# Placebo and cognitive enhancement: the cognitive and neurological effects of conditioning and verbal suggestion

Hiralal, Kamil

Universiteit Leiden

# Placebo and cognitive enhancement: the cognitive and neurological effects of conditioning and verbal suggestion

Master Thesis


Student: Kamil Hiralal

Student Number: 1672959

Supervisor: dr. Michiel van Elk

Unit: Cognitive Psychology


Word Count: 8789

Academic Year: 2021-2022

Abstract

Previous research has shown that subjective cognitive performance can be enhanced, and that sense of agency can be lowered via verbal suggestion of brain stimulation. Mixed results have been found on the effect of placebo stimulation on error-related negativity (ERN) amplitude, which is an implicit outcome measure. This EEG study focused on the role of verbal suggestion and associative learning in eliciting placebo effects in subjective performance, sense of agency, and ERN amplitude. Using a within-subject design, we recorded EEG while participants ($n$ = 19) performed in a simple cognitive task. Participants were told that a sham brain stimulation device would either enhance (placebo condition) or impair (nocebo condition) their cognitive performance. Next, we used a conditioning phase in which we altered the task difficulty according to the experimental block in order to induce the association between task difficulty and proposed stimulation. After this conditioning phase, the task difficulty was equal across conditions. We found increased subjective performance in the nocebo condition, but not the placebo condition, compared to control. We found a lower sense of agency in the placebo condition, but not the nocebo condition, compared to control. Finally, we found no difference in ERN amplitude throughout conditions. These results are not in line with previous research. Our conditioning phase did not work as intended and therefor the results are difficult to interpret. In addition, based on the results of our post-test questionnaire, our verbal suggestion might have been too weak. Future research should try to replicate the earlier results and continue investigating possible (other) implicit outcome measures.

*Keywords:* Placebo/Nocebo effects, Subjective performance, Sense of agency, Verbal Suggestion, Associative learning

While the Golden State Warriors were defending their NBA title back in 2016, one of their players made the news with a noteworthy story. James Michael McAdoo tweeted out a picture in which he was wearing a device that was pulsing electrical signals through his scalp. The physical staff of the Golden State Warriors confirmed that a couple of players on their roster had been using this brain stimulation device for months with the goal of optimizing their basketball performance (Hutchinson, 2016). Imagine watching your favourite sports player using brain stimulation. How would you evaluate the efficacy of brain stimulation after you witness another remarkable performance of your idol? What would you expect to happen when you ever get the opportunity to have your own brain stimulated?

As a scientific community it is our responsibility to properly disseminate scientific information to a lay audience and make sure findings are not interpreted disproportionately by the popular media (Ritchie, 2020). One line of research has investigated the neuroimage bias, which is valuing neuroscientific information as more valuable than "normal" information. Indeed, there are studies which found that when people are presented with explanations of scientific phenomena, they tend to believe bad explanations which include neuroscientific methods over explanations without neuroscientific methods (Ali et al., 2014; Hopkins et al. 2016; Michael et al., 2013; Weisberg et al., 2008). However, a different set of studies found no such effect (Fernandez-Duque et al., 2015; Gruber & Dickenson, 2012; Michael et al., 2013).

Baker and colleagues (2017) reason that this might have to do with the fact that most studies that found a neuroimage bias used a repeated measures design, opposed to studies that did not find an effect. In repeated measures designs, explanations containing neuroscientific information are presented alongside explanations without neuroscientific information. Thus, evaluations of the explanations are relative to each other, which might cause explanations without neuroscientific information to be evaluated as less credible than explanations including neuroscientific information (Baker et al., 2017). This theory is supported by findings from a

multi-experiment paper of Schweitzer and colleagues (2013) who only found a neuroimage bias when multiple explanations are presented together and not when they are presented in isolation. As of yet, it is not exactly clear when and for whom neuroimage bias occurs and future research can elucidate which individual traits might be of importance.

Ali et al. (2014) have coined the term *neuroenchantment* to describe when people tend to overestimate neuroscientific advances. For example, it is a myth that the current neuroscientific tools allow us to disentangle the entire human mind and enables us to read thoughts (Ali et al, 2014; Haynes, 2012). Neuroenchantment will raise the overall expectations of the efficacy of neuroscientific methods. It has been shown that expectations can elicit placebo effects (Rief & Petrie, 2016; Schwarz et al., 2016) In this study, we investigated if verbal suggestion about brain stimulation can evoke placebo effects on a subjective and a neurological level.

**Placebo effects**

Placebo effects are experiences of physical or cognitive benefits from an inactive treatment or drug (Gibbs, 2010; Price et al., 2008). Nocebo effects are the phenomena when someone experiences detrimental physical or cognitive effects from an inactive treatment or drug (Colloca & Miller, 2011). An example of a placebo effect was found in a double-blind study investigating whether transcranial magnetic stimulation (Conforto et al., 2014), which is a magnetic brain stimulation device (Klomjai et al., 2015), reduced chronic migraine in comparison to sham stimulation. Participants were blind to their condition but were aware of the possibility of receiving sham stimulation. To their surprise, Conforto and colleagues (2014) found a greater decrease in headache after placebo stimulation compared to actual stimulation. The authors speculate that they found this placebo effect because sham stimulation induces a tingling sensation at the scalp, which gives the suggestion of an active treatment. The study shows that the suggestion of being treated can reduce experienced symptoms.

One very important factor for eliciting placebo effects is context (Benedetti, 2002; Miller & Kaptchuck, 2008). In 1955, Balint defined context as everything that "surrounds the patient in treatment". An example of the influence of context comes from the classical study of Ulrich (1984), who found that patients with a window view on trees had a shorter post-surgical recovery period compared to patients who had a view of a brick wall. Ulrich (1984) theorized that nature scenes elicit positive feelings in comparison to urban scenes and that nature scenery will reduce stressful feelings. As a result, these positive feelings will stimulate a speedy recovery (Ulrich, 1984). Not only can context influence one's feelings after treatment, context can also change the expectancies of treatments, which is crucial for elicitation of placebo effects.

How context can shape the expectancies of a treatment is shown by the Conforto and colleagues (2014) study. Their use of brain stimulation induces a neuroscientific context which could have raised the overall belief in the efficacy of the treatment, especially since people tend to put their trust in neuroscientific methods (Ali et al., 2014; Hopkins et al., 2016; Michael et al., 2013; Weisberg et al., 2008). I will now discuss the theoretical framework that is commonly used to describe how expectancies elicit placebo effects.

**Predictive Processing Framework**

The predictive processing framework suggests that we make top-down predictions about the world and compare those predictions to the sensory input we receive. The predictions are based on earlier experiences and are updated when a mismatch occurs between our prediction and our sensory input (Clarke, 2013; Schwengerer, 2018). Consider the following example: you are excited about having a meal at your favourite restaurant, but once you take your first bite the food tastes horrible. Your previous experiences at this restaurant were great, but now you are very disappointed. You made top-down predictions about how the food would taste, but your sensory input (tasting the food) resulted in a prediction error. As a result, you update your

future predictions and will choose a different restaurant next time.

Ongaro and Kaptchuk (2019) gave an intuitive example of when prior experience affects future predictions. Imagine yourself in a snake-infested forest. You feel something moving around your feet and you feel startled. Based on your beliefs about the forest you think you feel a snake, but after closer examination you see that it is just a twig (Ongaro & Kaptchuk, 2019). From this example the importance of context on our predictions of the world becomes evident. Now imagine that you are in the same forest at night. You observe the same movement around your feet, and you think that you see a snake. In the dark you can't make the clear distinction between a snake and a twig, and you observe a snake, while in reality it was just another twig. The ambiguity of the situation has caused you to observe the world based on your top-down expectations of what the world is like. This is a key aspect of the predictive processing framework: we do not directly observe the world as it is, but we observe our best guess based on sensory information (Ongaro & Kaptchuk, 2019).

Two key factors in updating predictions are magnitude and precision (Büchel et al., 2014). Magnitude refers to the strength of a stimulus. When the food in a restaurant is extremely tasty, your predictions for the future will be better compared to when the food is only so-so. Precision relates to the certainty that something is going to happen. When you repeatedly find a twig at your feet in the forest, it will become less likely that you will predict to detect a snake in the future.

In relation to placebo effects, the predictive processing framework suggests that people experience treatment effects when they receive external cues of that an effect should be taking place (Ongaro & Kaptchuk, 2019). For example, it is likely that you experience that your fever is going down after you have measured that it is going down, while it is less likely that you experience that your fever is going down when you do not measure it. From this perspective, the experience of symptom reduction is related to the expectation that your symptoms should

be reducing ("I have seen that I do not have a fever anymore, so I should feel better").

There is strong evidence that visual cues can shape expectations. People associate colours with certain effects (Jacobs & Nordan, 1996; Tao et al., 2018; Wan et al., 2015). For example, people associate red pills with stimulant effects (Jacobs & Nordan, 1996; Tao et al., 2018) and white pills with anti-headache effects (Wan et al., 2015). Also, the expectation of the strength and effectiveness of a drug depends on other characteristics like size, packaging colour and brand (for a review, see: Meisnerr & Linde, 2018 or Spence, 2021). So, by altering the colour of a certain drug, the expectations about the drug can alter as well.

As a next step, expectations can elicit placebo effects. For instance, symptom relief occurs faster when people are given overt treatment versus covert treatment (Benedetti et al., 2011) and for people with high expectations of treatment efficacy, there is no significant difference in treatment effects between an active treatment versus a placebo treatment (Sanders et al., 2020). A meta-analysis of Vase and colleagues (2002) found that the strength of placebo effects systematically varies with the type of study; the placebo effect is smaller when a double-blind design is used compared to when deceptive suggestion is used. That is, when someone is told that there is a 50% chance of receiving a placebo pill (double-blind), the expectation that a treatment will be effective is lower compared to when it is suggested that an active substance is administered (deceptive suggestion) (Geers et al., 2010). These results suggest that expectations play a pivotal role in eliciting placebo effects. I will now discuss several tools that can be used to shape expectations.

**Shaping expectations**

Through associative learning, expectations can be shaped in a similar fashion as the saliva reaction of Pavlov's dog (Colloca & Miller, 2011; Pavlov, 1927). Cognitive associations form between the environmental cues before receiving treatment (e.g., presence of a doctor or ingesting a pill) and the physical or cognitive effects of that treatment (relief of symptoms).

