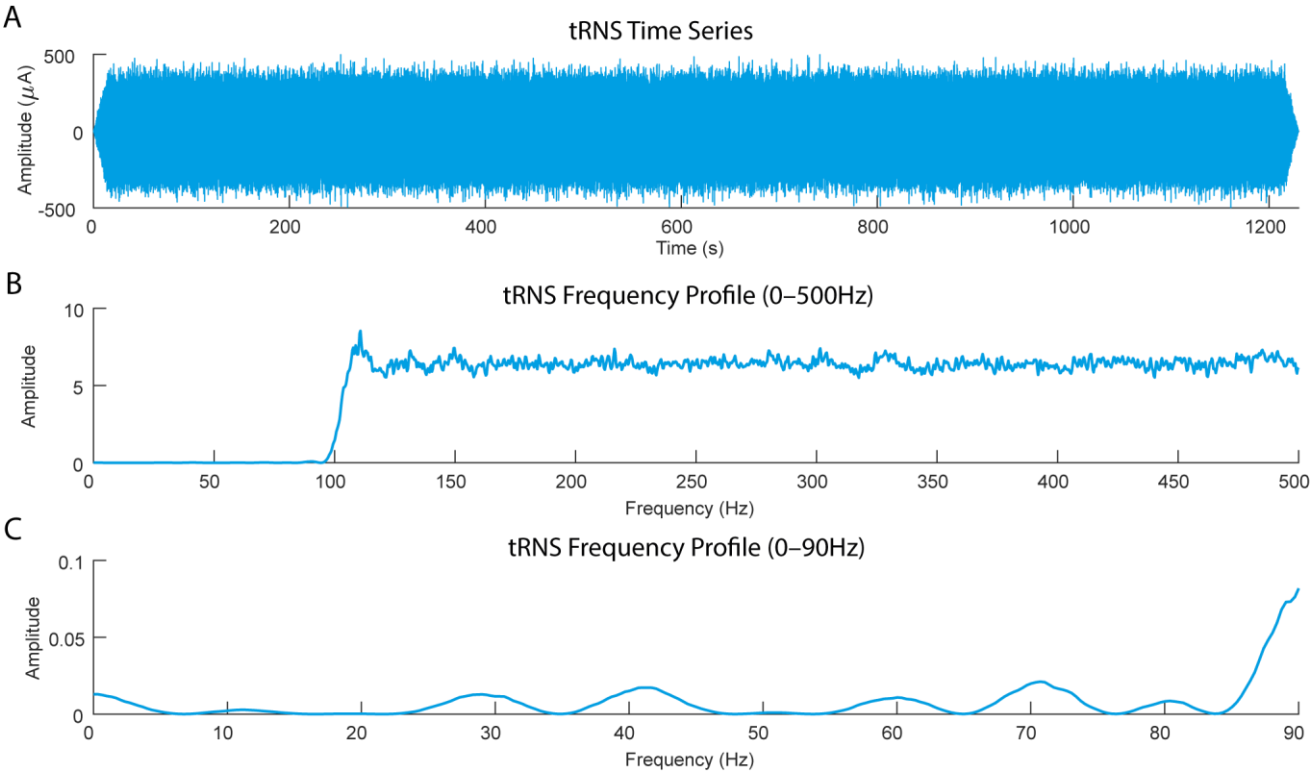


Figure 2: The profile of the tRNS stimulation. a) The time series of the stimulation over the 20-minute period, b) The full frequency profile (0–500Hz) of the tRNS stimulation, and c) the frequency profile in the 0–90Hz range. Despite the tRNS being filtered for higher frequencies (100–500Hz), peaks in the activity are visible in the lower frequency range (0–90Hz)

2.4 EEG collection

EEG data were collected via the stimulation unit from 28 sites whilst stimulation was ongoing and 30 sites once stimulation had ended (electrodes F3 and F4 were added) at a sampling rate of 500Hz. The data were recorded with Ag/AgCl electrodes. Similar to the stimulation preparation, EEG scalp sites were prepared by first abrading the scalp with abrasive gel, and then cleaned using surgical alcohol. The space between the scalp and electrodes was bridged using conductive gel.

2.5 EEG preprocessing

EEG data were imported into EEGLAB (Delorme & Makeig, 2004) for preprocessing and were then filtered using a 0.5–80Hz band-pass filter. The mean was removed from each channel, and then periods of non-stereotyped artefactual data were removed after identification via visual inspection. Infomax ICA was run, and any periods of non-independence between components were identified by eye. Following removal of these periods of non-independence, we ran a second infomax ICA. For analysis, two versions of the EEG data were created. In the first version, we removed all components associated with stereotyped noise (e.g., blinks, lateral eye movements, heart beat

artifacts), as well as a component with a bilateral activation pattern and spectral profile matching the tRNS spectral profile. In the second version, only components associated with typical stereotyped EEG artefacts were removed whilst the tRNS artefact components remained, allowing us to investigate the impact of tRNS-induced artefacts alone. For the ERP analysis, EEG data were grand averaged over trials. For the frequency analysis, the data were epoched into two-second windows with 50% overlap and a Hanning window was applied before being put through a FFT. Finally, ITPC values were computed following methods outlined in Cohen (2014).

2.6 Model selection

For RT, accuracy, and calculation vs. retrieval behavioural data, multiple Bayesian multilevel models were run and primarily evaluated using two methods. Firstly, model fit was tested by checking graphical posterior predictive checks and Markov chains traces to determine whether the predicted values were unusual for the data set and whether the Markov chains were unable to step incorrectly. Secondly, we tested the out-of-sample cross-validation (Lambert, 2018), selecting the model with the highest out-of-sample validation score, as indicated by the expected log pointwise density (Gelman, Hwang, & Vehtari, 2014; Vehtari, Gelman, & Gabry, 2017). Here, we calculated the elpd using stratified k-fold (k = 10) clustering. In the k-fold clustering, the data set was divided into 10 subsets. The model was then run on 9 of the subsets, and then the results of the 10th were predicted. This process was repeated 10 times, each time leaving out one of the data sets as a cross-validation set. The elpd value could then be calculated as the average log-likelihood of each of the cross-validation sets (specifically, the log of the average exponential log-likelihoods; Gelman et al., 2014). We refer to this as the $elpd_{10\text{-}fold}$. Higher (closer to zero) elpd values indicates that the estimated distribution is close to the true one.

To compare how well the models described the current data set, we also calculated the Bayes factor (BF) for the two models with the highest elpd values. The BF is the likelihood of the data under one model compared to another, giving us evidence for one model over another considering the current dataset.

As we were primarily interested in how the data might describe future data as opposed to the current data set, the $elpd_{10\text{-}fold}$ results are reported in the main text whilst the BF results are reported in the supplementary material. It should be noted that cross-validation approaches can be less sensitive to smaller effect sizes, when comparing models with similar terms (Piironen & Vehtari, 2017). Moreover, as BF values are sensitive to prior selection (O'Hara & Sillanpää, 2009) and we selected largely non-informative priors for this analysis, we chose to favour the elpd values here. That said, unlike more traditional analysis, both cross-validation (elpd) and BF results should be taken as part of a broader picture of the data

2.7 Analysis

Behavioural analyses were conducted using Bayesian multilevel models run via the *brms* package (Bürkner, 2017, 2018) in RStudio (Rstudio Team, 2016). Models were run (see Table 1 for the list of models) and compared using stratified 10-fold cross-validation. The best model is the one with the highest (closest to zero) expected

log pointwise predictive density (elpd). For the RT, accuracy, and calculation/retrieval analysis, weakly informative priors were selected for both the intercept and the b coefficients (normal (0,1)), and each model was run with 4 chains of 12,000 iterations (6,000 iterations being warmup).

Analysis comparing the ERPs and spectral profiles between the sham, tRNS removed, and non-removed data were compared using a Bayesian t-test as per the method outlined in Rouder et al. (2009). Using Bayesian, rather than frequentist statistics allowed us to avoid the multiple comparison problem, as this problem is irrelevant in the case of Bayesian statistics. As suggested by Jeffreys (1998), BFs above 3 were considered as moderate evidence (in favour of a model), values above 10 as strong evidence, and values above 100 as extremely strong evidence. Data were compared during both the online and offline periods. For both stimulation groups we defined the online period as the time in which the active tRNS group received stimulation, and the offline period as the time following tRNS until the end of the task.

3 Results

3.1 Learning slopes

To confirm that learning had taken place, we calculated the slope coefficient for block number on reaction time to identify participants with a non-negative slope (indicating no decrease in reaction time over the task period). In total, 5 participants demonstrated no decrease in reaction time, all from the sham group. A chi-squared test checking whether the two groups differed indicated that this was statistically significant ($X^2(1)=5.35$, $p=0.02$). However, Bayesian contingency tables run in JASP (JASP Team, 2019) suggested that the data were only 3.6 times more likely under the alternate hypothesis (there were more non-negative learning slopes amongst the sham participants than the active tRNS participants) compared to the null hypothesis (no difference between the stimulation groups).

3.2 Stimulation blinding

Analysis of the stimulation blinding indicated that the null model (participants in the active and sham tRNS groups were as likely to report that they were stimulated) was slightly favourable compared to the alternate hypothesis ($elpd_{10\text{-fold}}$ -39.991 vs. -40.65), and the posterior distribution of the group slope values were centred on 0 (95% highest density interval (HDI) = [-0.85, 1.01]). This suggests that the two groups were equally likely to indicate that they were in the active condition, showing that blinding was successful.

3.3 RT

As there were no time constraints on participant response, participants were able to take as long as needed to ensure an accurate answer. In such cases, we would expect to observe the effect in the reaction time data rather than the accuracy data (Pachella, 1974; Snowball et al., 2013). The results of the model comparisons revealed that the three-way interaction model (median reaction time ~ stimulation group * block

number * baseline reaction time) was the most favourable according to the elpd value (Table 1). The results of the Hamiltonian Monte Carlo (HMC) chain traces and the graphical posterior predictive checks (PPCs) indicated that the models converged successfully and that posterior sampling produced data that were a reasonable approximation of the actual data (see supplementary **Figure S1**). Visualisation of the marginal effects of the interaction terms (95% HDI values [0.001, 0.005]; see **Figure 3**) suggested that tRNS induced differences in reaction times during the training, with a greater increase in reaction time as the block number increases in participants who exhibited faster reaction times at baseline. There also appeared to be a similar, but reduced, effect for participants with an average baseline reaction time. This suggests that tRNS was more beneficial as participants' baseline reaction time was faster.

Response times

Model	Predictors	elpd	se	Δelpd	Δse
0	Intercept only	-2141.70	42.26	-	-
1 vs. 0	Block	-1575.60	43.01	566.00	26.20
2 vs. 0	Stimulation	-2149.27	43.33	-7.60	5.20
3 vs. 0	Baseline RT	-2144.95	42.21	-3.20	4.70
4 vs. 1	Block + Stimulation	-1581.39	43.49	-5.70	3.90
5 vs. 1	Block + Baseline RT	-1593.90	45.00	-18.20	5.70
6 vs. 1	Stimulation + Baseline RT	-2139.91	41.31	-564.20	26.20
7 vs. 1	Block * Stimulation	-1587.85	45.08	-12.20	6.20
8 vs. 1	Block * Baseline RT	-1573.19	44.08	2.50	10.00
9 vs. 1	Stimulation * Baseline RT	-2136.30	41.42	-560.60	25.90
10 vs. 1	Block + Stimulation + Baseline RT	-1584.80	43.54	-9.10	5.00
11 vs. 1	**Block * Stimulation * Baseline RT**	**-1567.26**	**44.57**	**8.40**	**10.40**

Table 1: Comparison of Bayesian multilevel models regressing reaction time back on different predictors. The best model (Model 11), which includes the 3-way interaction between stimulation group, block number, and baseline reaction time, has been bolded. The Δelpd and Δse columns represent the difference in the elpd (cross-validation accuracy) and se (cross-validation standard error) values between the two models noted at the start of the row.

