

Supplemental Material: Blinding is compromised for transcranial direct current stimulation at 1 mA for 20 minutes in young healthy adults

Zsolt Turi¹, Gábor Csifcsák², Nya Mehnwolo Boayue², Per Aslaksen², Andrea Antal^{1,3}, Walter Paulus¹, Josephine Groot^{2,4}, Guy E. Hawkins⁵, Birte Forstmann⁴, Alexander Opitz⁶, Axel Thielscher^{7,8}, Matthias Mittner^{2*}

¹*Department of Clinical Neurophysiology, University Medical Center Goettingen*

²*Department of Psychology, University of Tromsø*

³*Medical Psychology, Otto-Guericke University Magdeburg*

⁴*Integrative Model-based Cognitive Neuroscience Research Unit, University of Amsterdam*

⁵*School of Psychology, University of Newcastle*

⁶*Department of Biomedical Engineering, University of Minnesota*

⁷*Department of Electrical Engineering, Technical University of Denmark*

⁸*Danish Research Centre for Magnetic Resonance, Centre for Functional and Diagnostic Imaging and Research, Copenhagen University Hospital Hvidovre, Denmark*

1. Supplemental Analyses

1.1. Analysis methods – details

It has recently been pointed out that treating ordinal variables, such as responses to Likert-scales, as metric can have a strong impact on the validity of the statistical analyses: It can distort effect-size estimates, inflate error rates and cause other problems (Liddell and Kruschke, 2018, Bürkner and Vuorre (2018)). Despite these severe problems, treating ordinal variables as metric is extremely common in psychology and related disciplines (Liddell and Kruschke, 2018). A more appropriate approach uses specific ordinal

*Corresponding author: matthias.mittner@uit.no

models such as ordinal probit regression (Bürkner and Vuorre, 2018). Here we use a Bayesian approach to estimate such a model as implemented in the R-package `brms` (Bürkner, 2018). We used the default priors implemented in the `brms`-package (Bürkner, 2018). Those include uniform priors for all regression coefficients and a Students' t-prior centered at zero with $df=3$ and $sd=10$ for the thresholds. In addition, we use Bayesian methods for model selection based on the leave-one-out crossvalidation criterion (LOOIC; Vehtari et al. (2017)) procedure. Lower scores of the LOOIC imply a better model.

All Bayesian analyses used four independent chains with 2000 samples each. We treated the first 1000 iterations as a warm-up phase such that all inferences are based on 4000 iterations. We ensured that the Gelman-Rubin diagnostic (Gelman et al., 1992) $\hat{R} \leq 1.05$ for all variables in all models. In addition, traceplots for all variables were visually inspected and no anomalies were found.

1.2. Model-selection details

We tested a sequence of models of increasing complexity independently for each of the two outcome measures (guessing sham vs. anodal stimulation and discomfort caused by stimulation). We started with a model that contains no factors (Null-model) and then added each of the factors (lab: AMS, GOE and TRM; stimulation condition: sham vs. real) in various combinations. In addition, we tested models that included the achieved impedance at the electrodes (see next section for a rationale). We also relaxed the assumption of homoscedasticity by testing models that allowed for unequal variances across labs and stimulation conditions. A summary of all models is provided in Table 1 for the blinding question and in Table 2 for the discomfort question. For each model, we list the estimated effect of the anodal stimulation and the LOOIC (Vehtari et al. (2017)).

For the blinding question, model-selection favored a model that included stimulation and lab as factors. Including their interaction decreased the model-selection score ($\Delta LOOIC=-3.97$, $SE=1.77$). Allowing unequal variances across labs or stimulation did not appear to strongly improve model-fit. We therefore report the model that included only Stimulation and Lab as factors in the main paper. However, the conclusions drawn from this model, in particular with respect to the effect of sham vs. anodal stimulation on blinding efficacy and discomfort, were robust against model choice: All models (except models 8 and 9 that included higher-order interaction terms)

Table 1: Summary of models for the blinding question.

Model	Description	Stimulation Effect			LOO	
		b^a	SE(b) ^b	95% HDI ^c	LOOIC ^d	SE(LOO)
0	Null	–	–	–	691.17	16.72
<i>equal variances</i>						
1	Lab	–	–	–	689.80	16.48
2	Stim(ulation)	0.35	0.15	[0.06, 0.65]	687.95	17.24
3	Imp(edance)	–	–	–	693.40	16.66
4	Stim+Lab	0.35	0.15	[0.06, 0.64]	686.58	17.00
5	Stim+Imp	0.36	0.15	[0.06, 0.65]	690.04	17.15
6	Lab+Imp	–	–	–	688.35	16.83
7	Lab+Stim+Imp	0.31	0.15	[0.01, 0.62]	686.67	17.21
8	Stim×Lab+Imp	0.16	0.26	[−0.35, 0.66]	689.45	17.21
9	Stim×Lab×Imp	0.95	0.89	[−0.85, 2.65]	696.83	17.51
<i>unequal variances lab</i>						
10	Stim+Lab	0.35	0.15	[0.06, 0.64]	686.58	17.02
<i>unequal variances stimulation</i>						
11	Stim+Lab	0.35	0.15	[0.05, 0.64]	686.52	16.96

^a Effect of anodal stimulation on 7-point scale
^b SE=standard error
^c HDI=Highest Density Interval
^d LOOIC=Leave-One-Out Information Criterion

showed a positive effect of real stimulation on both questions with the posterior 95% interval excluding zero. We disregard models 8 and 9 here since they show worse model-selection scores and produce a much higher posterior uncertainty around the estimates for the effects.