Environmental cues are the conditioned stimuli and treatment effects the conditioned responses. Similar as Pavlov's dog salivating after hearing a bell ring, we expect a physical or cognitive effect to take place after seeing environmental cues of a treatment. We repeatedly associate treatment environments with symptom relieve, so the prediction that treatment environments will be beneficial for your health has a high precision. In this sense, placebo effects are the physical or cognitive effects that arise from the associations of the environmental cues when no active treatment is administered.

Verbal suggestion can induce expectations as well. When you are told that something is going to happen, then it is likely that you will expect this to happen. Imagine you have never been to the aforementioned restaurant before and a friend tells you that it is the best place they have ever eaten. Now it is likely that you expect great things from this restaurant when you first go there; the verbal suggestion has shaped your expectation. Placebo effects can be elicited using verbal suggestion (Benedetti & Amanzio, 2011; Colloca et al., 2013). If you are told that a pill will have a physical or cognitive effect, then you will expect this effect to take place after ingestion. Placebo effects then arise when there is no active treatment is administered but, through verbal suggestion, people still experience physical or cognitive effects. Verbal suggestion is a particularly strong elicitor for the nocebo effect (Bartels et al., 2014; Colloca & Miller, 2011). For example, Colloca and colleagues (2008) found that tactile stimuli are experienced as painful when they are verbally suggested to be painful. So verbally induced nocebo effects can turn seemingly neutral stimuli into painful.

It is important to note that the source of verbal suggestion plays an important role in the strength of placebo effects (Baskin et al., 2003; Locher et al., 2018). When your friend is a chef and tells you that the restaurant is great you would have higher expectations than when, for example, your friend only eats at fast-food chains. A recommendation from a chef will be stronger and more trustworthy. In terms of predictive processing, it will be of greater magnitude

and precision. Similarly, placebo effects will be stronger when the administer of the treatment is a credible source of information (e.g., a doctor).

Another tool to shape expectations we have already come across in the introduction is observational learning. We speak of observational learning when people shape their expectations by observing others (Bajcar & Bąbel, 2018). Recall the example of the basketball player from the introduction. There, your expectations might be shaped by observing the effects for that basketball player. Placebo effects are induced when your expectations are shaped by observing others and these expectations are enough to elicit physical or cognitive effects.

Although associative learning, verbal suggestion, and observational learning can be used individually to elicit placebo effects, a combination of these factors can elicit stronger placebo effects (Bartels et al., 2014). For example, a placebo/nocebo study investigating itchy sensations showed that using both verbal suggestion and associative learning elicited stronger placebo effects compared to using each tool individually (Bartels et al., 2014). In addition, the combination of verbal suggestion and conditioning also elicits more robust placebo effects (Benedetti & Amanzio, 2011).

Placebo effects are not exclusively found in the medical domain, but also in other domains (e.g., Bérdi et al., 2011). Historically there has been a strong focus on pain and motoric function in placebo studies (Price et al., 2008; Turi et al., 2018), but recently some studies have stepped outside the realm of pain research and investigated placebo effects on cognitive enhancement.

**Placebo Research on Cognitive Enhancement**

Several studies have shown that placebo stimulation can lead to cognitive enhancement. Cognitive enhancement occurs when a treatment leads to an improvement in cognitive processes (Dubljević et al., 2015). For example, Schwarz & Büchel (2015) showed enhanced subjective, but not objective, performance in a cognitive task when people received placebo

brain stimulation. Also, Magalhães de Saldanha de Gamma and colleagues (2013) showed less conflict-interference during a Stroop task after receiving placebo brain stimulation.

Using electroencephalography (EEG) we can measure electrical potential sourcing from the brain with high temporal resolution. By analysing event-related potentials (ERP), researchers have linked cognitive processes to distinct peaks in the ERP-waveform (for an overview, see: Sur & Sinha, 2009). Using EEG, we can determine whether placebo effects can not only be detected on a self-report level but also on an objective neurological level. Roseman and colleagues (2011) have argued that self-reports are often prone to bias and most studies mentioned in this paper have used these self-report measures. The addition of implicit physiological measures of the placebo effect gives us insight in how the reported placebo effects relate to cognitive processes.

Error-processing is well suited to be studied using EEG: it is characterized by a distinct peak in the ERP waveform and this peak occurs relatively early in the ERP, which means it is less likely to be affected by other cognitive processes. Error-processing is reflected by the error-related negativity (ERN) which occurs whenever a person makes a mistake or responds when not responding was appropriate (Crowley, 2013; Gehring et al., 1993; Gehring et al., 2018). The ERN is a large negative potential in the ERP originating from the anterior cingulate cortex and occurs between 80 and 150 milliseconds after an error response (Ito et al., 2003; Stevens et al., 2011). Inzlicht & Al-Khindi (2012) showed that people show a decreased sense of agency when receiving placebo cognitive enhancement and that this decreased sense of agency was related to a lower ERN amplitude. People attributed their errors to the sham stimulation instead of their own cognitive capacities and therefor their error-processing was less pronounced (these findings were not replicated in a study by Rodilla et al., 2016).

In addition, two studies have used the Eriksen flanker task (Eriksen & Eriksen, 1974) to investigate if placebo brain stimulation leads to altered error-processing (van Elk et al., 2020;

Hoogeveen et al., 2018). In both studies participants received placebo or nocebo brain stimulation and were told that this stimulation would lead to a better or worse performance in a cognitive task, respectively. Next to the ERN, subjective performance ("How well did you perform?") and sense of agency over errors ("Did the brain stimulation cause you to make errors?") was measured using self-reports.

In both studies strong placebo effects were found on the self-report measures. First, participants indicated higher subjective performance in the placebo condition compared to the other conditions. These results indicate that people's subjective performance was influenced by the verbal suggestion of brain stimulation. Second, participants reported a lower sense of agency over errors in the placebo and nocebo condition compared to the control condition, indicating that errors were misattributed to the brain stimulation (van Elk, Groenendijk, & Hoogeveen, 2020; Hoogeveen, Schjoedt, & van Elk, 2018). However, mixed results were found regarding the ERN amplitude.

Hoogeveen and colleagues (2018) found an increased ERN amplitude in the placebo condition compared to the control condition, which suggest that verbal suggestion of stimulation is enough to elicit neurological changes. The increased ERN amplitude in the placebo condition can be interpreted as increased error detection. When people believe that their cognition is enhanced, their prediction will be that they will less mistakes. Therefor a mistake will lead to a large prediction error reflected in increased ERN amplitude. In contrast, van Elk, Groenendijk, & Hoogeveen (2020) found no differences in ERN amplitude across conditions. The authors indicate that this might have to do with the strength of the manipulation. To effectively test the placebo effects of brain stimulation, participants need to be convinced of the efficacy of brain stimulation devices. In their article, van Elk, Groenendijk, & Hoogeveen (2020) explain that the verbal suggestions about the effects of the stimulation might not have been strong enough to induce strong expectations of the stimulation. As stated before, placebo

effects are more robustly found when a combination of verbal suggestion and conditioning is used which might explain these mixed results (Bartels et al., 2014).

In addition, the stability of the ERN depends on the number of observed error-trials (Clayson, 2020). In the Hoogeveen, Schjoedt, & van Elk (2018) study participants made more than 20 errors throughout conditions, but the participants in the van Elk, Groenendijk, & Hoogeveen (2020) study made less than 10 errors per condition. Although a minimum of 6 to 8 error-trials is often the cut-off for inclusion in ERN studies, a recent meta-analysis (Clayson, 2020) suggested that a higher number of error trials (+/-16) leads to higher internal consistency. Following this logic, the stability of individual ERNs might have been lower in the 2020 study, however this effect might have been balanced out by the larger sample size (57 vs 23).

To summarize, there have been mixed results in placebo research involving the ERN. Both studies tried to elicit placebo effects based on the verbal suggestion framework alone and this might explain these mixed results. Also, more error trials are needed to measure stable ERNs. In this study, we used a combination of verbal suggestion and associative learning to investigate whether we can find placebo effects in the ERN. The study will provide insights whether we can use implicit physiological measures to investigate the placebo effect in addition to self-reports.

**Current Study**

We tried to answer the research question if placebo and nocebo effects on a subjective and a neurophysiological level can be elicited by a combination of verbal suggestion and associative learning. We let participants perform a cognitive task after three conditions: placebo brain stimulation, nocebo brain stimulation, and a control condition. We told our participants (verbal suggestion) which effects the brain stimulation would have: enhancing or impairing their cognitive performance. Then, we used a dynamic flanker task (Eriksen & Eriksen, 1974) to shape participants' expectations through associative learning. By adding a phase in which

the task was made easier or harder depending on the condition, we induced the association between task difficulty and condition. We were interested if this combination of verbal suggestion and associative learning led to placebo effects measured using EEG recordings and self-report scales. We have built on earlier studies (Hoogeveen et al., 2018; van Elk et al., 2020) by using a combination of verbal suggestion and associative learning, rather than solely verbal suggestion.

The study design and analysis plan have been pre-registered on the open science framework prior to conducting the study (https://osf.io/j7z45). We expected H1) increased subjective performance in the placebo condition and decreased subjective performance in the nocebo condition compared to control, H2) decreased sense of agency in both the placebo and nocebo condition compared to control, and H3) increased ERN amplitude in the placebo condition compared to the other conditions. Exploratively, we were also interested in other ERP peaks to find potential placebo effects on different cognitive domains. We looked at the N2c, which is believed to reflect conflict processing (West et al., 2005), and the feedback-related negativity (FRN). The N2c peak occurs between 250 and 350ms after a response and increases in amplitude when conflict is higher (Kopp et al., 1996; West et al., 2005). The FRN peak occurs approximately 250ms after feedback onset and increases in amplitude when performance is worse than expected (Crowley, 2013) We didn't specify any prior expectations about condition effects on N2c or FRN peaks.