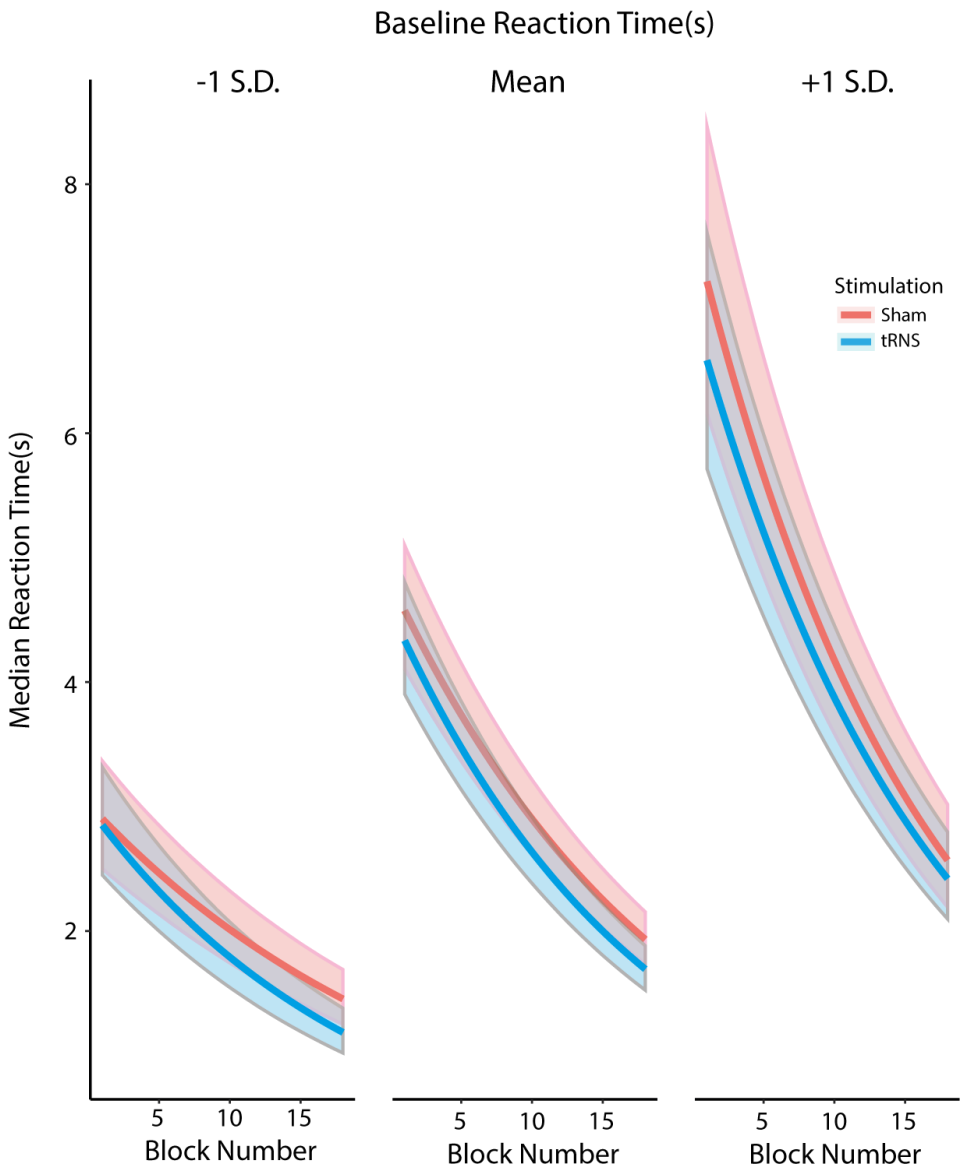


Figure 3: Plotting the 3-way interaction between stimulation group, block number, and baseline reaction time interaction effect of RT as a function of one standard deviation (SD) above and below the mean. The bars represent the 95% credible interval.

3.4 Accuracy

The model with the greatest estimated out-of-sample predictive accuracy was the model including the main effects of both block number and stimulation group (see Table 2). However, the model did not suggest any significant impact of stimulation group on accuracy, though accuracy was slightly higher in the stimulation group compared to the sham group). The similar $elpd_{10\text{-fold}}$ values of the models containing block number suggested that the likelihood of each model was largely determined by the inclusion of block number, with other variables having negligible effects

Accuracy

Model	Predictors	elpd	se	Δelpd	Δse
0	Intercept only	-1123.50	30.55	-	-
1 vs. 0	Block	-1097.63	30.56	25.90	8.80
2 vs. 0	Stimulation	-1132.47	31.03	-9.00	3.10
3 vs. 0	Baseline Accuracy	-1118.67	30.05	4.80	2.90
4 vs. 1	**Block + Stimulation**	**-1094.18**	**30.52**	**3.50**	**3.30**
5 vs. 1	Block + Baseline Accuracy	-1100.03	30.39	-2.50	3.70
6 vs. 1	Stimulation + Baseline Accuracy	-1123.20	30.51	-25.60	9.30
7 vs. 1	Block * Stimulation	-1099.56	30.86	-1.90	3.80
8 vs. 1	Block * Baseline Accuracy	-1094.53	30.18	3.10	5.00
9 vs. 1	Stimulation * Baseline Accuracy	-1117.40	30.24	-19.80	9.00
10 vs. 1	Block + Stimulation + Baseline Accuracy	-1102.04	30.44	-4.40	3.70
11 vs. 1	Block * Stimulation * Baseline Accuracy	-1099.26	30.61	-1.60	5.10

*Table **2**: Comparison of Bayesian multilevel models regressing accuracy back on different predictors. The best model (Model 4), which contains main effect terms of block number and stimulation group, has been bolded. The Δelpd and Δse columns represent the difference in the elpd and se values between the two models noted at the start of the row.*

3.5 Calculation/retrieval results

Similar to the RT results, the most favourable model was the model with the three-way interaction (see Table 3). However, similar to the results for accuracy, the most favourable models all included a block number, indicating that the likelihood of each model is largely determined by the presence of the block number. This was supported by the largely limited impact of stimulation and baseline. As stimulation appeared to have little effect on calculation/retrieval, the full model results are reported in the supplementary materials (see **Figure S2** in the supplementary material).

Retrieval score

Model	Predictors	elpd	se	Δelpd	Δse
0	Intercept only	-2325.06	25.00	-	-
1 vs. 0	Block	-2038.67	27.47	286.40	15.80
2 vs. 0	Stimulation	-2326.81	25.09	-1.70	2.90
3 vs. 0	Baseline RT	-2324.73	24.99	0.30	2.80
4 vs. 1	Block + Stimulation	-2043.65	27.65	-5.00	2.70
5 vs. 1	Block + Baseline RT	-2035.32	27.53	3.30	2.90
6 vs. 1	Stimulation + Baseline RT	-2327.12	25.06	-288.40	15.90
7 vs. 1	Block * Stimulation	-2040.64	27.65	-2.00	2.90
8 vs. 1	Block * Baseline RT	-2038.90	27.54	-0.20	3.00

9 vs. 1	Stimulation * Baseline RT	-2328.76	25.25	-290.10	16.00
10 vs. 1	Block + Stimulation + Baseline RT	-2039.61	27.68	-0.90	2.90
11 vs. 1	**Block * Stimulation * Baseline RT**	**-2033.51**	**27.47**	**5.20**	**3.30**

*Table **3**: Comparison of Bayesian multilevel models regressing calculation/retrieval scores back on different predictors. The best model (Model 11), which includes the 3-way interaction between stimulation group, block number, and baseline reaction time, has been bolded. The Δelpd and Δse columns represent the difference in the elpd and se values between the two models noted at the start of the row.*

3.6 tRNS removal results

The results of the Bayesian t-tests indicated that whilst the grand average ERPs did not differ between the active and sham stimulation groups, regardless of whether we removed the tRNS artefact from the data or not, the standard deviations did. This problem was solved when we removed the tRNS artefact (BF_{10}'s < 3; **Figure 5**). Similarly, the frequency domain analysis of the data revealed substantial differences between the active and sham stimulation groups before tRNS scrubbing was applied. Specifically, tRNS was associated with several broad peaks (larger than the underlying power evoked in neural circuits) at approximately every 10Hz from 30–80Hz (aside from 50Hz). These peaks were not present in the sham data. Following ICA removal of the tRNS signal, the active tRNS data was largely comparable to the sham data but notably different from the non-removed data in both frontal and occipital sites (**Figure 6**).

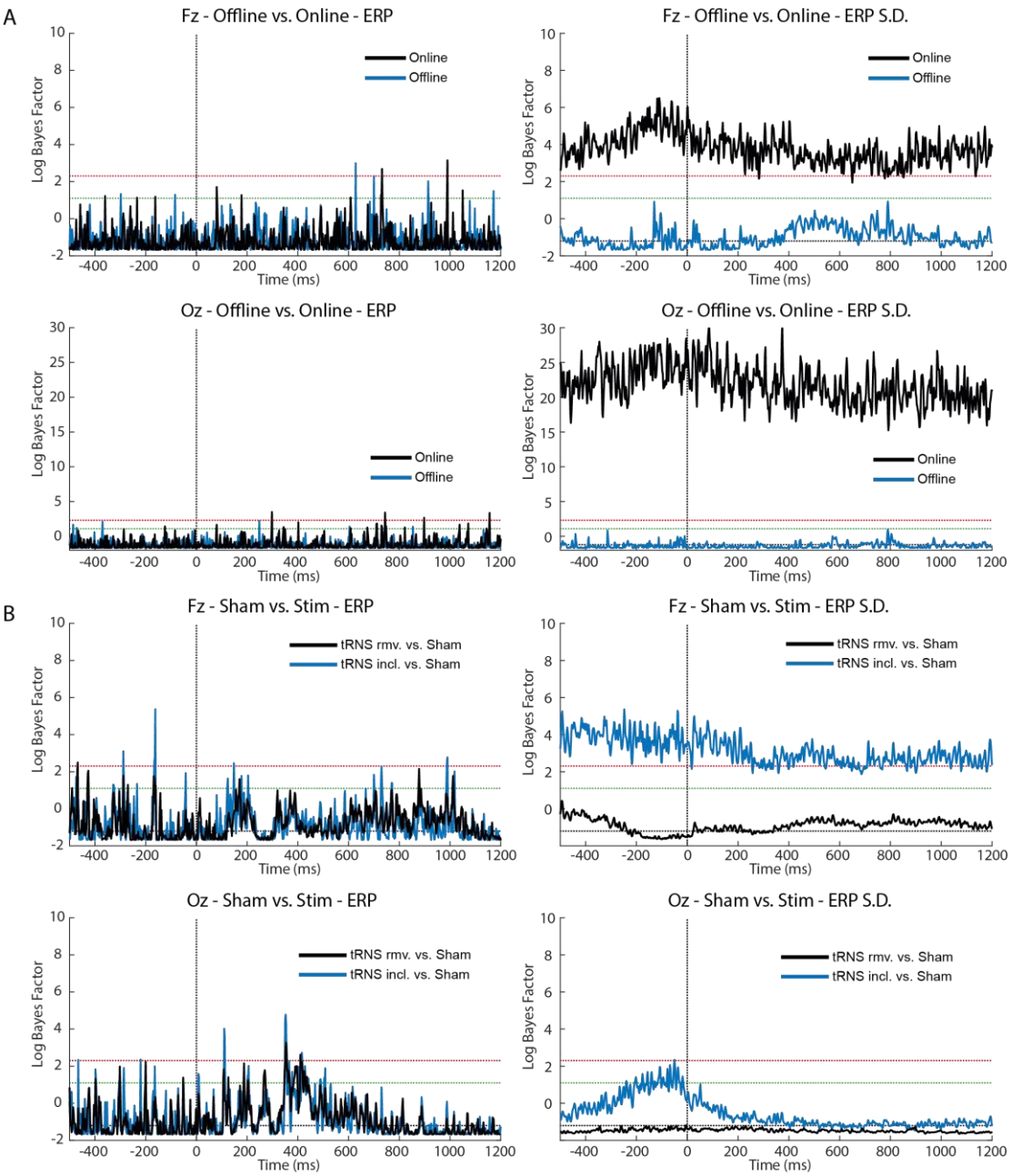


Figure 5: Time-domain results of Bayesian t-tests after tRNS artefact removal. a) Results of Bayesian t-tests (log BF_{10}) comparing ERPs and ERP standard deviations for the data with the tRNS removed against the data with the tRNS artefact included. Results are for both electrodes Fz (closest to the tRNS electrodes over F3–F4) and Oz (furthest away from the tRNS electrodes over F3–F4). Red and green dotted lines represent a BF of log(3) and log(10), respectively, whereas the black dotted line represents a value of log(0.3). Note that the scale at Oz is larger than that of Fz. These results suggest that there were no differences in the ERP between online and offline periods regardless of whether the stimulation artefact was included or not, but that there was substantially more variance (standard deviations (S.D.)) in the data when stimulation was ongoing. b) ERPs and ERP standard deviations for the data sets with the tRNS artefact included (tRNS incl..) and tRNS artefact removed (tRNS rmv.) compared against the sham participants for both Fz and Oz electrodes. As above, red, green, and black dotted lines represent a BF of log(3), log(10), and log(0.3), respectively. Similar to the online vs. offline comparison, there was little difference in the ERP between the sham and active stimulation conditions regardless of whether stimulation noise was removed or not. However, when the tRNS artefact was included, there was a noticeable increase in variance compared to the sham condition, but no increase in

variance was noticeable when the artefact was removed.