For the discomfort question, model selection also favored the inclusion of both Lab and Stimulation, while including their interaction decreased the score ($\Delta\text{LOOIC}=-2.42$, $\text{SE}=2.76$). For the discomfort question, including the impedance as a covariate resulted in slightly lower model-selection scores, though that improvement was not strong ($\Delta\text{LOOIC}=0.87$, $\text{SE}=3.48$). That pattern was consistent when allowing for heteroscedasticity (models 11 and 13, Table 2). The estimate for the effect of interest (i.e., the effect of stimulation condition on discomfort) was changed slightly when including impedance (model 4 without impedance: $b = 0.34$, $[0.04, 0.63]$, model 7 with impedance: $b = 0.29$ $[-0.02, 0.6]$) to the point that the 95% did include zero for the model including impedance. The effect of impedance was estimated close to zero, $b = 0.07$ $[-0.01, 0.14]$. We note that the effect of stimulation condition was positive across all model specifications, with the lower boundary of the effect fluctuating around zero. In addition, participants in both Göttingen and Tromsø reported more discomfort when compared to Amsterdam ($b_{\text{GOE}} = 0.43$ $[0.05, 0.80]$, $b_{\text{TRM}} = 0.67$ $[0.31, 1.06]$).

Diagnostic plots as well as posterior-predictive checks for all tested models can be found at our github-repository: https://github.com/ihrke/2018_tdcs_blinding/tree/master/analysis.

1.3. Achieved impedance as a moderator

Our experimental protocol required the impedance of the electrodes to be below 10 k Ω before the stimulation could be started (Boayue et al., in press). Even though that value is quite low in comparison to what other studies in the field use, it is possible that differences in impedance may have had effects on discomfort or the blinding properties of the protocol. In order to investigate whether there were differences in what impedances were realized across labs and stimulation condition, we ran a Bayesian ANOVA (Rouder et al., 2012) with impedance at the start of the stimulation as dependent variable and factors lab and stimulation condition (prior r -scale=0.5). Descriptive statistics for this variable are summarized in Figure 1.

We found that there were clear differences in the achieved impedances across labs ($\text{BF}_{10}=7.57 \times 10^{21}$). The Göttingen lab achieved the lowest impedances ($M = 3.3\text{k}\Omega$, $\text{SD}= 2.13$), followed by Amsterdam ($M = 6.05\text{k}\Omega$, $\text{SD}=\$

Table 2: Summary of models for the discomfort question.

Model	Description	Stimulation Effect			LOO	
		b^a	SE(b) ^b	95% HDI ^c	LOOIC ^d	SE(LOO)
0	Null	–	–	–	637.87	20.23
<i>equal variances</i>						
1	Lab	–	–	–	629.83	19.37
2	Stim(ulation)	0.33	0.15	[0.02, 0.63]	635.22	20.57
3	Imp(edance)	–	–	–	635.81	20.06
4	Stim+Lab	0.34	0.15	[0.04, 0.63]	627.33	19.70
5	Stim+Imp	0.30	0.16	[–0.01, 0.61]	634.56	20.32
6	Lab+Imp	–	–	–	627.38	18.90
7	Lab+Stim+Imp	0.29	0.16	[–0.02, 0.60]	626.46	19.29
8	Stim×Lab+Imp	-0.06	0.28	[–0.60, 0.48]	629.35	19.93
9	Stim×Lab×Imp	0.65	0.94	[–1.21, 2.44]	637.34	21.10
<i>unequal variances lab</i>						
10	Stim+Lab	0.34	0.16	[0.04, 0.64]	627.02	19.63
11	Stim+Lab+Imp	0.29	0.16	[–0.02, 0.60]	626.87	19.32
<i>unequal variances stimulation</i>						
12	Stim+Lab	0.34	0.15	[0.04, 0.63]	627.03	19.58
13	Stim+Lab+Imp	0.29	0.16	[–0.02, 0.60]	626.93	19.36