## Methods

### *Pre-registration*

The study was pre-registered on the open science framework prior to conducting the study (https://osf.io/j7z45). However, we deviated from our pre-registration due to some unforeseen circumstances. First, we noticed a programming error after the 11$^{th}$ participant (we didn't specify an upper bound for the response window and stimulus presentation; see

Procedure). We updated the experimental paradigm and checked for differences in outcome measures due to this update (See appendix A). Because of this we excluded the first 11 participants from further analysis. The final sample size thus deviates from our original pre-registered sample size.

Next, we inspected our data after the 30$^{th}$ participant (as pre-registered: https://osf.io/wqtc6) to see how our experimental procedure affected task performance during the conditioning phase (See Appendix B). The experimental paradigm did not result in a conditioning phase as intended, so we decided to terminate data collection.

Finally, we deviated from our original analysis plan. After inspection of the intra-class correlation for each outcome measure, we came to the realization that it would be more appropriate to fit multi-level models than repeated measures ANOVAS to respect the structure of the data. When the intraclass-correlation coefficient is high, meaning that there is a relatively high between-subject variance and relatively low within-subject variance, repeated measure ANOVAs will lead to too small standard errors and, as a result, inflated *p*-values (Roback & Legler, 2021). In addition, we decided to only use the last 6 blocks of the experimental phase for our analysis, since only in the last 6 blocks the performance data was comparable between blocks.

***Participants***

30 participants completed the experimental procedure (18 female, mean age = 37.6 years, SD = 16.5, range = 19-63 years). The final sample consisted of 19 participants (13 female, mean age = 34 years, SD = 12.4, range = 19-63 years). Participants were recruited online and via advertisements in a local newspaper. Initially, we excluded students since they might be too familiar with placebo research. However, the COVID-19 pandemic made recruiting troublesome and we dropped this rule during data collection. For our final sample we included participants that were: 18 years or older, were non-psychology students, had no history of

psychological disorders, had no history of neurological trauma, were no drug or alcohol abusers, and had not participated in placebo research before. Participants were pre-screened on these criteria before study admission.

***Procedure***

EEG was recorded after questionnaire administration. Participants were placed 60 cm in front of a computer screen (1920x1080 resolution) and responded using response boxes that were attached to the armrests of a chair. The experiment was programmed using Presentation® software (version 22.1; Neurobehavioral Systems, Inc, Berkeley, CA). See Figure 1 for a graphical overview of the experimental procedure.

After EEG set-up, we told participants that we would run a low electrical current through their brain via two electrodes to target specific brain regions to either enhance or impair their cognitive performance. However, at no point during the entire experiment we used electrical stimulation. Verbal suggestion was provided before each of the three experimental blocks. We told participants that the stimulation could influence their task performance during the upcoming block. Specifically, we told them that the stimulation could lead to less errors in the placebo condition and more errors in the nocebo condition. A third control condition was used in which participants were told that their brain would not be stimulated.

Next, instructions were shown on the screen about the stimulation effects, repeating our verbal suggestion. Then, a two-minute timer was shown during which the alleged stimulation took place was presented on the screen. Next, the adjusted Flanker task (Eriksen & Eriksen, 1974) started. In each trial a fixation cross was presented for 450-ms after which distractor arrows were presented for 200-ms, followed by the target arrow. Participants indicated the direction of the target arrow while ignoring the distractor arrows. After the response, a blank screen was presented for approximately 1350-ms. The task used four different stimuli that were either congruent (“<<<<<”, “>>>>>”) or incongruent (“<<><<”, “>><>>”). Participants

received 500-ms feedback after their response (“Correct!”, “Incorrect!”, “Too late!”).

We adjusted several parameters based on task performance to force error responses. First, the response window was adjusted with 50-ms steps with a minimum and maximum response window of 100-ms and 1000-ms. Second, the stimulus presentation time was adjusted with 5-ms steps a minimum stimulus presentation of 100-ms. Third, the brightness of the target stimulus was adjusted with the RGB-color code ranging from (188,188,188) to (255,255,255).

Participants completed 13 blocks (20 trials per block). The first four blocks served as a conditioning phase. In this conditioning phase we adjusted the parameters based on a condition-specific accuracy range. When the accuracy exceeded the boundaries of the range, the task parameters were adjusted. The accuracy ranges for the conditioning blocks were as follows: enhancement range: 70-90%, impairment range: 30-50%, control range: 60-80%. Participants had a short break after the conditioning phase. During the last nine blocks the accuracy range was set to 60-80% throughout conditions.

After each block participants were asked to answer the question: “Were your mistakes influenced by the brain stimulation?” on a 7-point scale, ranging from 1 (“not at all”) to 7 (“completely”). For this scale, higher scores indicated less sense of agency over errors. In addition, participants were asked to answer the question: “How well did you perform during the last block?” on a 5-point scale, ranging from 1 (“Very bad”) to 5 (“Very good”). These questions were also asked during the conditioning phase so participants could not notice procedural differences between the first 4 and last 9 blocks. The order of the experimental procedure was counterbalanced across participants and participants practiced with the task at the start of the experiment. After the experiment, we used a post-test questionnaire to assess if participants believed the brain stimulation was real and to assess the participants’ experience during the experiment.

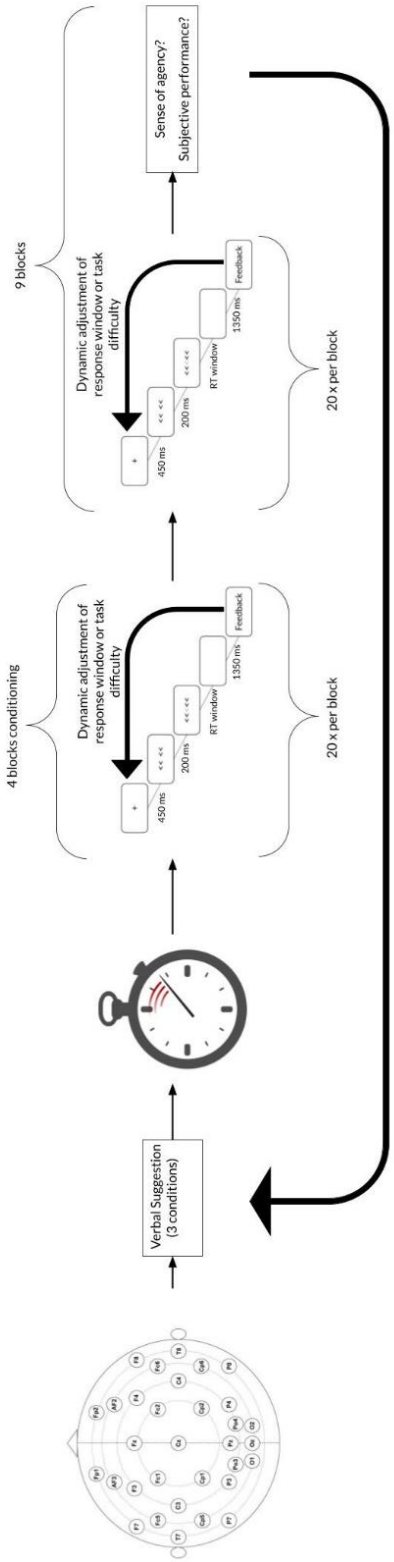


*Figure 1***.** Experimental procedure. After EEG set-up was complete received verbal suggestion dependent on the current condition. Then a 2-minute timer was presented on the screen, after which the conditioning phase of the flanker task started. After the conditioning phase, participants performed in an additional 9 blocks. After each block of 20 trials, sense of agency and subjective performance was assessed.

### *Measurements*

We used three main outcome measures: 1) subjective performance, 2) sense of agency, and 3) ERN. In addition, we used the Tellegen absorption scale (Tellegen & Atkinson, 1974) to assess trait absorption. This can be considered as a measure of susceptibility for losing yourself in absorbing events. Examples from this questionnaire are: "When I listen to music, I can get so caught up that I do not notice anything else" and "Sometimes I experience things as if they are doubly real". Participants must indicate how strongly they agree with the items on a 5-point Likert scale. Also, we used the Neuromyths questionnaire (van Elk, 2019) to assess susceptibility for neuromythcial statements ("to uncover people's true thought, brain scans are more suited than questionnaires") and how strongly people believe in cognitive enhancement techniques ("brain stimulation can be used to enhance cognitive performance). For all statements participants are asked to indicate how strongly they agree on a scale from 0 to 100. In addition, we also asked for age, gender, nationality, and occupation.

After the experimental procedure, participants indicated if they experienced any side effects of the brain stimulation (e.g., headache or nausea), to what extent they believed they received brain stimulation, and to what extent they believe brain stimulation is an effective way to enhance cognitive performance.

### *EEG Measurement*

During the experiment, we measured EEG at a 2048 Hz sampling rate, 100 Hz online low-pass filter, and .16 Hz online high-pass filter using BioSemi Actiview (BioSemi, 2020). 32 electrodes (See Appendix B) were connected to an EEG cap. VEOG and HEOG were measured by placing electrodes at the subjects' outer canthi and above and below the right eye, respectively. Two electrodes were placed at the mastoids that served as reference electrodes. EEG data was analyzed using Brain Vision Analyzer (Brain Vision Analyzer, Version 2.2.0, Brain Products GmbH, Gilching, Germany). First, all channels were re-referenced to the

mastoid electrodes. Then, a high-pass filter of 0.1 Hz and low-pass filter of 100 Hz were applied to the channels with a notch filter of 50 Hz. Data was then segmented into epochs that were time-locked to responses (-100-ms to 1000-ms from onset) for each condition starting from the 8$^{th}$ (out of 13) block, so the responses from the last 6 blocks were analyzed. An automatic ocular correction was applied using the Gratton & Coles algorithm (Gratton et al., 1983), after which epochs containing artifacts were automatically rejected based on the following criteria: 100 µV gradient change within epoch, a difference between minimum and maximum amplitude of 200 µV, a minimum and maximum amplitude of +/- 150 µV, and 0.5 µV during an interval of 100-ms.

Then, ERPs were averaged and baseline corrected (-200ms to 0ms) per participant per condition. To detect the ERN, we used automatic peak detection for the global maximum during a period of 50-ms to 150-ms after the response. To detect the N2c, we used automatic peak detection for the global maximum during the period of 250-ms to 350-ms after stimuli onset.