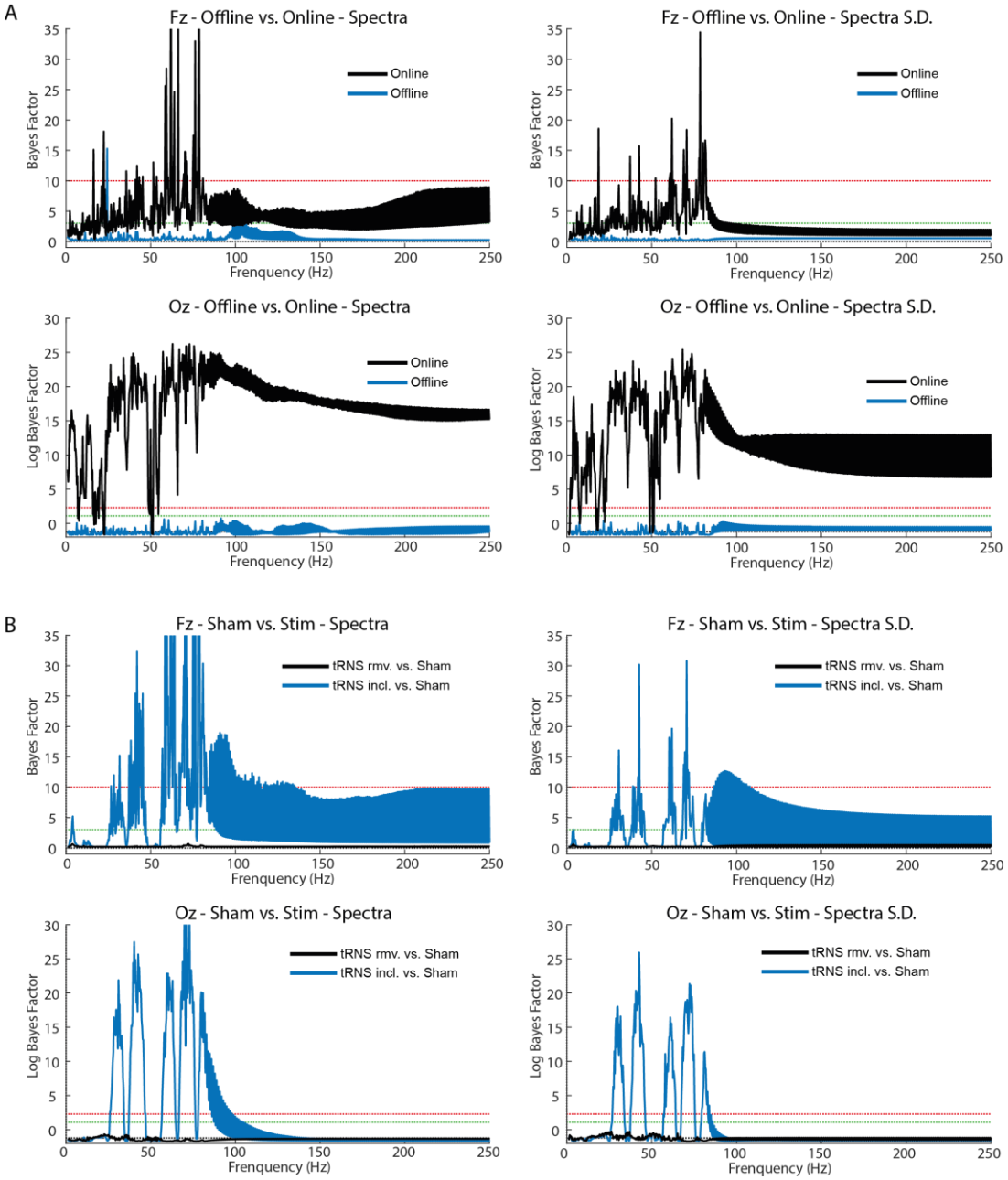


Figure 6: Frequency-domain results of Bayesian t-tests after tRNS artefact removal. a) Results of Bayesian t-tests (log BF_{10}) comparing the spectral profile and their standard deviations (S.D.) for the data with the tRNS removed against the data with the tRNS artefact included. Results are for electrodes Fz (closest to the tRNS electrodes over F3–F4) and Oz (furthest midline electrode from the tRNS electrodes over F3–F4). Red and green dotted lines represent a BF of log(3) and log(10), respectively, whereas the black dotted line represents a value of log(0.3). Note that the scale at Fz is different from that of Oz, with the latter being a log scale. Unlike the ERP results, both the periodograms and their standard deviations were notably different between the data with the tRNS artefact included and the tRNS artefact removed during the online period but not during the offline period. This is despite tRNS activity only being delivered in the 100–500Hz range. b) The spectra and their standard deviations for the data sets with the tRNS artefact included

(tRNS incl.) and tRNS artefact removed (tRNS rmv.) against the sham participants for both Fz and Oz electrodes. As above, red, green, and black dotted lines represent a BF of log(3), log(10), and log(0.3), respectively. Note that the Oz results are reported on a log scale, whilst the Fz results are not. These results indicate that the presence of 100–500Hz tRNS noise led to substantial noise in the 0–100Hz range, typically beginning at approximately 25Hz. However, this noise could be removed, leading to a spectral profile largely comparable to sham.

3.7 ERP results

Following the successful removal of the tRNS artefact, ERP plots of the tRNS-removed data from two midline sites revealed two main differences between the active and sham tRNS groups. First, over frontal sites, an early negative potential, peaking 100ms following stimulus presentation, was greater amongst the active tRNS group compared to the sham group. Notably, this difference was not seen following the presentation of a fixation cross (**Figure 7**). This indicates that the effect is specific to stimulus presentation, rather than a general tRNS-induced artefact, and highlights our success in removing the tRNS-induced artefact.

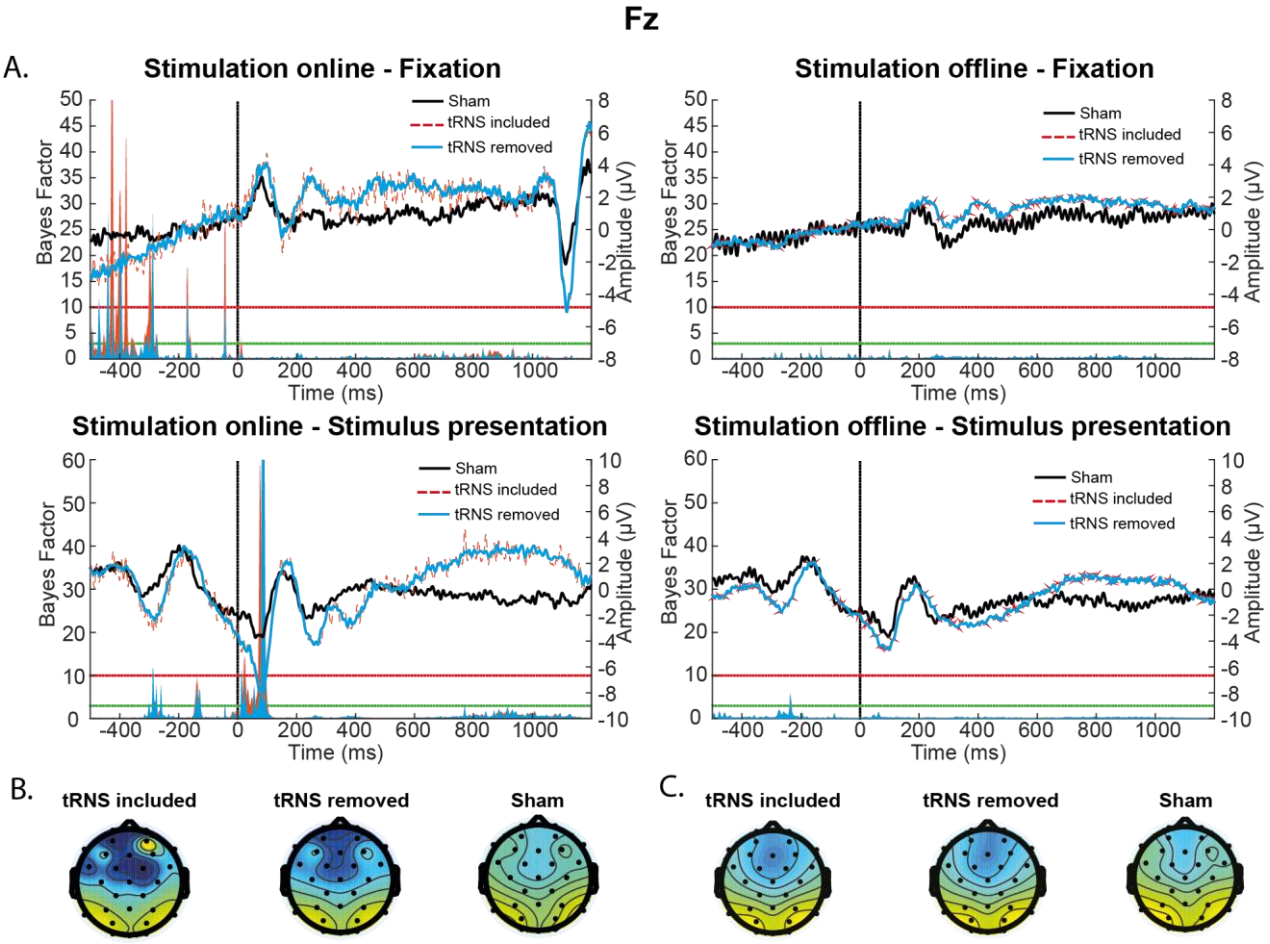


Figure 7: Task-related ERP effects at Fz following stimulus presentation. a) ERPs following stimulus and fixation cross

presentations showing a stimulation group effect during the online, but not the offline, period following stimulus presentation. Waveforms are plotted for both the tRNS removed data and the data with the tRNS artefact included. Underneath the ERP trace plots, the Bayes factor comparisons of the tRNS removed and the tRNS artefact included groups against the sham group. These values are plotted in the same colours as the ERPs. They highlight differences in the ERP waveform between the tRNS and sham groups following stimulus presentation, primarily in early ERPs (~100ms following stimulus presentation). b) Topographic distribution of activity (in µV) for the sham, removed tRNS, and non-removed tRNS data at the 100ms time point during the online period. c) Topographic distribution of activity (in µV) for the sham, removed tRNS, and non-removed tRNS data at the 100ms time point during the offline period. The colour bar range for all topographies is the same.

Second, at Oz, a greater positive inflection in the waveform 200ms after stimulus presentation was seen in the active tRNS group compared to the sham group (**Figure 8**). In contrast to the N1 effect, the group difference over occipital sites 200ms after stimulus presentation shows a lasting effect, as reflected in a similar effect during the offline period. During the offline period, the active tRNS group demonstrated greater positivity over occipital sites beginning approximately 200ms after stimulus presentation and continuing for another 300ms (**Figure 9**), likely reflecting the late positive potential (LPP).