^a Effect of anodal stimulation on 7-point scale

^b SE=standard error

^c HDI=Highest Density Interval

^d LOOIC=Leave-One-Out Information Criterion

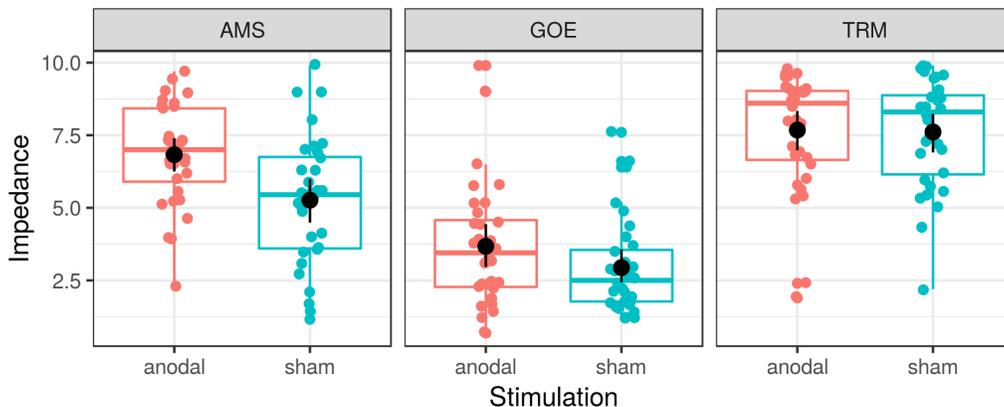


Figure 1: Impedance at the start of the experiment across labs and stimulation conditions. The protocol required the impedance to be below $10\text{k}\Omega$, placing a ceiling on the observed values. Black points and bars are means and confidence intervals. Colored dots show impedance for individual participants and boxplots summarize the distributions.

1.94) and Tromsø ($M = 7.65\text{k}\Omega$, $SD = 1.94$). The reason for the different impedance levels among the laboratories remains elusive. We speculate that this might be due to the different amount of conductive paste used by the data collectors or other circumstantial differences, such as the amount of time spent by the data collectors with the preparation of the electrode-skin interface. Note, however, that the stimulation equipment was identical and that the exact procedure was pre-registered and pre-described in fine detail to ensure comparability across labs. There was no evidence for differences between the stimulation conditions ($BF_{10} = 1.11$).

In order to ensure that the differences between labs in achieved impedance of the stimulation was not causing differences in reported discomfort or above-chance levels of guessing the real stimulation condition, we included impedance as a variable in our analysis (see Table 1 and Table 2). When including this variable, model-selection criteria were not generally improved: Model-selection seemed to be largely agnostic to the distinction between including or not including impedance as a covariate. The estimated positive effect of real stimulation on the scores on both answers was only minorly affected when including impedance.

1.4. Frequentist analysis treating ordinal scales as metric variables

Even though we believe that our reported Bayesian analyses are much more suitable for the analysis of ordinal-scale data (Liddell and Kruschke, 2018) we also report standard frequentist analyses that do not take the ordinal scale into account. We do this mostly of a desire to convince the reader that we did not cherry-pick our analysis method to achieve a particular result. Simple independent t-tests comparing sham vs. anodal stimulation find significant effects both for subjects guessing the stimulation condition ($t(185.9) = -2.38, p = 0.018, d = 0.34$) and skin sensations ($t(189.03) = -2.0, p = 0.047, d = 0.29$). The effect of actual stimulation condition on guessing stimulation condition is significant also when including lab-location in a 2 (sham vs. anodal) \times 3 (lab: GOE, TRM, AMS) ANOVA, $F(1, 186) = 5.74, p = 0.018, \eta^2 = 0.03$. In this analysis, neither the main effect of lab ($F(2, 186) = 2.67, p = 0.07, \eta^2 = 0.028$) nor the lab \times stimulation interaction were significant ($F(2, 186) = 0.25, p = 0.78, \eta^2 = 0.003$). A corresponding ANOVA with reported skin-sensations as dependent variable also showed stronger sensations during anodal stimulation ($F(1, 186) = 4.31, p = 0.039, \eta^2 = 0.023$). For this question, the main effect of lab was significant ($F(2, 186) = 8.46, p = 0.00030, \eta^2 = 0.083$) but the interaction between lab and stimulation was not, $F(2, 186) = 0.94, p = 0.39, \eta^2 = 0.01$.

References

- Boayue, N. M., Csifcsák, G., Aslaksen, P., Turi, Z., Antal, A., Groot, J. M., Hawkins, G., Forstmann, B., Thielscher, A., Opitz, A., Thielscher, A., Mittner, M., in press. Increasing propensity to mind-wander by transcranial direct current stimulation? a registered report. *European Journal of Neuroscience*.
- Bürkner, P.-C., 2018. Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal* 10 (1), 395–411.
URL <https://journal.r-project.org/archive/2018/RJ-2018-017/index.html>
- Bürkner, P. C., Vuorre, M., Sep 2018. Ordinal regression models in psychology: A tutorial.
URL psyarxiv.com/x8swp

- Gelman, A., Rubin, D. B., et al., 1992. Inference from iterative simulation using multiple sequences. *Statistical science* 7 (4), 457–472.
- Liddell, T. M., Kruschke, J. K., 2018. Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology* 79, 328–348.
- Rouder, J. N., Morey, R. D., Speckman, P. L., Province, J. M., 2012. Default bayes factors for anova designs. *Journal of Mathematical Psychology* 56 (5), 356–374.
- 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.