We used the same procedure to detect the feedback related negativity (FRN), except ERPs were time-locked to feedback onset. Since we did not have a marker for feedback onset during the experiment, feedback onset was estimated using the experimental script. We shifted the codes that indicated the response type (correct, incorrect, and miss) with 1300-ms. Then we segmented the data per condition time-locked relative to these shifted markers and applied the artifact rejection and baseline correction. Finally, we used automatic peak detection for the global maximum during the period of 225-ms to 275-ms after feedback onset to detect the FRN.

***Data analysis***

For testing the effect our experimental manipulations on subjective performance, sense of agency, ERN amplitude, N2c amplitude, and FRN amplitude we fitted multilevel models with random effects being dependent on participant ID and experimental condition as main predictor. In addition, we controlled for absorption ratings and susceptibility for neuromyths.

Inferences in multilevel models can be based on whether the 95%-confidence interval of the point estimate for the coefficient contain zero (Roback, 2021) or can be based on a $\chi^2$-test of a bootstrapped distribution when the assumptions of the analysis are not met. We provide both options in the results section. Note that we reverse coded the sense of agency scale, so that lower scores on this scale reflect a lower sense of agency over the errors made during the last block.

To test if we found a significant ERN- and N2c-effect, we used paired sample t-tests. For the ERN, we tested the difference between ERP amplitudes of correct and incorrect trials. For the N2c, we tested the difference between ERP amplitudes of congruent and incongruent trials.

**Results**

The results from our post-test questionnaire are depicted in Table 1. Each item was assessed on a 5-point Likert scale. The items regarding side effects were answered on a scale with an answer of 1 indicating "not applicable at all" and 5 indicating "completely applicable". The items regarding the effects of neurostimulation were answered on a scale with an answer of 1 indicating "completely not" and 5 "completely". So, on all questions, a higher score indicates a stronger experience of the proposed item. There was a low amount of reported side effects. On average, participants were not strongly convinced that we used brain stimulation ($M$ = 3.08, $SD$ = 1.26). Also, participants were not strongly convinced about the effectiveness of brain stimulation ($M$ = 3.16, $SD$ = 1.28).

*Table 1*. Results of post-test questionnaire

| Item | Mean | SD |
|---|---|---|
| Headache | 1.84 | 1.28 |
| Neck pain | 1.32 | .75 |
| Feeling of nausea | 1.16 | .47 |
| Tension in neck | 1.52 | .96 |
| Tingling sensation at electrodes | 1.56 | 1.04 |
| Burning sensation at electrodes | 1.00 | .00 |
| Uncomfortable feeling | 1.68 | .80 |
| Experienced power of stimulation | 2.72 | 1.37 |

| | | |
|---|---|---|
| Nervousness before stimulation | 1.64 | 1.15 |
| Effectiveness of stimulation | 3.16 | 1.28 |
| Stimulation was used | 3.08 | 1.26 |

*Note*. The first seven items assessed whether participants experienced any side-effects from the sham stimulation and the final four items assessed the participant's experience of brain stimulation

***Prespecified results***

Based on exploration of the performance data (Appendix B), we decided to only include the last 6 blocks for our data analysis. Performance data across participants of these blocks are depicted in Table 2. We found no differences in accuracy across conditions, $F$ (1,55) =.192, $p$=.663, or reaction times across conditions, $F$ (1,55) < 0.01, $p$=.987, indicating that performance during the last 6 blocks was comparable across conditions. The results for the subjective scales are depicted in Table 2 and Figure 2. The control condition served as reference condition in all analyses.

We fitted a random-intercepts model with subjective performance (H1) rating as dependent variable and only a random effect of participant ID, which showed an intraclass correlation of .311 ($n$ = 19). This indicates that 31.1% of the variance in subjective performance ratings can be explained by differences between participants. We then added condition to model predicting subjective performance, allowing for random effects between participants. Contradicting our hypothesis (H1), we found higher subjective performance ratings in the nocebo condition ($M$ = 3.53) compared to control ($M$ = 3.05), $b$=.47, $t$ = 3.91, 95% CI [.24, .71], but not for the placebo condition ($M$ = 3.20) compared to control, $b$ = .15, $t$ = 1.23, 95% CI [-.09, .39]. The higher accuracy in the nocebo condition might explain these found differences. However, when added to the model, the nocebo condition effect remained significant, $b$=.28, $t$ = 2.68, 95% CI [.07, .48], indicating that participants on average reported a .28-point higher subjective rating in the nocebo condition compared to control while controlling for absorption ratings, susceptibility for neuromyths, and accuracy. Note that the effect of accuracy is also significant, $b$=2.34, $t$ = 11.64, 95% CI [1.95, 2.73], indicating that on

average the subjective rating goes up by .234-points when the accuracy increases with 10%, keeping all other variables constant. We found no significant main-effect of absorption ratings, $b$ = -.001, $t$ = -.32, 95% CI [-.01, .01], but we found a significant main-effect of susceptibility for neuromyths, $b$ = .02, $t$ = 2.20, 95% CI [.00, .04].

*Table 2*. Performance data and subjective ratings

| | Placebo | Nocebo | Control |
|---|---|---|---|
| Reaction Time (ms) | 327 | 353 | 326 |
| Accuracy | 70.9% | 80.9% | 72.4% |
| Error Responses | 1.5% | 7.2% | 9.5% |
| Misses | 11.7% | 7.4% | 13.5% |
| Subjective Performance | 3.20 | 3.53 | 3.05 |
| Sense of Agency | 3.61 | 3.15 | 2.94 |

*Note.* This table depicts performance data and subjective ratings across all subjects for each condition.

The assumption of linearity is difficult to check since our main predictor is categorical.

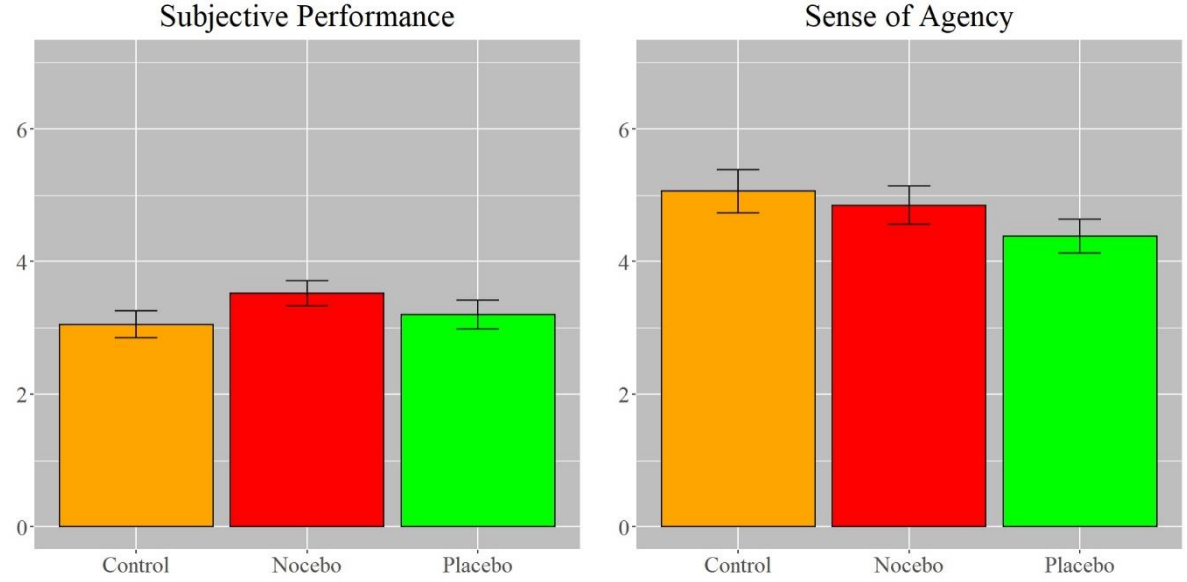


*Figure 2*. Bar charts of sense of agency ratings and subjective performance ratings.

Therefor, a parametric bootstrap can be used to confirm our results. A 1000-sample bootstrap showed that a model including accuracy significantly improves the model fit over a more parsimonious model, $\chi^2(1) = 114.5$, $p < 0.001$, and fits significantly better than a model with no predictors, $\chi^2(5) = 135{,}32$, $p < 0.001$. Against our expectations (H1), after nocebo stimulation

subjective performance was higher than control and there was no difference in subjective performance after placebo stimulation compared to control.

We then fitted a random intercepts model with sense of agency (H2) as dependent variable and ID as sole predictor ($n$ = 19). This model indicates that 40.6% of the variance in sense of agency ratings can be explained by differences between persons. We then fitted a model with sense of agency as dependent variable and condition as independent variable whilst allowing for random effects between participants. Confirming our expectations (H2), we found a significantly lower sense of agency in the placebo condition ($M$ = 3.61) compared to the control condition ($M$ = 2.94), $b$ = -.68, $t$ = -3.89, 95% CI [-1.02, -.34]. However, we found no significant differences between the nocebo condition ($M$ = 3.15) and the control condition, $b$ = -.21, $t$ = -1.21, 95% CI [-.55, .13]. The effects remain when controlled for absorption ratings and susceptibility for neuromyths: $b$ = -.68, $t$ = -3.89, 95% CI [-1.02, -.34] for placebo vs control and $b$ = -.21, $t$ = -1.21, 95% CI [-.55, .13] for nocebo vs control. Placebo brain stimulation led to lower sense of agency over errors in our sample, which is in line with our hypothesis. However, there are no differences in sense of agency between the control condition and the nocebo condition, where we expected a lower sense of agency for the nocebo condition compared to control. We found no significant main-effect for absorption ratings, $b$ = .00, $t$ = .38, 95% CI [-.02, .03], but we did find a significant main-effect of susceptibility for neuromyths, $b$ = -.04, t = -1.94, 95% CI [-.02, .03] A parametric bootstrap confirmed these results, $\chi^2(4) = 15.57$, $p < 0.001$.

*ERN Analysis*

After data-processing, we had 15 participants for which we could compute an ERN for at least one condition. A paired-sample t test showed a significant main effect of response type on response-locked negativity, $t$ (14) =2.71, $p$ = 0.017, 95% CI [.64, 5.49], illustrated in a stronger negativity after an error-response ($M$ = -6.05 μV, $SD$ = 4.09) compared to a correct

response (*M* = -2.99 µV, *SD* = 3.02). For a paired-sample t-test we assume independent observations and normally distributed measurement. Although we can assume independent observations in this sample, it is hard to determine normality of observations in such a small sample.