At both sites, ERP differences existed between the active and sham tRNS groups regardless of whether tRNS scrubbing had taken place, though this was less clear in the topographical maps due to stimulation noise.

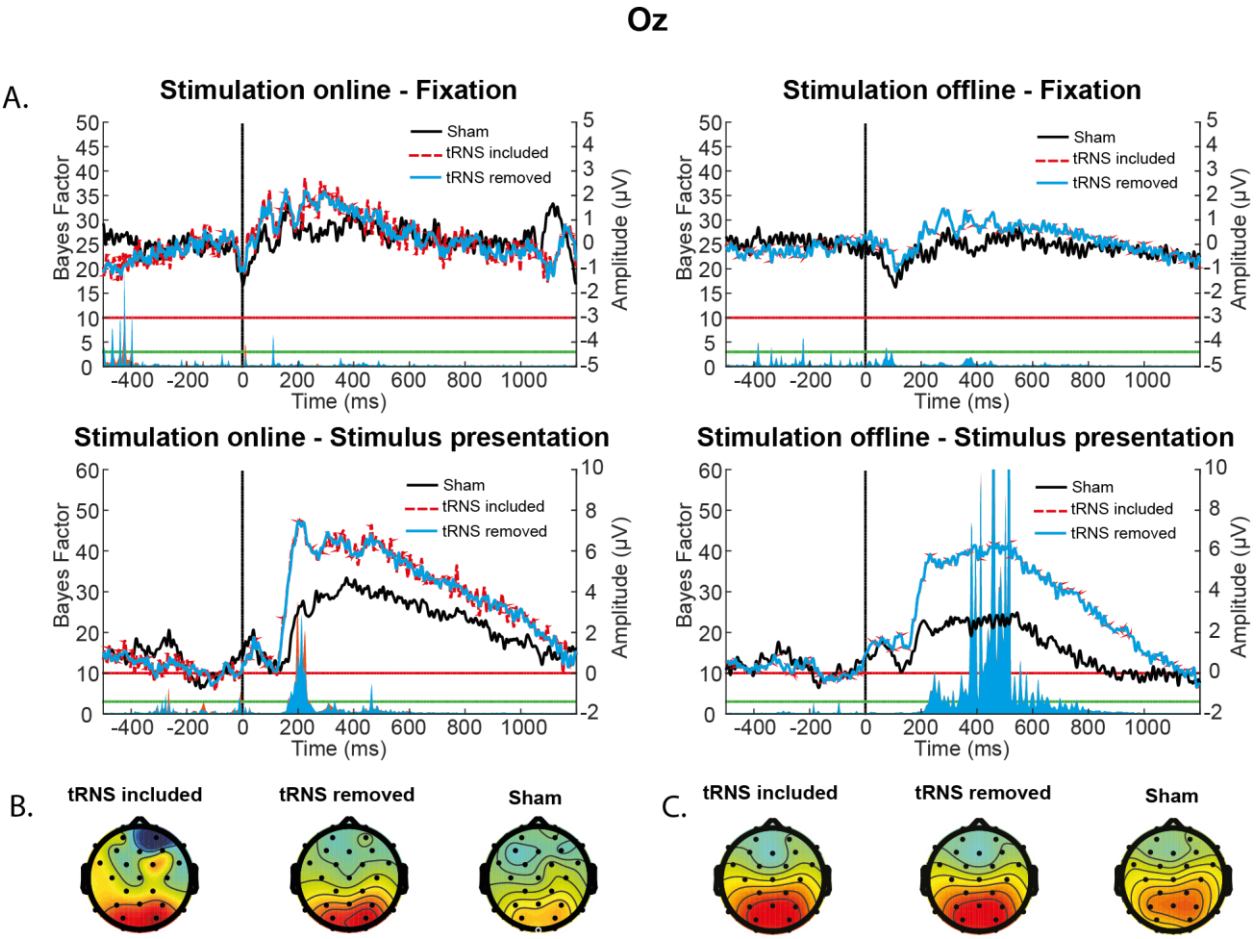


Figure 8: Task-related ERP effects at Oz following stimulus presentation. a) ERPs following stimulus and fixation cross presentations during both the online and offline periods. Waveforms are plotted for both the removed tRNS data and the non-removed tNRS data (with the tRNS artefact remaining). Underneath the ERP trace plots, the Bayes factor comparisons of with removed artefact and tRNS data included against the sham group. These values are plotted in the same colours as the ERPs. They highlight differences in the ERP waveform between the tRNS and sham groups following stimulus presentation in both early ERPs (~200ms following stimulus presentation) and late ERPs (~400ms following stimulus presentation). b) Topographic distribution of activity (in μV) for the sham, removed tRNS, and non-removed tRNS non-cleaned data at the 200ms time point during the online period. c) Topographic distribution of activity (in μV) for the tRNS removed data, tRNS included data, and the sham data at the 400ms time point during the offline period. The colour bar range for all topographies is the same.

3.8 ITPC results

Given previous research relating P2 amplitude to increased theta-band intertrial phase coherence (ITPC), we calculated the ITPC values for the theta band over Oz. These were then tested using Bayesian t-tests. The results indicated that whilst there were differences between active and sham tRNS groups in the P2 amplitude, there were no differences between active and sham tRNS groups in the theta-band ITPC over the same time period. There was, however, a difference between the tRNS artefact included conditions and both the sham group and tRNS removed condition during the online period. The tRNS artefact included data demonstrated lower ITPC values than both the sham and tRNS artefact removed data ($BF_{10} > 10$; **Figure 9**).

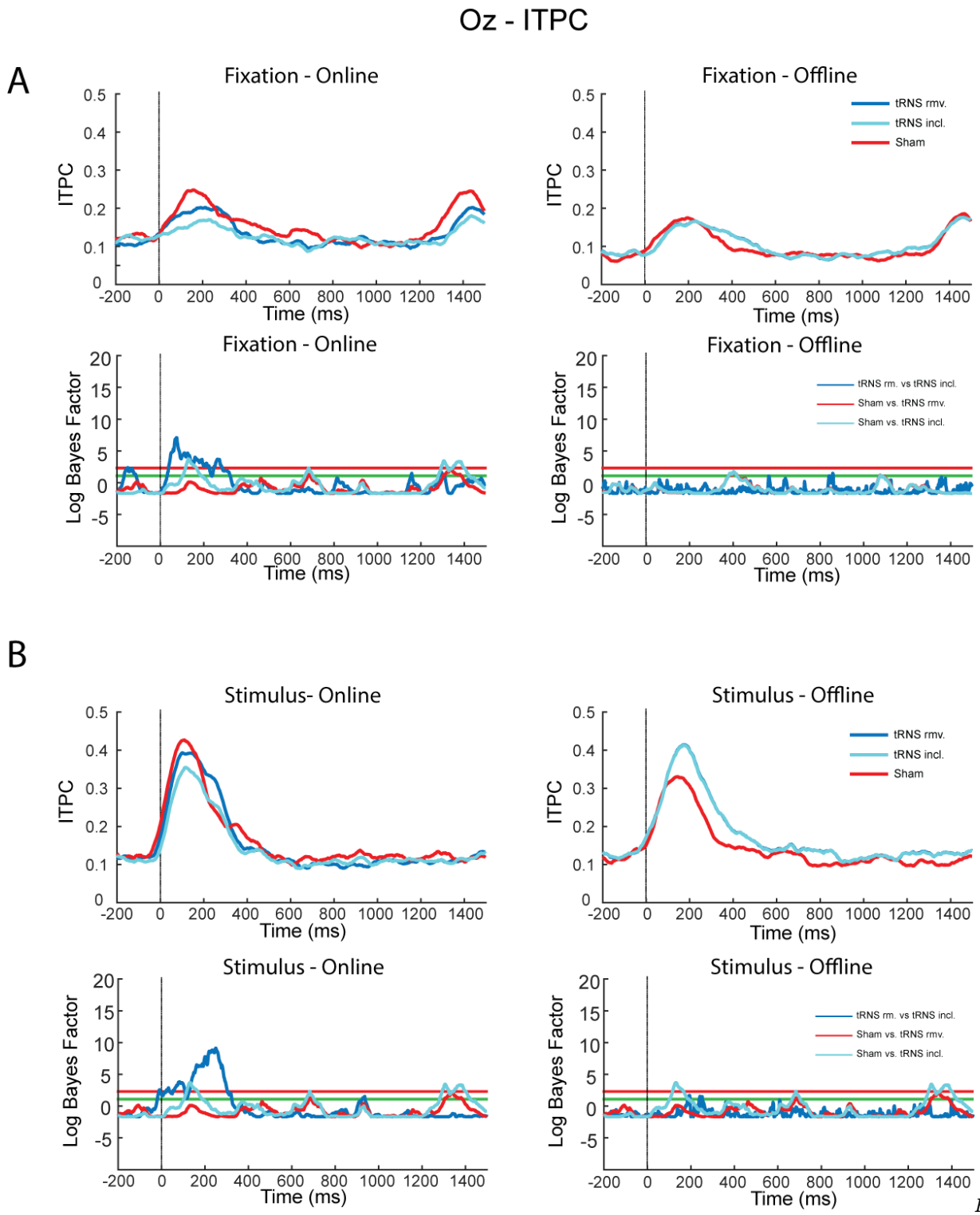


Figure 9: Changes in 6Hz ITPC at Oz following both fixation (panel A) and stimulus (panel B) presentations. Both fixation and stimulus presentations were associated with an increases in theta coherence across trials, peaking approximately 200ms after the event, in line with the P2 peak. Bayesian t-test results indicate that during the offline period, the data with the non-removed tRNS artefact differed substantially from both the data with the removed tRNS artefact and the sham data. Red and green lines represent a BF of log(3) and log(10), respectively.

4 Discussion

Given the aforesaid beneficial effects of tRNS on human behaviour, including cognitive training, we used mathematical training to understand how tRNS may alter neural signals associated with changes in performance. Participants were asked to complete a mathematical learning task whilst undergoing concurrent tRNS-EEG. During analysis, we removed the tRNS artefact from the ongoing EEG data and examined the underlying EEG signal change due to tRNS.

At the behavioural level, we found that tRNS influenced performance as was highlighted by learning slopes and most strongly in RTs. The RTs effect was qualified by individual differences, as seen in recent tRNS studies (Evans et al., 2018; Frank et al., 2018; Harty & Cohen Kadosh, 2019). Notably, in contrast to previous literature, we found the greatest improvement over time between the stimulation and sham groups in participants with the better baseline performance, with stimulation participants outperforming sham participants. This effect was not seen amongst participants with lower baseline performance. Our results suggest that tRNS can benefit individuals who already demonstrate high levels of performance. These results seem incongruous with stochastic resonance; instead, stimulation is honing the neural mechanisms of individuals demonstrating high levels of performance. In this case, tRNS may be boosting an already relatively high signal-to-noise ratio leading to greater performance. By comparison, participants with lower baseline performance have higher noise levels attenuating the effects of tRNS, or deficits outside the stimulated region, leading to little observable effect of stimulation. However, this explanation is ad-hoc and highlights the need for a more rigorous account of higher-performance individuals within the stochastic resonance framework, or an alternative framework, with clear predictions and testable hypotheses.

However, it should be noted that while the effect of tRNS is significant, there is still noticeable overlap between the sham and active groups after 18 blocks. Potentially, this may indicate that the benefits to learning induced by stimulation are small. Alternatively, it may indicate that the stimulation parameters selected were suboptimal. Whilst the parameters chosen here are representative of those common in the literature, there is little evidence to suggest that these are the ideal tRNS parameters for cognitive training. This excludes further parameter changes that might be needed on either a task or participant level.

In contrast to the RT data, neither participants' accuracy nor participants' reports on how much they used a retrieval vs. a calculation strategy appeared to differ substantially between stimulation groups over the course of the study. Instead, task-based (such as block number) effects and individual effects appeared to lead to the greatest changes in these two measures. As we would expect, participants' performance improved, and participants depended more on retrieval than calculation in the latter part of the task. Moreover, retrieval was more common amongst participants who demonstrated greater performance at baseline. By comparison, individuals demonstrating lower performance at baseline were more likely to rely on calculation during the earlier blocks than the high-performance individuals. This is consistent with the RT data as we would expect individuals relying on calculation to respond more slowly when solving arithmetic problems. It should be noted, however,

that as participants were given as much time as needed to answer questions, accuracy may not be the most sensitive measure for identifying differences between stimulation groups.