53.6% of the variation in ERN was explained by variation between participants. We fitted a model with condition as a predictor of ERN amplitude, allowing for random effects between participants. Contradicting our hypothesis (H3), we found no significant difference in ERN amplitude for the placebo condition (*M* = -8.16 µV, *SD* = 6.00) compared to the control condition (*M* = -7.66 µV, *SD* = 5.87), *b* = -.26, *t* = -.17, 95% CI [-3.26, 2.68], or for the nocebo condition (*M* = -8.44 µV, *SD* = 4.52) compared to the control condition , *b* = -1.56, *t* = -.83, 95% CI [-5.19, 2.17] (Figure 3). The findings were confirmed by a parametric bootstrap: $\chi^2(2) = .77$, $p = 0.719$.

These results suggest that there is no effect of placebo or nocebo stimulation on the ERN

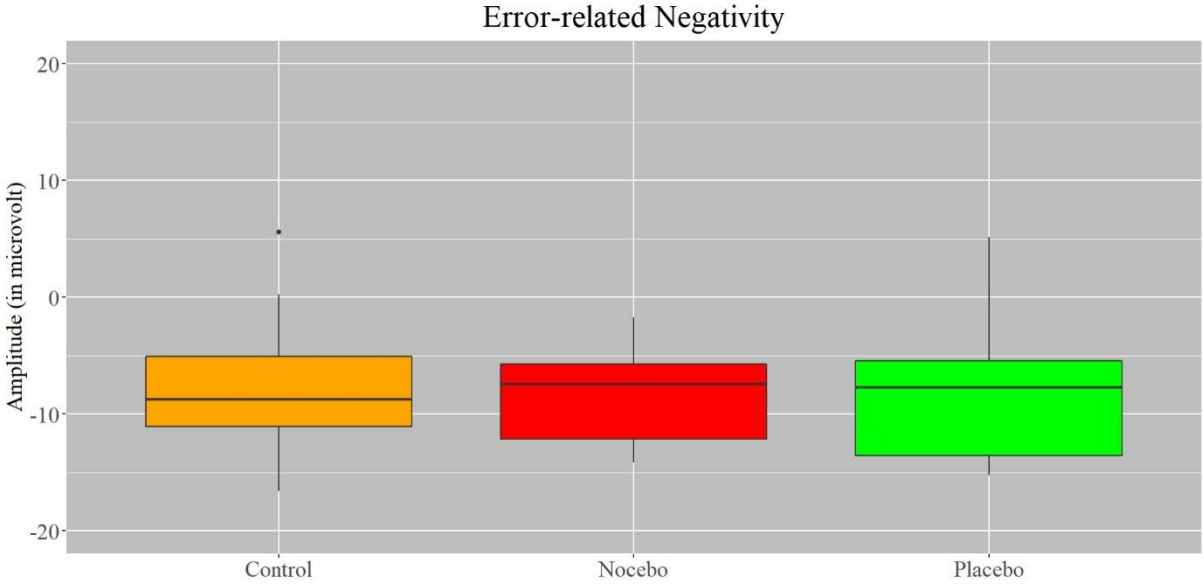


*Figure 3.* Box plots of ERN amplitude (in µV). Dots indicate outliers and the arms outside the boxes represent the outer 50% of the observations. Thick black lines within the boxes represent the mean ERN amplitude per condition.

amplitude. Our regression estimates lack precision, reflected by the large margin of errors of the confidence intervals. This is most likely due to our small sample size (*n* = 15).

***Exploratory Results***

A paired-sample t-test ($n$ = 15) showed a significant difference in negativity time-locked to stimulus onset between congruent ($M$ = 0.49 µV, $SD$ = 4.59) and incongruent ($M$ = -1.98 µV, $SD$ = 5.49) stimuli, $t$ (14) = 3.00, $p$ = .009, 95% CI [.71, 4.23], indicating increased conflict processing for incongruent stimuli. Figure 4 (left panel) displays boxplots of N2c amplitude for incongruent stimuli for each condition. The absolute N2c amplitude is somewhat larger in the placebo condition ($M$ = -3.11 µV, $SD$ = 5.93) compared to the control condition ($M$ = -1.83 µV, $SD$ = 6.09), but not significantly so: $b$ = -1.28, $t$ = -1.47, 95% CI [-2.98, .42]. Also, the nocebo condition ($M$ = -1.08 µV, $SD$ = 5.59) resulted in a somewhat smaller absolute N2c amplitude to control, but not significantly so: $b$ = .89, $t$ = .99, 95% CI [-.86, 2.63]. These results were confirmed by a parametric bootstrap: $\chi^2(2) = 5.80$, $p = 0.061$.

Figure 4 (right panel) shows boxplots for FRN amplitude after missed responses ($n$ =15). We found a significantly smaller FRN amplitude in the nocebo condition ($M$ = 1.72 µV, $SD$ = 7.56) compared to control ($M$ = -4.14 µV, $SD$ = 4.81), $b$ = 5.71, $t$ = 2.20, 95% CI [.64, 10.90], and in the nocebo condition compared to the placebo condition. ($M$ = -5.93 µV, $SD$ = 7.76), $b$ = 9.37, $t$ = 2.50, 95% CI [2.07, 16.38]. There is no significant difference in FRN amplitude between the placebo condition and the control condition, $b$ = -1.41, $t$ = -.64, 95% CI [-6.09, 2.87]. A parametric bootstrap confirmed these results: $\chi^2(2) = 7.76$, $p = 0.026$. These results suggest altered feedback processing in the nocebo condition compared to control, but again our regression estimates lack precision due to the small sample size. Appendix C contains correlations between all outcome measures across conditions

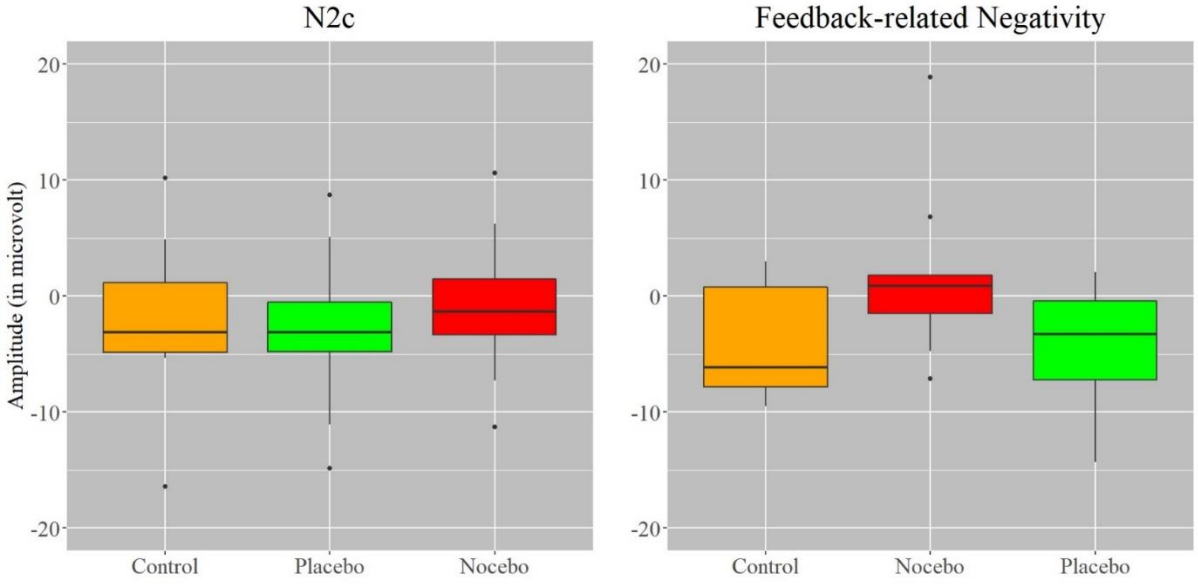


Figure 4. Boxplots per condition of N2c amplitude (in µV) on the left panel and FRN amplitude (in µV) on the right panel. Dots indicate outliers and the arms outside the boxes represent the outer 50% of the observations. Thick black lines within the boxes represent the mean ERN amplitude per condition.

**Discussion**

We investigated the effect of nocebo and placebo brain stimulation on subjective performance, sense of agency, and error processing by making use of verbal suggestion and associative learning. Exploratively, we also investigated the effects on conflict processing and feedback processing.

Against our hypothesis (H1), we found enhanced subjective performance in the nocebo condition compared to the control condition. In addition, we did not find enhanced subjective performance in the placebo condition compared to the control condition. These results contradict existing literature in which placebo and nocebo effects of sham brain stimulation (Hoogeveen et al., 2018; van Elk, 2020) and other placebo stimulations (Schwarz & Büchel, 2015; Winkler & Herman, 2019) on reported performance were repeatedly found. The enhanced subjective performance after nocebo stimulation can be explained by the analysis of the conditioning phase (Appendix B). During the conditioning phase, there was a significant difference in difficulty between the nocebo condition and the other two conditions. As a result, there was a large difference in performance between the conditioning phase and the last nine

blocks in the nocebo condition. Although we intended to make the conditioning phase of the nocebo condition difficult, the difference in difficulty might have been too large which may have led to a contrast effect between the conditioning blocks and the last nine blocks. Because this contrast was so large, participants might have felt like they performed far better during the last nine blocks of the nocebo condition compared to the conditioning phase. The fact that we found no significant objective performance difference during the conditioning phase between the placebo condition and control condition explains why we did not find placebo effects on subjective performance: after the placebo brain stimulation there was no difference in performance compared to control, so participants were not conditioned that placebo brain stimulation would lead to better performance.

Partly confirming our hypothesis (H2), sense of agency was significantly lower in the placebo condition compared to control, but not in the nocebo condition compared to control. The former result shows the tendency of people to attribute their errors to external factors when they get the chance and is in line with previous research (Hoogeveen et al., 2018; van Elk et al., 2020). Why we did not find an effect of nocebo stimulation on sense of agency again might be due to the conditioning phase. After the nocebo stimulation people perceived their performance during the last nine blocks as better than expected based on the suggestion of cognitive impairment. This might have led to no perceived effects of the nocebo stimulation and, if people do not believe that the stimulation is working, then errors should be due to their own performance instead of the nocebo stimulation. This explanation is supported by our post-test questionnaire (see Results), which showed that our participants were not strongly convinced that we stimulated their brains. Moreover, it might have been confusing for participants to indicate the influence of brain stimulation in the control condition since we instructed that no stimulation would be applied. This makes interpretation of sense of agency for the control condition difficult.

Also against expectations (H3), we found no differences in ERN amplitude between conditions. Participants did not process errors differently throughout conditions, which suggest that the ERN might not be a suitable objective measure of placebo/nocebo effects. However, these results have to be interpreted with caution since the experimental paradigm did not lead to a sufficient amount of error trials within each condition to compute reliable ERNs (Clayson, 2020). For example, only 1.5% of the trials resulted in an error in the placebo condition. The reason for this low number of error trials seems to be the nature of the used experimental paradigm. Most of the non-correct trials are missed trials due to the shortened response window and errors rarely occurred. Solely based on this study, the ERN should not be ruled out as a potential objective measure of placebo effects.

Our exploratory analysis showed no altered conflict processing reflected by peak N2c-amplitude throughout conditions. The study was not primarily designed to investigate conflict processing and might not have been optimal to investigate this. We only had one level of conflict (congruent vs incongruent) and to thoroughly investigate conflict processing it would have been fruitful to have differing levels of conflict (e.g., "> > > > >" vs. "< < > < <" vs. "> < > < >"). We found a significantly smaller FRN amplitude after missed trials in the nocebo condition compared to the control condition. The FRN peaks when an individual receives external feedback that performance is worse than expected (Crowley, 2013). This indicates that in the nocebo condition negative performance feedback is not perceived worse than expected, which is in line with the predictive processing framework. When you expected to perform badly and this is confirmed by feedback, this will result in a smaller prediction error compared to when you expected to perform better. In our study, the FRN is a more reliable measure than the ERN since a higher number of trials have been used to compute the FRN.

The effect on FRN amplitude seems to contradict with the result from our post-test questionnaire: people were not convinced about the effectiveness of brain stimulation, so why

would they expect negative feedback? One explanation might be the difference in explicit and implicit measures. By explicitly by asking participants about their beliefs about brain stimulation after the experiment, we might have induced demand effects. If you ask participants if they believe that we used brain stimulation, they might adjust their answers to what they think the study is about (Mummolo & Peterson, 2017) and this might not reflect their true beliefs. On the other hand, the FRN is an implicit measure which might better reflect the true beliefs of the participants.

Some serious limitations of this study need to be discussed. First, our conditioning did not have its intended effect (Appendix B). We found no statistical difference in performance between the placebo and control condition. Also, the difference in difficulty between the conditioning phase and last nine blocks in the nocebo block was likely too large which might have led to a contrast effect. Moreover, throughout conditions most non-correct trials were missed trials and it is debatable if missed trials reflect task difficulty. Since these flaws in the conditioning phase, it can be argued that we cannot test the effect of associative learning in this study. Second, the results of the post-test questionnaire suggest that participants were not convinced that we used brain stimulation. Thus, our verbal suggestion might have been too weak in order for us to elicit placebo/nocebo effects. This might also explain the mixed results for subjective performance and sense of agency. Finally, we used a within-subject design because it requires less participants. This is beneficial for EEG research since the time it takes to test a single participant: it would not have been feasible to conduct this study with a between-subject design. However, various other studies use a between-subject design to investigate placebo/nocebo effects (Lee & Suhr, 2019; Kuenzel et al., 2011; Winkler & Hermann, 2019). As mentioned earlier Baker and colleagues (2017) proposed that discrepancies of results in studies investigating the effect of including neuroscientific explanations have to do with the study design. When consecutive conditions are presented, this might induce contrast-effects

between those conditions that would not be present in a between-subject design. The same might hold for placebo/nocebo effect. When people are receiving placebo stimulation in comparison to nocebo stimulation, they might evaluate the effectiveness of this stimulation in comparison to the other condition. In addition, within-subjects designs might lead to participants guessing the aim of the study which is a smaller problem in between-subject designs (Charness et al., 2012). But as noted before, a between-subject design would not have been feasible for this study.

Future studies should focus on the convincingness of the brain stimulation (e.g., using sham-tDCS) to strengthen the verbal suggestion. For example, the ramp-up phase of a tDCS device can be used as sham stimulation to induce a tingling sensation at the scalp which might strengthen the verbal suggestion. Also, future research should continue to focus on implicit measures like ERN, FRN, and N2c amplitude. The analysis of the FRN amplitude showed a discrepancy between implicit and explicit measures. It would be fruitful to continue investigating this discrepancy using EEG as an implicit measure. Based on the mixed results between our study and previous studies (Hoogeveen et al., 2018; van Elk et al., 2020) I would suggest going back to the effect of verbal suggestion in order to inform us about the replicability of the previously found effects (Hoogeveen et al., 2018; van Elk et al., 2020). Thus, future studies should start with only using verbal suggestion before including associative learning to elicit placebo effects. Not only would this inform us about the replicability of the previously found effects, but the current study has also highlighted the complications of including a conditioning phase in the experimental design.

To conclude, the mixed results of this study should be interpreted with care. We found enhanced subjective performance in the nocebo condition, decreased sense of agency in the placebo condition, and decreased FRN amplitude in the nocebo condition. However, the effects on subjective performance are not in line with earlies findings (Hoogeveen et al., 2018; van Elk

et al., 2020): we found enhanced subjective performance after nocebo stimulation, no enhanced subjective performance after placebo stimulation, and no decreased sense of agency after nocebo stimulation. This might have been caused by our experimental design and our small sample size. Despite this, we found the FRN as a potential implicit measure that reflects people's expectations. Future research should further elucidate the power of the mind over objective and implicit outcome measures and how the mind can be influenced by altering expectations. This brings us closer to explaining why fans of James Michael McAdoo expect that using headphones that pulse electrical signals through your brain will lead to better basketball performance.

**References**


Ali, S.S., Lifshitz, M., & Raz, A. (2014). Empirical neuroenchantment: from reading minds to thinking critically. *Frontiers in Human Neuroscience, 8 (375).* doi: 10.3389/fnhum.2014.00357

Balint, J. (1955). *The Doctor, His Patient and The Illness* (2nd ed.). Elsevier Health Sciences.

Bartels, D.J.P., van Laarhoven, A.I.M., Haverkamp, E.A., Wilder-Smith, O.H., Donders, A.R.T., van Middendorp, H., van de Kerkhof, P.C.M., & Evers, A.W.M. (2014). Role of conditioning and verbal suggestion in placebo and nocebo effects on itch. *PLoS One, 9*(3). doi: 10.1371/journal.pone.0091727

Baker, D.A., Ware, J.M., Schweitzer, N.J., & Risko, E.F. (2017). Making sense of research on the neuroimage bias. *Public Understanding of Science, 26* (2), 251-258. doi: 10.1177/0963662515604975

Baskin, T.W., Tierney, S.C., Minami, T., & Wampold, B.E. (2003). Establishing specificity in psychotherapy: a meta-analysis of structural equivalence of placebo controls. *Journal of Consulting and Clinical Psychology, 71*, 973-979. doi: 10.1037/0022-006X.71.6.973

Benedetti, F. (2002). How the doctor's words affect the patients' brain. *Evaluation & the Health Professions, 25* (4), 369-386. doi: 10.1177/0163278702239051

Benedetti, F., Amanzio, M. (2011). The placebo response: how words and rituals change the patient's brain. *Patient Education and Counselling, 84,* 413-419. doi: 10.1016/j.pec.2011.04.034

Benedetti, F., Carlino, E., & Pollo, A. (2011). Hidden administration of drugs. *Clinical Pharmacological Therapy, 90 (5),* 651-661. doi: 10.1038/clpt.2011.206

Bérdi, M., Köteles, F., Szabó, A., & Bárdos, G. (2011). Placebo effects in sport and exercise: a meta-analysis. *European Journal of Mental Health, 6,* 196-212. doi: 10.5708/EJMH.6.2011.2.5

BrainVision Analyzer (Version 2.2.0) [Software]. (2019). Gilching, Germany: Brain Products GmbH.

Charness, G., Gneezy, U., & Kuhn, M.A. (2012). Experimental methods: between-subject and within-subject design. *Journal of Economic Behavior & Organization, 81,* 1-8. doi: 10.1016/j.jebo.2011.09.009

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(3), 181-204. doi:10.1017/S0140525X12000477

Clayson, P.E. (2020). Moderators of the internal consistency of error-related negativity scores: A meta-analysis of internal consistency estimates. Psychophysiology, *57*(8). doi: 10.1111/psyp.13583

Colloca, L., Sigaudo, M., & Benedetti, F. (2008). The role of learning in nocebo and placebo effects. *Pain, 136*(1-2), 211-218. doi: 10.1016/j.pain.2008.02.006.

Colloca, L., Klinger, R., Flor, H., & Bingel, U. (2013). Placebo analgesia: psychological and neurobiological mechanisms. *Pain, 154*(4), 511-514. doi: 10.1016/j.pain.2013.02.002

Colloca, L., & Miller, F.G. (2011). How placebo responses are formed: a learning perspective. *Philosophical Transactions of the Royal Society B, 366*(1572), 1859-1869. doi: 10.1098/rstb.2010.0398

Colloca, L., & Miller, F.G. (2011). The nocebo effect and its relevance for clinical practice. *Psychosomatic Medicine, 73*(7), 598-603. doi:10.1097/PSY.0b013e3182294a50.

Conforto, A.B., Amaro, E., Goncalves, A.L., Mercante, J.P.P., Guendler, V.Z., Ferreira, J.R., Kirschner, C.C.F.B., & Peres, M.F. (2014). Randomized, proof-of-principle clinical trial of active transcranial magnetic stimulation in chronic migraine. *Cephalalgia, 34* (6), 464-472. doi: 10.1177/0333102413515340.