Together, the results suggest that tRNS might be increasing calculation speeds as training is progressing, mainly in those with better arithmetic performance at baseline. Alternatively, tRNS might impact another set of cognitive processes that support mathematical learning.

To understand the neural mechanisms underlying the effect of tRNS on cognitive training we effectively removed the tRNS artefact from the concurrently recorded EEG activity using ICA. When looking at the time-domain data, we note that the lack of difference between the stimulation data when the tRNS artefact is left in and when it is removed fits with our expectations. This is partially because the ongoing EEG signal underwent a band-pass filter to remove the highest power frequencies of the tRNS signal. Additionally, the grand averaging procedure used to create ERPs acts as a low-pass filter further attenuating the high-frequency tRNS. This was confirmed when investigating the standard deviation of the ERPs as the amount of variance in the data was considerably larger when the tRNS artefact remained in the data than when it was removed.

Similarly, the tRNS artefact removal appeared successful in the frequency domain. Prior to tRNS artefact removal, the EEG data were characterised by large peaks in the spectra at approximately 30Hz and every 10Hz following (aside from 50Hz), which were removed from subsequent ICA, leading to a periodogram similar to that seen in the sham group. This indicates that ICA can be used to clean up both the time and frequency domains, though if the analysis of interest is purely focused on average in the time domain, the ICA may not be necessary. Our suggested procedure allowed us to remove substantial differences between the sham group and the active tRNS group due to the tRNS artefact. Such a procedure allowed us the sensitivity to detect differences between the online and offline spectral activities, which were significant after the tRNS artefact was removed.

There were notable differences in the time-domain EEG signal following stimulus presentation between the stimulation and sham groups, which depended on whether the stimulus was presented during the online or offline period. The dissociation between the online and offline periods highlights the importance of removing tRNS artefacts in order to examine their immediate, rather than delayed, effect. During the online period, participants receiving stimulation demonstrated a greater negative potential over frontal sites approximately 100ms following stimulus presentation than their sham counterparts, which was not seen following fixation cross presentation. The N1 has been considered an attentional component reflecting perceptual attentional processes (Hillyard & Anllo-Vento, 1998), with higher attentional focus amongst participants with greater N1 amplitudes. In our sample, the increased N1 amplitude amongst tRNS participants following stimulus presentation may indicate greater attentional resources devoted to the problem presented. At a later time point, over occipital sites, a larger P2 was observed amongst tRNS participants compared to sham participants. In contrast to the N1 tRNS-induced effect that was limited to the online period, this effect was observed during both the online and offline periods. Similar to N1, P2 has often been associated with attentional processes (Hillyard & Anllo-Vento, 1998), with

previous work identifying a parietal N1-P2 complex associated with early perceptual attention gating (Singhal, Doerfling, & Fowler, 2002). Freunberger, Klimesch, Doppelmayr, and Höller (2007) found that P2 was associated with increased theta phase locking over parietal sites. According to Freunberger et al. (2007), this finding is related to attentional and working memory processes. However, whilst applying tRNS resulted in increased P2 amplitude in the tRNS group compared to the sham group, this alteration in P2 was not associated with similar differences in the theta-band ITPC. The theta ITPC values were reduced in the tRNS condition compared to the sham condition when the tRNS artefact remained in the data, suggesting that the increased P2 amplitude amongst the tRNS group was independent of increased phase locking as the P2 effect was seen in both the tRNS removed and the tRNS included datasets. In addition, the present results suggest that the association between P2 and ITPC, which has been based on a correlational data, might not be due to the causal effect of P2 on ITPC. Future studies could shed further light on the association between P2 and ITPC and the potential causal effect of ITPC on P2.

Overall, these results indicate that the effect of tRNS when applied above the dlPFC is characterised by EEG markers that are associated with greater attentional resources following the presentation of the operand, when the task demands greater working memory load. Given that we matched our participants based on their baseline arithmetic performance and that generally active tRNS yielded better performance than sham, a likely interpretation is that improvement in performance caused by tRNS reflects greater allocation of attentional resources during the training. Such mechanistic explanation clarify the relative broad improvement seen by tRNS in different tasks, including its effect on learning, where attentional resources play a key role (Simonsmeier et al., 2018), as well as its potential benefit in attention-deficit/hyperactivity disorder (Berger, Dakwar-Kawar, Grossman, Nahum, & Kadosh, 2019).

Finally, we also saw greater slow wave components, likely reflecting the late positive potential (LPP), in the active tRNS over occipital sites compared to the sham tRNS. However, this effect was only observed during the offline period. LPP has been previously identified in memory recognition experiments and is thought to be reflective of conscious recollection (Curran & Cleary, 2003; Curran, Schacter, Johnson, & Spinks, 2001). For example, larger amplitudes have been observed when participants were presented with information they knew compared to unfamiliar information (Turk et al., 2018). Whilst LPP is less well investigated in the mathematical learning and mental arithmetic literatures, studies from other domains have highlighted its modulation by top-down control processes and have suggested that it originates from a bidirectional prefrontal-occipitoparietal codependency (Moratti, Saugar, & Strange, 2011). This is reflective of the mathematical learning literature, which has suggested that the movement from calculation to retrieval is associated with changes in neural activation from frontal to parietal sites (Butterworth et al., 2011; Delazer et al., 2003; Menon, 2015; Laura Zamarian & Delazer, 2015). However, whilst this interpretation could explain the faster RT amongst the tRNS participants, it does not correspond to the participants' self-reported retrieval.

In contrast to the time-domain analysis, we did not see any effects of tRNS in the frequency-domain data. There are two possible explanations for this. Firstly, our analyses were limited to mid-line sites, whereas these changes may have been

lateralised. As noted earlier, switching from calculation to retrieval is associated with a shift from frontal to parietal areas, and this effect is mostly lateralised to the left hemisphere (Butterworth et al., 2011; Zamarian, Ischebeck, & Delazer, 2009; Zamarian & Delazer, 2015). Therefore, it is possible that we might see an effect over left parietal sites. Secondly, these changes may be either temporally or spatially transient, and therefore may not be obvious in the frequency-domain analysis.

Whilst the present results give a basis of understanding about how tRNS may lead to cognitive changes, in this case in mathematics, it also opens up a range of questions for further study. As learning is dependent on interpreting feedback signals and adjusting behaviour accordingly, comparison of EEG markers following correct and incorrect performances can provide a robust image of endogenous and exogenous feedback-driven learning mechanisms. This could further inform us about how the updating of outcome expectations is impacted by stimulation. Unfortunately, this could not be investigated here due to a low number of error trials (5.2% of trials). Furthermore, we note changes in attentional mechanisms associated with a 100–500Hz filter following application of 1mA stimulation. However, given the predicted curvilinear effect of tRNS on performance (Harty & Cohen Kadosh, 2019; van der Groen & Wenderoth, 2016b), there is much to gain by investigating how stimulation parameter changes lead to alterations in the event-related physiological response. Finally, whilst our findings do not fit a stochastic resonance interpretation of tRNS mechanism, as stimulation leads to the greatest improvement amongst high-performance individuals, they call for possible alternative theories to elegantly explain the effect of tRNS on the human brain.

References

Ahn, S., Mellin, J. M., Alagapan, S., Alexander, M. L., Gilmore, J. H., Jarskog, L. F., & Fröhlich, F. (2019). Targeting reduced neural oscillations in patients with schizophrenia by transcranial alternating current stimulation. *NeuroImage*, *186*(May 2018), 126–136. https://doi.org/10.1016/j.neuroimage.2018.10.056

Ambrus, G. G., Paulus, W., & Antal, A. (2010). Cutaneous perception thresholds of electrical stimulation methods: Comparison of tDCS and tRNS. *Clinical Neurophysiology*, *121*(11), 1908–1914. https://doi.org/10.1016/j.clinph.2010.04.020

Auvichayapat, N., Rotenberg, A., Gersner, R., Ngodklang, S., Tiamkao, S., Tassaneeyakul, W., & Auvichayapat, P. (2013). Transcranial direct current stimulation for treatment of refractory childhood focal epilepsy. *Brain Stimulation*, *6*(4), 696–700. https://doi.org/10.1016/j.brs.2013.01.009

Berger, I., Dakwar-Kawar, O., Grossman, E. S., Nahum, M., & Kadosh, R. C. (2019). Scaffolding the Attention-Deficit/Hyperactivity Disorder Brain Using Transcranial Direct Current and Random Noise Stimulation. *MedRxiv*, 19005983. https://doi.org/10.1101/19005983

Brem, A. K., Almquist, J. N. F., Mansfield, K., Plessow, F., Sella, F., Santarnecchi, E., … Erdogmus, D. (2018). Modulating fluid intelligence performance through combined cognitive training and brain stimulation. *Neuropsychologia*, *118*, 107–114. https://doi.org/10.1016/j.neuropsychologia.2018.04.008

Bulletin, P., Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., Iverson, G., & Bulletin, P. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin and Review*, *16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1). https://doi.org/10.18637/jss.v080.i01

Bürkner, P. C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *R Journal*, *10*(1), 395–411. https://doi.org/10.32614/RJ-2018-017

Butterworth, B., Varma, S., & Laurillard, D. (2011). Dyscalculia: From Brain to Education. *Science*, *332*(6033), 1049–1053. https://doi.org/10.1126/science.1201536

Cappelletti, M., Gessaroli, E., Hithersay, R., Mitolo, M., Didino, D., Kanai, R., … Walsh, V. (2013). Transfer of Cognitive Training across Magnitude Dimensions Achieved with Concurrent Brain Stimulation of the Parietal Lobe. *The Journal of Neuroscience*, *33*(37), 14899–14907. https://doi.org/10.1523/JNEUROSCI.1692-13.2013

Cappelletti, M., Pikkat, H., Upstill, E., Speekenbrink, M., & Walsh, V. (2015). Learning to integrate versus inhibiting information is modulated by age. *Journal of Neuroscience*, *35*(5), 2213–2225. https://doi.org/10.1523/JNEUROSCI.1018-14.2015

Cohen, M. X. (2014). *Analyzing Neural Time Series Data. Igarss 2014*. https://doi.org/10.1007/s13398-014-0173-7.2

Cragg, L., & Gilmore, C. (2014). Skills underlying mathematics: The role of executive function in the development of mathematics proficiency. *Trends in Neuroscience and Education*, *3*(2), 63–68. https://doi.org/10.1016/j.tine.2013.12.001

Curran, T., & Cleary, A. M. (2003). Using ERPs to dissociate recollection from familiarity in picture

recognition. *Cognitive Brain Research*, *15*(2), 191–205. https://doi.org/10.1016/S0926-6410(02)00192-1

Curran, T., Schacter, D. L., Johnson, M. K., & Spinks, R. (2001). Brain potentials reflect behavioral differences in true and false recognition. *Journal of Cognitive Neuroscience*, *13*(2), 201–216. https://doi.org/10.1162/089892901564261

Delazer, M., Domahs, F., Bartha, L., Brenneis, C., Lochy, A., Trieb, T., & Benke, T. (2003). Learning complex arithmetic - An fMRI study. *Cognitive Brain Research*, *18*(1), 76–88. https://doi.org/10.1016/j.cogbrainres.2003.09.005

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Dowker, A., Sarkar, A., & Looi, C. Y. (2016). Mathematics anxiety: What have we learned in 60 years? *Frontiers in Psychology*, *7*(APR). https://doi.org/10.3389/fpsyg.2016.00508

Evans, C., Banissy, M. J., & Charlton, R. A. (2018). The efficacy of transcranial random noise stimulation (tRNS) on mood may depend on individual differences including age and trait mood. *Clinical Neurophysiology*, *129*(6), 1201–1208. https://doi.org/10.1016/j.clinph.2018.03.012

Fertonani, A., Pirulli, C., & Miniussi, C. (2011). Random Noise Stimulation Improves Neuroplasticity in Perceptual Learning. *Journal of Neuroscience*, *31*(43), 15416–15423. https://doi.org/10.1523/JNEUROSCI.2002-11.2011