Crowley, M.J. (2013). Error-Related negativity. *Volkmar F.R. (eds) Encyclopedia of Autism Spectrum Disorders.* New York, NY: Springer. https://doi.org/10.1007/978-1-4419-1698-3_724

Crowley, M.J. (2013). Feedback-Related negativity. *Volkmar F.R. (eds) Encyclopedia of Autism Spectrum Disorders.* New York, NY: Springer. https://doi.org/10.1007/978-1-4419-1698-3_731

Dubljević, V., Venero, C., & Knafo, S. (2015). What is cognitive enhancement? In S. Knafo & C. Venero (Eds.), *Cognitive enhancement: pharmacologic, environmental, and genetic factors.* Academic Press. doi: 10.1016/B978-0-12-417042-1.00001-2

Eriksen, B.A., & Eriksen, C.W. (1974). Effects of noise letters upon the identification of a target letter in a non-search task. *Perception & Psychophysics, 16,* 143-149. doi: 10.3758/BF03203267

Fernandez-Duque, D., Evans, J., Christian, C., & Hodges, S.D. (2015). Superfluous neuroscience information makes explanations of psychological phenomena more appealing. *Journal of Cognitive Neuroscience, 27* (5), 926–944. doi: 10.1162/jocn_a_00750

Geers, A.L., Wellman, J.A., Fowler, S.L., Rasinski, H.M., & Helfer, S.G. (2010). Placebo expectations and the detection of somatic information. *Journal of Behavioural Medicine, 34* (3), 208-217. doi: 10.1007/s10865-010-9301-9

Gehring, W.J., Goss, B., Coles, M.G.H., Meyer, D.E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science, 4*(6), 385-390. https://doi.org/10.1111/j.1467-9280.1993.tb00586.x

Gehring, W.J., Goss, B., Coles, M.G.H., Meyer, D.E., & Donchin, E. (2018). The error-related negativity. Perspectives on Psychological Science, 13(2), 200-204. doi: 10.1177/1745691617715310

Gibbs, N.M. (2010). Placebos and placebo effects. *Anaesthesia and Intensive Care, 38,* 424. doi: 10.1177/0310057X1003800303

Gratton, G., Coles, M.G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology, 55*(4), 468-484. doi: 10.1016/0013-4694(83)90135-9

Gruber, D., & Dickerson, J.A. (2012). Persuasive images in popular science: testing judgments of scientific reasoning and credibility. *Public Understanding of Science*, *21*(8), 938–948. doi: 10.1177/0963662512454072

Haynes, J. D. (2012). "Brain reading" in *I know what you're thinking: brain imaging and mental privacy*. Oxford: Oxford University Press.

Hoogeveen, S., Schjoedt, U., & van Elk, M. (2018). Did I do that? Expectancy of brain stimulation on error-related negativity and sense of agency. *Journal of Cognitive Neuroscience, 30(11),* 1-14. doi: 10.1162/jocn_a_01297

Hopkins, E.J., Weisberg, D.S., & Taylor, J.C.V. (2016). The seductive allure is a reductive allure: people prefer scientific explanations that contain logically irrelevant reductive information. *Cognition, 155,* 67-76. doi: 10.1016/j.cognition.2016.06.011

Hutchinson, A. (2016, June 15). For the Golden State Warriors, brain zapping could provide an edge. *The New Yorker*. https://www.newyorker.com/tech/annals-of-technology/for-the-golden-state-warriors-brain-zapping-could-provide-an-edge.

Inzlicht, M., & Al-Khindi, T. (2012). ERN and the placebo: a misattribution approach to studying the arousal properties of the error-related negativity. *Journal of Experimental Psychology: General, 141,* 799-807. doi: 10.1037/a0027586

Ito, S., Stuphorn, V., Brown, J.W., & Schall, J.D. (2003). Performance monitoring by the anterior cingulate cortex during saccade countermanding. *Science, 302*(5642), 120-122. doi: 10.1126/science.1087847

Jacobs, K.W., & Nordan, F.M (1979). Classifcation of placebo drugs: effect of color. *Perceptual and Motor Skills, 49,* 367-372. doi: 10.2466/pms.1979.49.2.367

Kopp, B., Rist, F., & Mattler, U. (1996). N200 in the flanker task as a neurobehavioral tool for investigating executive control. *Psychophysiology, 33,* 282-294. doi: 10.1111/j.1469-8986.1996.tb00425.x`

Kuenzel, J., Blanchette, I., Zandstra, E.H., Thomas, A., & El-Deredy, W. (2011). Awareness changes placebo effects for feeling relaxed, but not for liking. *Journal of Marketing Communications, 18*(5), 379-396. doi: 10.1080/13527266.2010.548009

Lee, G.J., & Suhr, J.A. (2019). Expectancy effects on self-reported attention-deficit/hyperactivity disorder symptoms in simulated neurofeedback: a pilot study. *Archives of Clinical Neuropsychology, 34*(2), 200-205. doi: 10.1093/arclin/acy026

Locher, C., Gaab, J., & Blease, C. (2018). When a placebo is not a placebo: problems and solutions to the gold standard in psychotherapy research. *Frontiers in Psychology, 9,* Article 2317. doi: 10.3389/fpsyg.2018.02317

Magalhães De Saldanha da Gamma, P.A., Slama, H., Caspar, E.A., Gevers, W., & Cleeremans, A. (2013). Placebo-suggestion modulates conflict resolution in the Stroop task. *PlosOne, 8* (10). doi: 10.1371/journal.pone.0075701

Meissner, K., & Linde, K. (2018). Are blue pills better than green? How treatment features modulate placebo effects. *International Review of Neurobiology, 139,* 357-378. doi: 10.1016/bs.irn.2018.07.014

Michael, R.B., Newman, E.J., Vuorre, M., Cumming, G., & Garry, M. (2013). On the (non)persuasive power of a brain image. *Psychonomic Bulletin Review, 20,* 720-725. doi: 10.3758/s13423-013-0391-6

Miller, F.G., & Kaptchuck, T.J. (2008). The power of context: reconceptualizing the placebo effect. *Journal of the Royal Society of Medicine, 101,* 222-225. doi: 10.1258/jrsm.2008.070466

Mummolo, J., & Peterson, E. (2017). Demand effects in survey experiments: an empirical assessment. *American Political Science Review, 113*(2), 517-529. doi: 10.1017/s000305541800037.

Ongaro, G., & Kaptchuk, T.J. (2019). Symptom perception, placebo effects, and the Bayesian brain. *Pain, 160 (1),* 1-4. doi: 10.1097/j.pain.0000000000001367

Pavlov, I.P. (1927). *Conditioned Reflexes: an Investigation of the Physiological Activity of the Cerebral Cortex* (ed. Anrep, G.V.). Oxford, UK: University Press.

Price, D.D., Finniss, D.G., & Benedetti, F.A. (2008). A comprehensive review of the placebo effect: recent advances and current thought. *Annual Review of Psychology, 59,* 565-590. doi: 10.1146/annurev.psych.59.113006.095941

Ritchie, S. (2020). Hype. In *Science fictions: exposing fraud, bias, negligence, and hype in science.* London, UK: The Bodley Head.

Rief, W., & Petrie, K.J. (2016). Can psychological expectation models be adapted for placebo research? *Frontiers in Psychology, 7.* doi: 10.3389/fpsyg.2016.01876

Roback, P., Legler, J. (2021). Correlated data. In *Beyond multiple linear regression: applied generalized linear models and multilevel models in R.* New York, NY: Chapman and Hall.

Rodilla, C.C., Beauducel, A., & Leue, A. (2016). Error-related negativity and the misattribution of state-anxiety following errors: on the reproducibility of Inzlicht and Al-Khindi (2012). *Frontiers in Human Neuroscience, 10.* Article 475. doi: 10.3389/fnhum.2016.00475

Rosenman, R., Tennekoon, V., & Hill, L.G. (2011). Measuring bias in self-report data. *International Journal of Behavioural and Healthcare Research, 2*(4), 320-332. doi: 10.1504/IJBHR.2011.043414

Roullet, B., & Droulers, O. (2005). Pharmaceutical packaging color and drug expectancy. *Advances in Consumer Research, 32*, 164-171.

Sanders, A.E., Slade, G.D., Fillingim, R.B., Ohrbarch R., Arbes, S.J., & Tchivileva, I.E. (2020). Effect of treatment expectation on placebo response and analgesic efficacy: a secondary aim in a randomized clinical trial. *JAMA Network Open, 3* (4). doi: 10.1001/jamanetworkopen.2020.2907

Schwarz, K.A., & Büchel, C. (2015). Cognition and the placebo effect – dissociating subjective perception and actual performance. *PlosOne*. doi: 10.1371/journal.pone.0130492

Schwarz, K.A., Pfister, R., & Büchel, C. (2016). Rethinking explicit expectations: connecting placebos social cognition, and contextual perception. *Trends in Cognitive Sciences, 20* (6), 469-480. doi: 10.1016/j.tics.2016.04.001

Schweitzer, N.J., Baker, D.A., & Risko, E.F. (2013). Fooled by the brain: re-examining the influence of neuroimages. *Cognition, 129* (3), 501-511. doi: 10.1016/j.cognition.2013.08.009

Schwengerer, L. (2019). Self-knowledge in a predictive processing framework. *Review of* Philosophy and Psychology, 10, 563-585. https://doi.org/10.1007/s13164-018-0416-

Stevens, F.L., Hurley, R.A., & Taber, K.H. (2011). Anterior cingulate cortex: unique role in cognition and emotion. Journal of Neuropsychiatry and Clinical Neuroscience, 23(2), 121-125. doi: 10.1176/jnp.23.2.jnp121

Spence, C. (2021). The multisensory design of pharmaceuticals and their packaging. *Food Quality and Preference, 91.* doi: 10.1016/j.foodqual.2021.104200

Sur, S., & Sinha, V.K. (2009). Event-related potential: an overview. *Industrial Psychiatry Journal, 18* (1), 70-73. doi: 10.4103/0972-6748.57865

Tao, D., Wang, T., Wang, T., & Qu, X. (2018). Influence of drug colour on perceived drug effects and efficacy. *Ergonomics, 61* (2), 284-294. doi: 10.1080/00140139.2017.1349935