Frank, B., Harty, S., Kluge, A., & Cohen Kadosh, R. (2018a). Learning while multitasking: short and long-term benefits of brain stimulation. *Ergonomics*, *61*(11), 1454–1463. https://doi.org/10.1080/00140139.2018.1563722

Frank, B., Harty, S., Kluge, A., & Cohen Kadosh, R. (2018b). Learning while multitasking: short and long-term benefits of brain stimulation. *Ergonomics*, *61*(11), 1454–1463. https://doi.org/10.1080/00140139.2018.1563722

Freunberger, R., Klimesch, W., Doppelmayr, M., & Höller, Y. (2007). Visual P2 component is related to theta phase-locking. *Neuroscience Letters*, *426*(3), 181–186. https://doi.org/10.1016/j.neulet.2007.08.062

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, *24*(6), 997–1016. https://doi.org/10.1007/s11222-013-9416-2

Harty, S., & Cohen Kadosh, R. (2019). Suboptimal Engagement of High-Level Cortical Regions Predicts Random-Noise-Related Gains in Sustained Attention. *Psychological Science*, 095679761985665. https://doi.org/10.1177/0956797619856658

Herpich, F., Melnick, M. D., Agosta, S., Huxlin, K. R., Tadin, D., & Battelli, L. (2019). Boosting Learning Efficacy with Noninvasive Brain Stimulation in Intact and Brain-Damaged Humans. *The Journal of Neuroscience*, *39*(28), 5551–5561. https://doi.org/10.1523/jneurosci.3248-18.2019

Hillyard, S. A., & Anllo-Vento, L. (1998). Event-related brain potentials in the study of visual selective attention. *Proceedings of the National Academy of Sciences*, *95*(3), 781–787. https://doi.org/10.1073/pnas.95.3.781

Inukai, Y., Saito, K., Sasaki, R., Tsuiki, S., Miyaguchi, S., Kojima, S., … Onishi, H. (2016). Comparison of

three non-invasive transcranial electrical stimulation methods for increasing cortical excitability. *Frontiers in Human Neuroscience*, *10*(DEC2016), 1–7. https://doi.org/10.3389/fnhum.2016.00668

JASP Team. (2019). JASP. *[Computer Software]*.

Javadi, A. H., & Cheng, P. (2013). Transcranial direct current stimulation (tDCS) enhances reconsolidation of long-term memory. *Brain Stimulation*, *6*(4), 668–674. https://doi.org/10.1016/j.brs.2012.10.007

Jeffreys, H. (2006). Theory of probability. *Community Care*. Clarendon Press. Retrieved from https://global.oup.com/academic/product/the-theory-of-probability-9780198503682?cc=gb&lang=en&

Lambert, B. (2018). *A student's guide to Bayesian statistics.* Sage.

Menon, V. (2015). Arithmetic in child and adult brain. In R. Cohen Kadosh & A. Dowker (Eds.), *The Oxford handbook of numerical cognition*. Oxford University Press.

Moratti, S., Saugar, C., & Strange, B. A. (2011). Prefrontal-Occipitoparietal Coupling Underlies Late Latency Human Neuronal Responses to Emotion, *31*(47), 17278–17286. https://doi.org/10.1523/JNEUROSCI.2917-11.2011

O'Hara, R. B., & Sillanpää, M. J. (2009). A review of bayesian variable selection methods: What, how and which. *Bayesian Analysis*, *4*(1), 85–118. https://doi.org/10.1214/09-BA403

Pachella, R. G. (1974). The interpretation of reaction time in information processing research. *Human Information Processing: Tutorials in Performance and Cognition*, (November), 41–82.

Parsons, S., & Bynner, J. (2006). Does numeracy matter more? national research and development centre for adult literacy and numeracy. *Institute of Education*. Retrieved from www.ioe.ac.uk/bedfordgroup

Pasqualotto, A. (2016). Transcranial random noise stimulation benefits arithmetic skills. *Neurobiology of Learning and Memory*, *133*, 7–12. https://doi.org/10.1016/j.nlm.2016.05.004

Piironen, J., & Vehtari, A. (2017). Comparison of Bayesian predictive methods for model selection. *Statistics and Computing*, *27*(3), 711–735. https://doi.org/10.1007/s11222-016-9649-y

Pirulli, C., Fertonani, A., & Miniussi, C. (2013). The role of timing in the induction of neuromodulation in perceptual learning by transcranial electric stimulation. *Brain Stimulation*, *6*(4), 683–689. https://doi.org/10.1016/j.brs.2012.12.005

Popescu, T., Krause, B., Terhune, D. B., Twose, O., Page, T., Humphreys, G., … Cohen Kadosh, R. (2016). Transcranial random noise stimulation mitigates increased difficulty in an arithmetic learning task. *Neuropsychologia*, *81*, 255–264. Retrieved from https://www.sciencedirect.com/science/article/pii/S0028393215302682?via%3Dihub

Popescu, T., Krause, B., Terhune, D. B., Twose, O., Page, T., Humphreys, G., & Kadosh, R. C. (2016). Neuropsychologia Transcranial random noise stimulation mitigates increased dif fi culty in an arithmetic learning task. *Neuropsychologia*, *81*, 255–264. https://doi.org/10.1016/j.neuropsychologia.2015.12.028

Prichard, G., Weiller, C., Fritsch, B., & Reis, J. (2014). Effects of different electrical brain stimulation protocols on subcomponents of motor skill learning. *Brain Stimulation*, *7*(4), 532–540. https://doi.org/10.1016/j.brs.2014.04.005

Rickard, T. C. (1997). Bending the Power Law: A CMPL Theory of Strategy Shifts and the Automatization of Cognitive Skills. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/0096-3445.126.3.288

Rstudio Team. (2016). RStudio: Integrated development for R. RStudio, Inc., Boston MA. *RStudio*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-20966-6

Rufener, K. S., Ruhnau, P., Heinze, H., & Zaehle, T. (2017). Transcranial Random Noise Stimulation (tRNS) Shapes the Processing of Rapidly Changing Auditory Information. *Frontiers in Cellular Neuroscience*, *11*(June), 1–11. https://doi.org/10.3389/fncel.2017.00162

Santarnecchi, E., Brem, A.-K. K., Levenbaum, E., Thompson, T., Kadosh, R. C., Pascual-Leone, A., … Pascual-Leone, A. (2015). Enhancing cognition using transcranial electrical stimulation. *Current Opinion in Behavioral Sciences*, *4*, 127–178. https://doi.org/10.1016/j.cobeha.2015.06.003

Sella, F., Raz, G., & Kadosh, R. C. (2019). When randomisation is not good enough: Matching groups in intervention studies. *Submitted*. Retrieved from https://osf.io/6jfvk/?view_only=8d405f7b794d4e3bbff7e345e6ef4eed

Sella, F., Sader, E., Lolliot, S., & Cohen Kadosh, R. (2016). Basic and advanced numerical performances relate to mathematical expertise but are fully mediated by visuospatial skills. *Journal of Experimental Psychology: Learning Memory and Cognition*, *42*(9), 1458–1472. https://doi.org/10.1037/xlm0000249

Simonsmeier, B. A., Grabner, R. H., Hein, J., Krenz, U., & Schneider, M. (2018). Electrical brain stimulation (tES) improves learning more than performance: A meta-analysis. *Neuroscience and Biobehavioral Reviews*, *84*(November 2017), 171–181. https://doi.org/10.1016/j.neubiorev.2017.11.001

Singhal, A., Doerfling, P., & Fowler, B. (2002). Effects of a dual task on the N100-P200 complex and the early and late Nd attention waveforms. *Psychophysiology*, *39*(2), 236–245. https://doi.org/10.1111/1469-8986.3920236

Sitzer, D. I., Twamley, E. W., & Jeste, D. V. (2006). Cognitive training in Alzheimer's disease: A meta-analysis of the literature. *Acta Psychiatrica Scandinavica*, *114*(2), 75–90. https://doi.org/10.1111/j.1600-0447.2006.00789.x

Snowball, A., Tachtsidis, I., Popescu, T., Thompson, J., Delazer, M., Zamarian, L., … Cohen Kadosh, R. (2013). Report Long-Term Enhancement of Brain Function and Cognition Using Cognitive Training and Brain Stimulation. *Current Biology*, *23*(Figure 1), 987–992. https://doi.org/10.1016/j.cub.2013.04.045

Terney, D., Chaieb, L., Moliadze, V., Antal, A., & Paulus, W. (2008). Increasing Human Brain Excitability by Transcranial High-Frequency Random Noise Stimulation. *Journal of Neuroscience*, *28*(52), 14147–14155. https://doi.org/10.1523/JNEUROSCI.4248-08.2008

Treasure, T., & MacRae, K. D. (1998). Extending of benefits of breadt cancer screening. *BMJ (Clinical Research Ed.)*, *317*(August), 360–362. https://doi.org/10.1136/bmj.317.7155.362

Tsai, N., Buschkuehl, M., Kamarsu, S., Shah, P., Jonides, J., & Jaeggi, S. M. (2018). (Un)Great Expectations: The Role of Placebo Effects in Cognitive Training. *Journal of Applied Research in Memory and Cognition*, (2017). https://doi.org/10.1016/J.JARMAC.2018.06.001

Turk, K. W., Elshaar, A. A., Deason, R. G., Heyworth, N. C., Nagle, C., Frustace, B., … Budson, A. E. (2018). Late Positive Component Event-related Potential Amplitude Predicts Long-term Classroom-based Learning. *Journal of Cognitive Neuroscience*, *30*(9), 1323–1329. https://doi.org/10.1162/jocn_a_01285

van der Groen, O., Tang, M. F., Wenderoth, N., & Mattingley, J. B. (2018). Stochastic resonance enhances the rate of evidence accumulation during combined brain stimulation and perceptual decision-making. *PLOS Computational Biology*, *14*(7), e1006301. https://doi.org/10.1371/journal.pcbi.1006301

van der Groen, O., & Wenderoth, N. (2016a). Transcranial random noise stimulation of visual cortex: Stochastic resonance enhances central mechanisms of perception. *Journal of Neuroscience*, *36*(19), 5289–5298. https://doi.org/10.1523/JNEUROSCI.4519-15.2016

van der Groen, O., & Wenderoth, N. (2016b). Transcranial random noise stimulation of visual cortex: Stochastic resonance enhances central mechanisms of perception. *Journal of Neuroscience*, *36*(19), 5289–5298. https://doi.org/10.1523/JNEUROSCI.4519-15.2016

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Zamarian, L, Ischebeck, A., & Delazer, M. (2009). Neuroscience and Biobehavioral Reviews Neuroscience of learning arithmetic — Evidence from brain imaging studies, *33*, 909–925. https://doi.org/10.1016/j.neubiorev.2009.03.005

Zamarian, Laura, & Delazer, M. (2015). Arithmetic learning in adults: Evidence from brain imaging. In R. Cohen Kadosh & A. Dowker (Eds.), *The Oxford handbook of numerical cognition*. Oxford University Press.

5 Supplementary

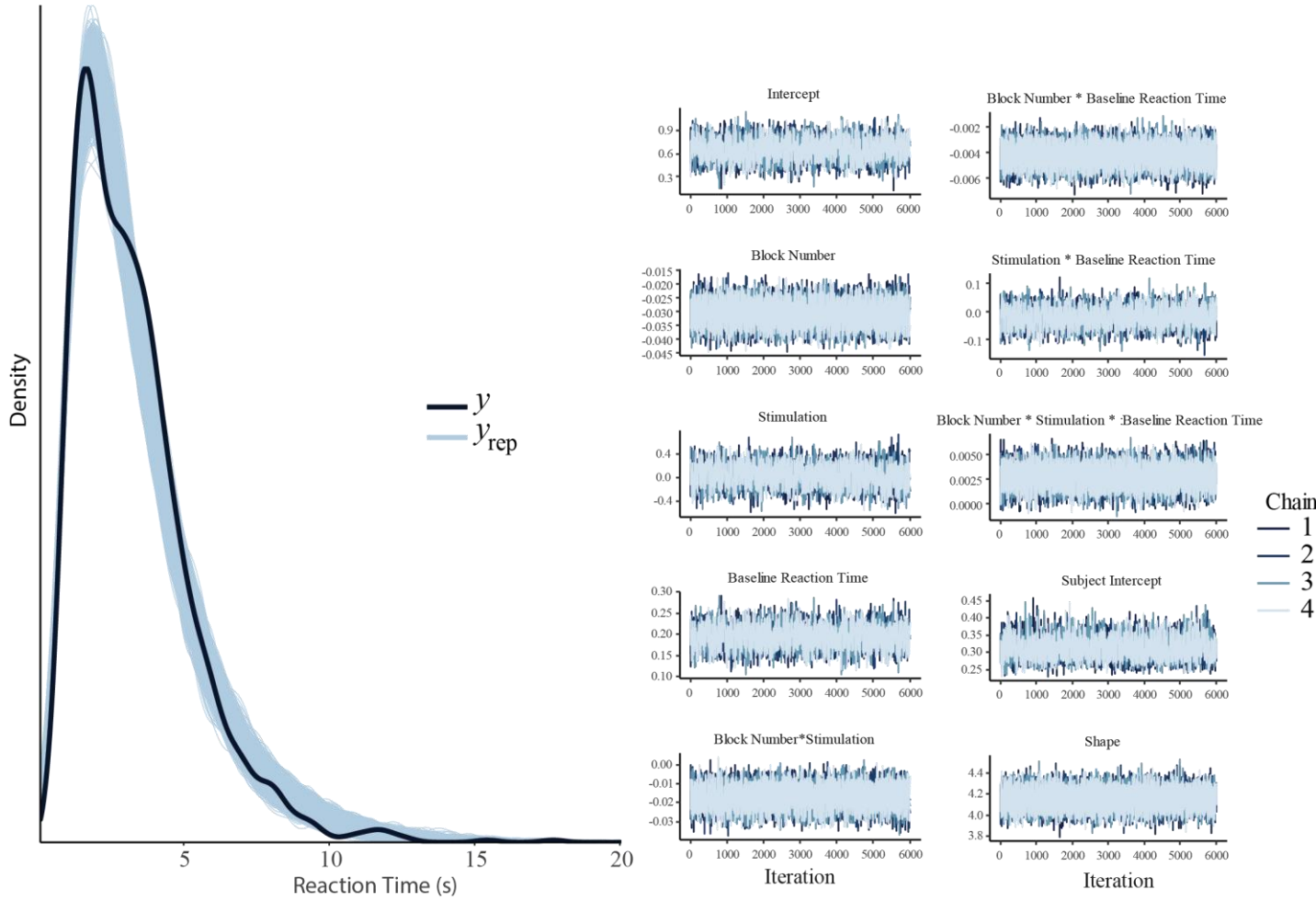
Table S1: List of potential operands that can be combined to create the arithmetic problems presented to the participants.

First Operand
29
23
13
19
28
29
17
26
Second Operand
3
4
2
6
8
7

RT

The 95% HDI values for the parameters indicated decreasing reaction times over blocks ([-0.04, -0.02]) and a positive relationship between baseline reaction time and task reaction time ([0.15, 0.24]). The 95% HDI values for the interaction between block number and stimulation ([-0.03, -0.01]) and block number and baseline reaction time ([-0.01, -0.002]).

Figure S1: Graphical post-posterior check (left) and HMC chain trace plots (right) for the selected median RT multilevel model. The results of both tests suggest that the model fits the data well. In the post-posterior check plot, y represents the actual data, whereas each line in y_{rep} represents a simulated data set from the posterior predictive distribution.



Calculation/Retrieval results

The 95% HDI values for the parameters suggested that only the online period ([-2.79, -1.54]) and the interaction between the online period and baseline RT ([-0.26, -0.0001]) had a notable impact on whether participants were more likely to report using calculation or retrieval. Specifically, the offline period was associated with greater retrieval than calculation compared to the online period, and individuals with higher baseline RTs were more likely to depend on calculation in the online period (at the beginning of the task) than those with lower baseline RTs.

Figure S2: Graphical post-posterior check (left) and HMC chain trace plots (right) for the selected retrieval vs. calculation multilevel model. The results of both tests suggest that the model fits the data well. In the post-posterior check plot, y represents the actual data, whereas each line in y_{rep} represents a simulated data set from the posterior predictive distribution.

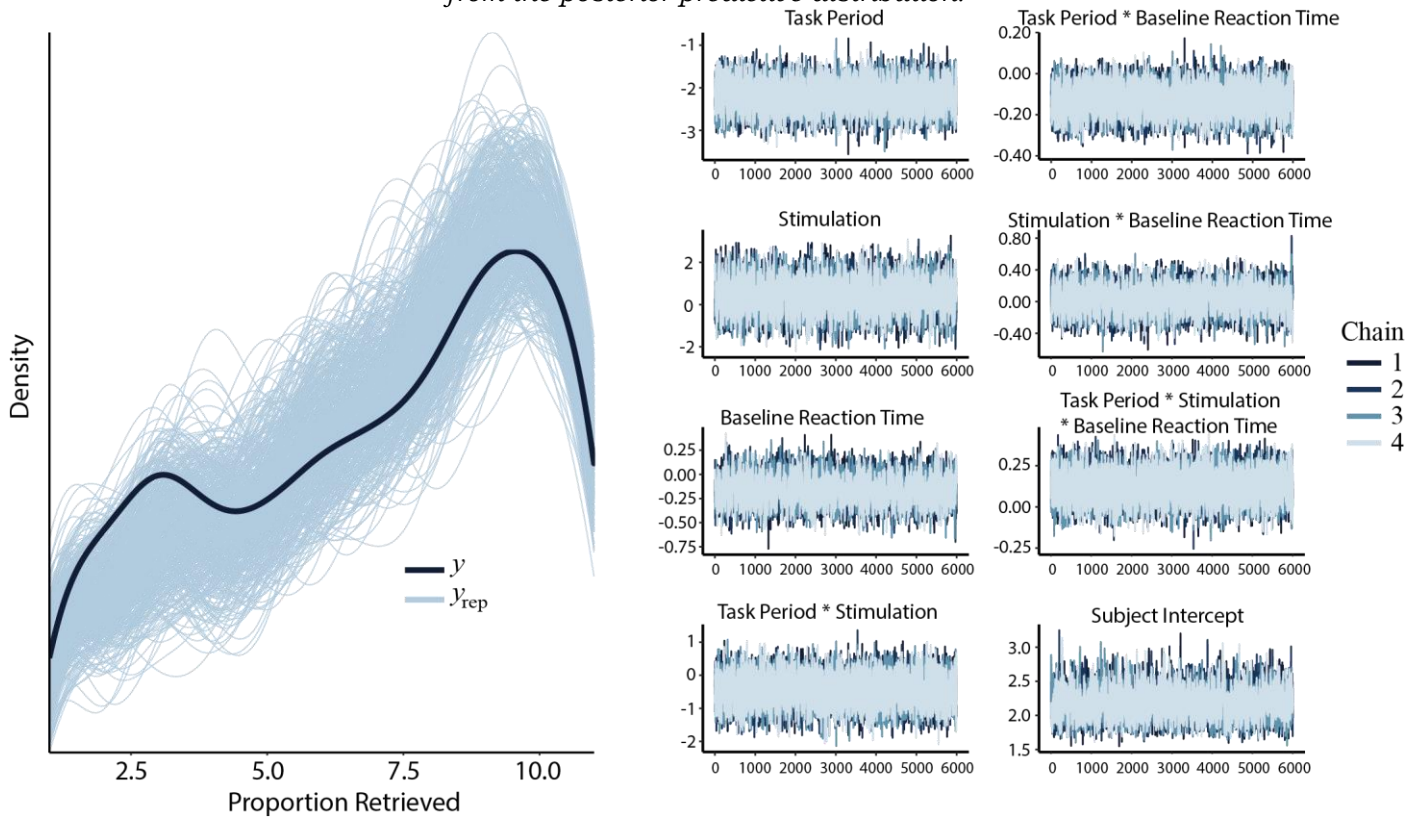


Figure S3: The marginal effects of task period (online vs. offline) and the two-way interaction between baseline

RT and task period on retrieval scores.

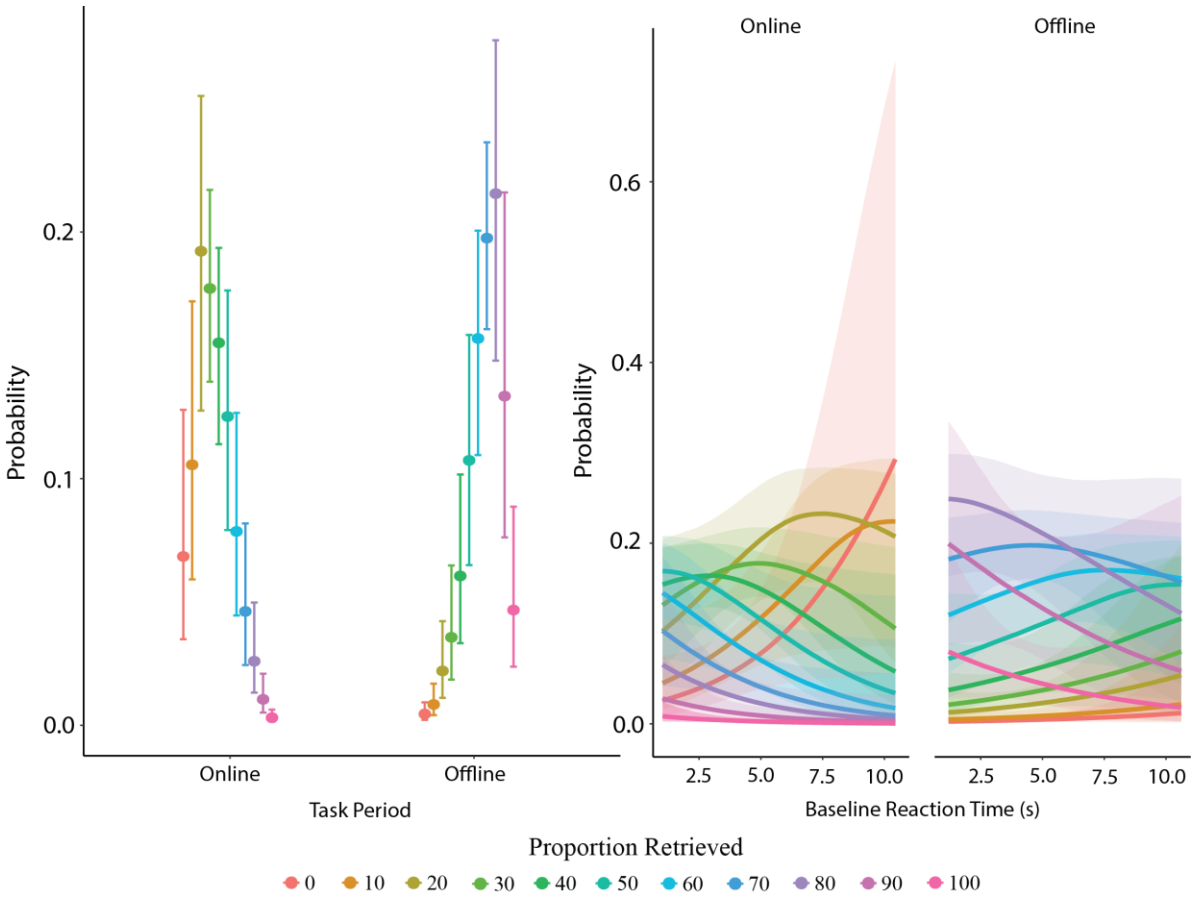
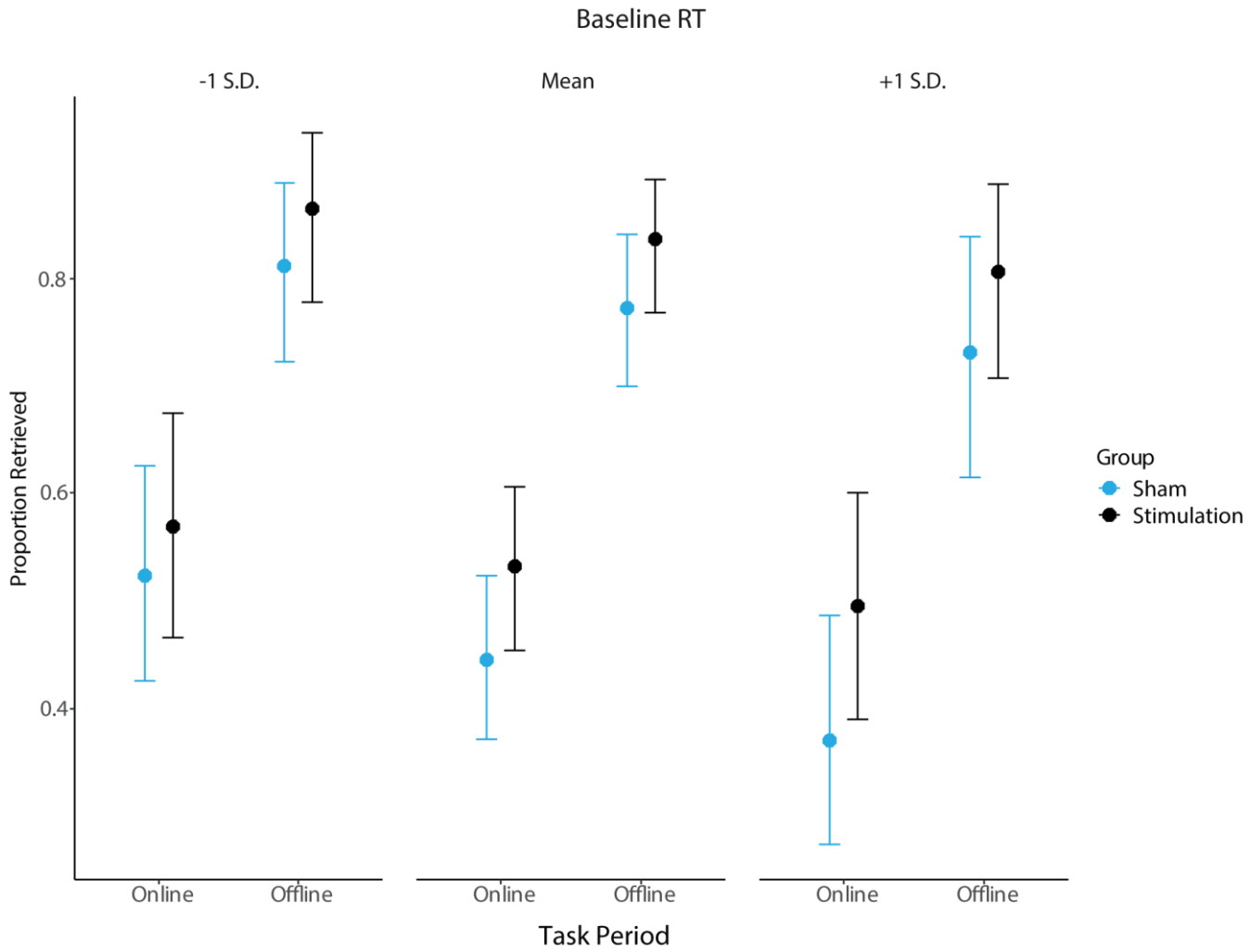


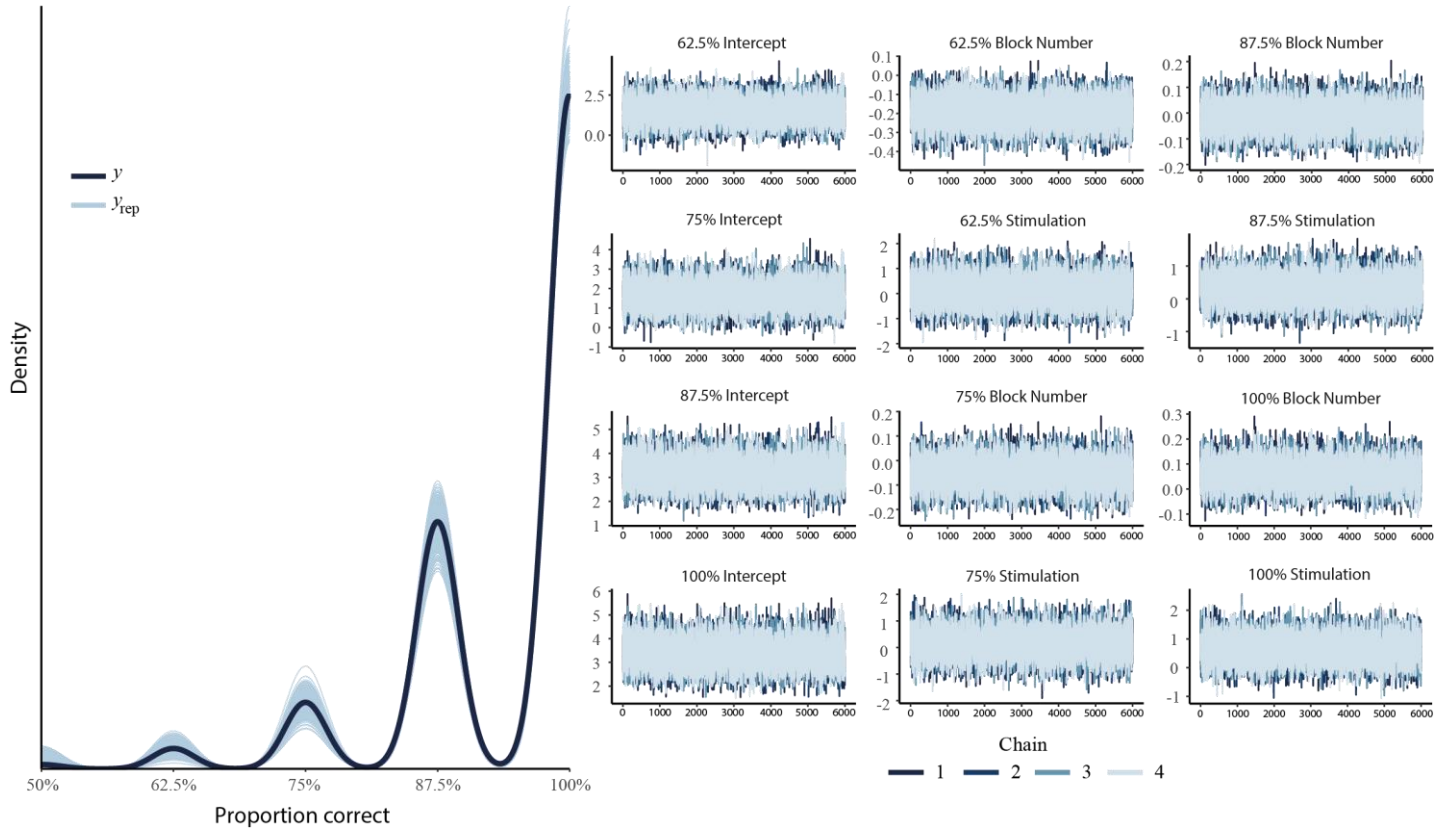
Figure S4. Marginal effects for the proportion retrieved (vs. calculated) regressed back on the 3-way interaction between

stimulation, task period (online vs. offline), and baseline RT.



Accuracy

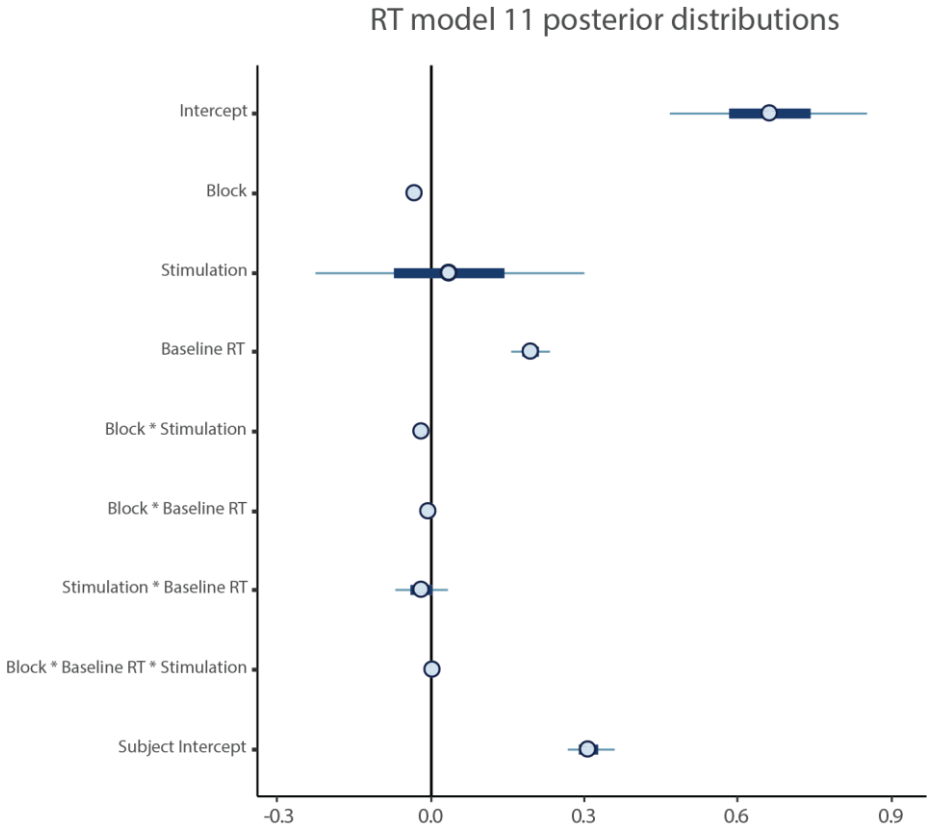
Figure S5: Graphical poster posterior check (left) and HMC trace plots (right) for the selected accuracy model. The results of both tests suggest that the model fits the data well. In the post-posterior check plot, y represents the actual data, whereas each line in y_{rep} represents a simulated data set from the posterior predictive distribution.



Bayes factors

In contrast to the RT out-of-sample models, the BF values indicated that the reaction time data were considerably more likely under the simple model (Model 8: RT ~ Block * Baseline RT) compared to the complex model (Model 11: RT ~ Block * Baseline RT * Stimulation; $BF_{10} = 3.92^{e5}$). This may due to the relatively limited influence of stimulation on reaction time data, with no apparent main effect on the data (see **Figure S5**), and considerable overlap between the two stimulation groups' performance except at the end of the learning period. The additional complexity of the model may not be compensated for by the relatively minor impact of stimulation on the current data. However, the reaction time data were more likely under the model containing the two-way interaction between block and baseline RT compared to the next best model (Model 1: RT ~ block; $BF_{10} = 7.67^{e15}$), which reflects the differences between participants of different baseline performance across task blocks.

Figure S6: Draws from the posterior distributions for all predictors from Model 11 of the reaction time data. The thick bars indicate the 50% interval and the thin bars indicate the 90% interval.



When comparing the BF values for the accuracy data, we tested Model 4 (accuracy ~ stimulation + block) against Model 1 (accuracy ~ block) to investigate the benefits of including stimulation for participants' accuracy. The accuracy data were equally likely under either model $BF_{10} = 0.97$,

providing no compelling evidence in favour of the more complex model over the simple model. This is consistent with our out-of-sample model cross-validation results, indicating that the largest contributor to participants' accuracy is block number.

Similar to the RT data, for calculation/retrieval results the three-way interaction was the best predictor when using the out-of-sample value. The calculation/retrieval data were more likely under the next best model (Model 5: Calculation/Retrieval ~ Online * Baseline RT; BF_{10} = 761.63). However, given that the results of the out-of-sample cross-validation suggested that block number was the most important predictor of retrieval (given that models that included online effects with a considerable drop-off were better at predicting out-of-sample data than models without online effects as a predictor), we tested whether a simple block-only model predicted the data better. The calculation/retrieval data were considerably more likely under the block-only model compared to the three-way interaction mode (BF_{10} = 2220.92), though the likelihood of the data was largely comparable under either of the more simple models, but slightly in favour of the simple Online * Baseline RT interaction model (BF_{10} = 2.97).