Tellegen, A., & Atkinson, G. (1974). Openness to absorbing and self-altering experiences ("absorption"), a trait related to hypnotic susceptibility. *Journal of Abnormal Psychology, 83,* 268-277. doi: 10.1037/h0036681

Turi, Z., Bjørkedal, E., Gunkel, L., Antal, A., Paulus, W., & Mittner, M. (2018). Evidence for cognitive placebo and nocebo effects in healthy individuals. *Scientific Reports, 8,* 17443. doi: 10.1038/s41598-018-35124-w

van Elk, M. (2019). Socio-cognitive biases are associated to belief in neuromyths and cognitive enhancement: a pre-registered study. *Personality and Individual Differences, 147,* 28-32. doi: 10.1016/j.paid.2019.04.014

van Elk, M., Groenendijk, E., & Hoogeveen, S. (2020). Placebo brain stimulation affects subjective but not neurocognitive measures of error processing. Journal of Cognitive Enhancement, 4, 389-400. doi: 10.1007/s41465-020-00172-6

Vase, L., Riley III, J.L., & Price, D.D. (2002). A comparison of placebo effects in clinical analgesic trials versus studies of placebo analgesia. *Pain, 99* (3), 443-452. doi: 10.1016/s0304-3959(02)00205-1

Wan, X., Woods, A.T., Salgado-Montejo, A., Velasco, C., & Spence, C. (2015). Assessing the expectations associated with pharmaceutical pill colour and shape. *Food Quality and Preference, 45,* 171-182. doi: 10.1016/j.foodqual.2015.06.009

Weisberg, D.S., Keil, F.C., Goodstein, J., Rawson, E., & Gray, J.R. (2008). The seductive allure of neuroscience explanations. *Journal of Cognitive Neuroscience, 20* (3), 470-477. doi: 10.1162/jocn.2008.20040

West, R., Jakubek, K., Wymbs, N., Perry, M., & Moore, K. (2005). Neural correlates of conflict processing. *Experimental Brain Research, 167,* 38-48. doi: 10.1007/s00221-005-2366-y

Winkler, A., & Hermann, C. (2019). Placebo- and nocebo-effects in cognitive neuroenhancement: when expectations shape perception. *Frontiers in Psychiatry, 10:498.* doi: 10.3389/fpsyt.2019.00498

## APPENDIX A – Experimental differences between groups

We checked for differences in performance data and subjective measures between experimental groups 0 (first 11 participants) and 1 (last 19 participants). We performed six repeated measure ANOVAs with condition and time as within-subject factor and experimental group as between-subjects factor. We found statistically significant three-way condition*time*group interactions for percentage of incorrect responses ($F$ (16,448) = 1.75, $p$ = 0.036) and subjective performance ($F$ (14,392) = 2.81, $p < 0.001$). We did not find statistical differences for percentage of correct trials ($F$ (16,448) = 1.21, $p$ = 0.26), percentage of missed trials ($F$ (16,448) = 1.76, $p$ = 0.28), reaction times ($F$ (16,448) = 1.18, $p$ = 0.28), or sense of agency ($F$ (14,448) = 1.68, $p$ = 0.058).

These results indicate that, at least to some extent, the different experimental procedure had a significant effect on experimental outcomes. Therefore, we believe the most justifiable choice is to not include the first 11 participants in our final data analysis. The relationship between the variables across time and conditions are depicted in Figure Appendix A.

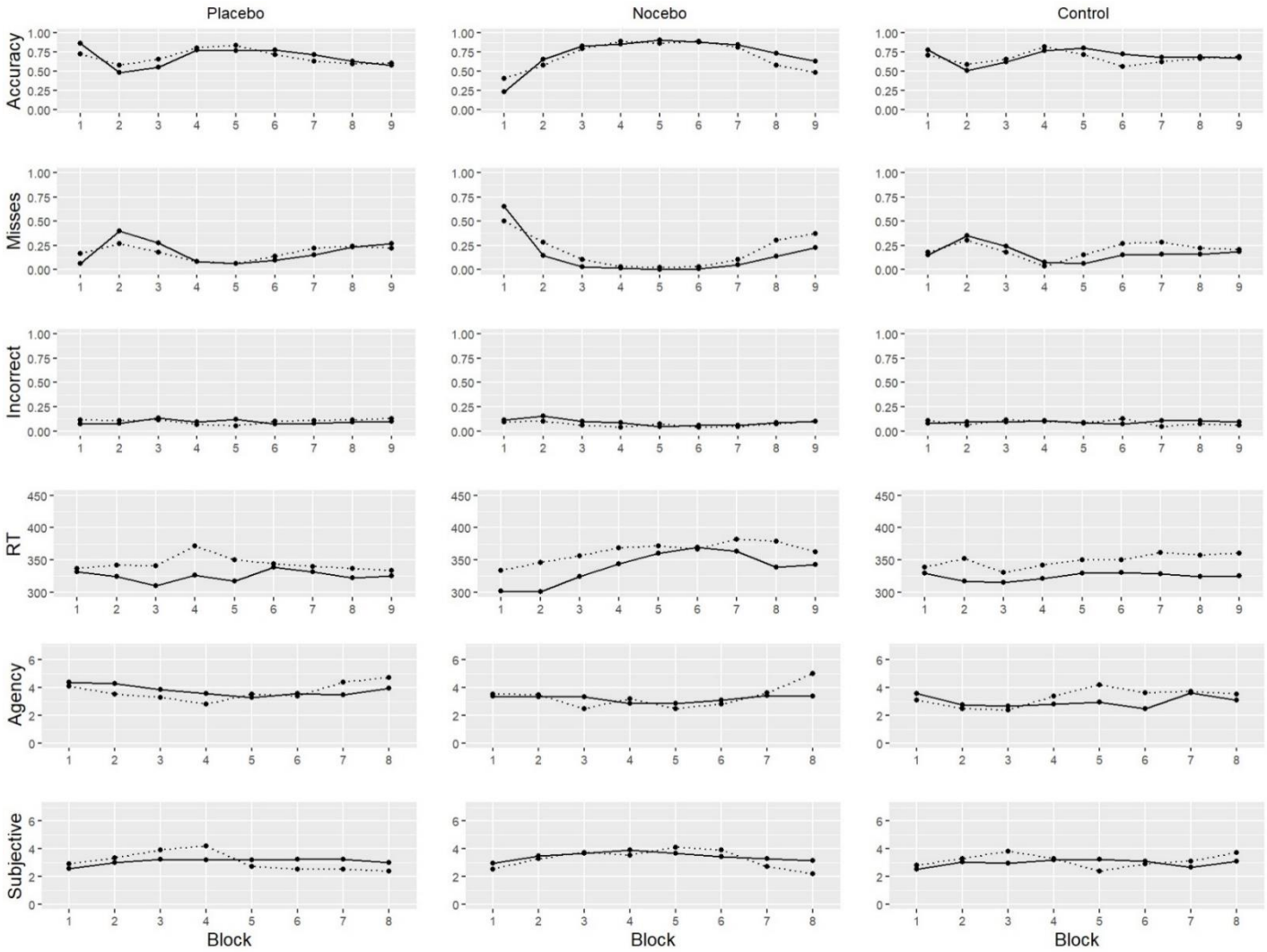


*Figure Appendix A.* Behavioural data measures across time and blocks. The left column contains the behavioural data for the placebo condition, the middle column for the nocebo condition, and the right column for the control condition. For the top three rows, the y-axis depicts a proportion. In the row containing the RT plots, the y-axis depicts milliseconds. The bottom two rows have scale ratings depicted on their y-axes. All the x-axes depict time in blocks. The dotted lines represent the group means of the first experimental group. The solid lines represent the group means of the second experimental group

**APPENDIX B – Performance data conditioning**

After 19 participants we checked if our experimental conditioning did what it was intended to do. Our intentions were to manipulate accuracy in such a way that during the placebo condition, accuracy would be higher than control. In addition, accuracy during nocebo was intended to be lower than during the control condition. Recall that our manipulation consisted out of adjusted reaction times, stimuli presentation time, and stimuli brightness. When looking at the accuracy percentages, we only see a clear distinction between the nocebo conditions and the other conditions. However, the accuracy percentages from the placebo condition do not seem to differ from the control condition.

In addition, we adjusted the parameters based on the accuracy over the entire conditioning phase. This means that when people perform well during the first block, the task would become almost impossible during the entire conditioning phase. This is since after 3 blocks the total accuracy is likely to still be above 50%, which would lead to increasing difficulty. Due to the nature of the task, few mistakes were made during the first block (88.6% accuracy). This led to a conditioning phase that was relatively easy during the first block and extremely hard during the final block. This difficulty seemed to be mainly reflected in missed trials.

We conducted a series of repeated measures ANOVAs with condition as within-subject factor to statistically investigate the effect of conditioning on proportion accuracy, missed trials, and incorrect trials. We found a significant effect of condition on accuracy, $F$ (2,36) = 156.32, $p < .001$, and a significant condition*block effect, $F$ (6,108) = 9.68, $p < .001$. However, a post-hoc inspection revealed a significant difference between the nocebo and control condition, $t$ = -14.21, $p < .001$, but not the placebo and control condition, $t$ = 2.02, $p$ = .051. These results suggest that there was no statistical difference in accuracy between the control and the placebo condition. Next, we found no statistical effect of conditioning on proportion incorrect trials, $F$

(2,36) = 1.13, $p$ = .335, nor did we find a significant condition*block effect, $F$ (6,108) = .398, $p$ = .879. These results suggest that there was no significant difference between conditions on proportion of incorrect trials. Finally, we found a significant effect of condition on proportion of missed trials, $F$ (2,36) = 90.64, $p$ < .001, and a significant condition*block effect, $F$ (6, 108) = 10.51, $p$ < .001. A post-hoc inspection revealed a significant difference between the nocebo and control condition, $t$ = -12.46, $p$ < .001, but not between the placebo and control condition, $t$ = -.05, $p$ = .238. Again, this indicates no significant difference in performance between the placebo condition and the control condition.

Based on this analysis, we concluded that we did not condition the participants correctly and should have adopted a much more dynamic experimental paradigm. Therefor, we decided to terminate data collection after the 30$^{th}$ participant and refine our experimental paradigm. Figure Appendix B depicts performance data over time during the conditioning phase for each condition.

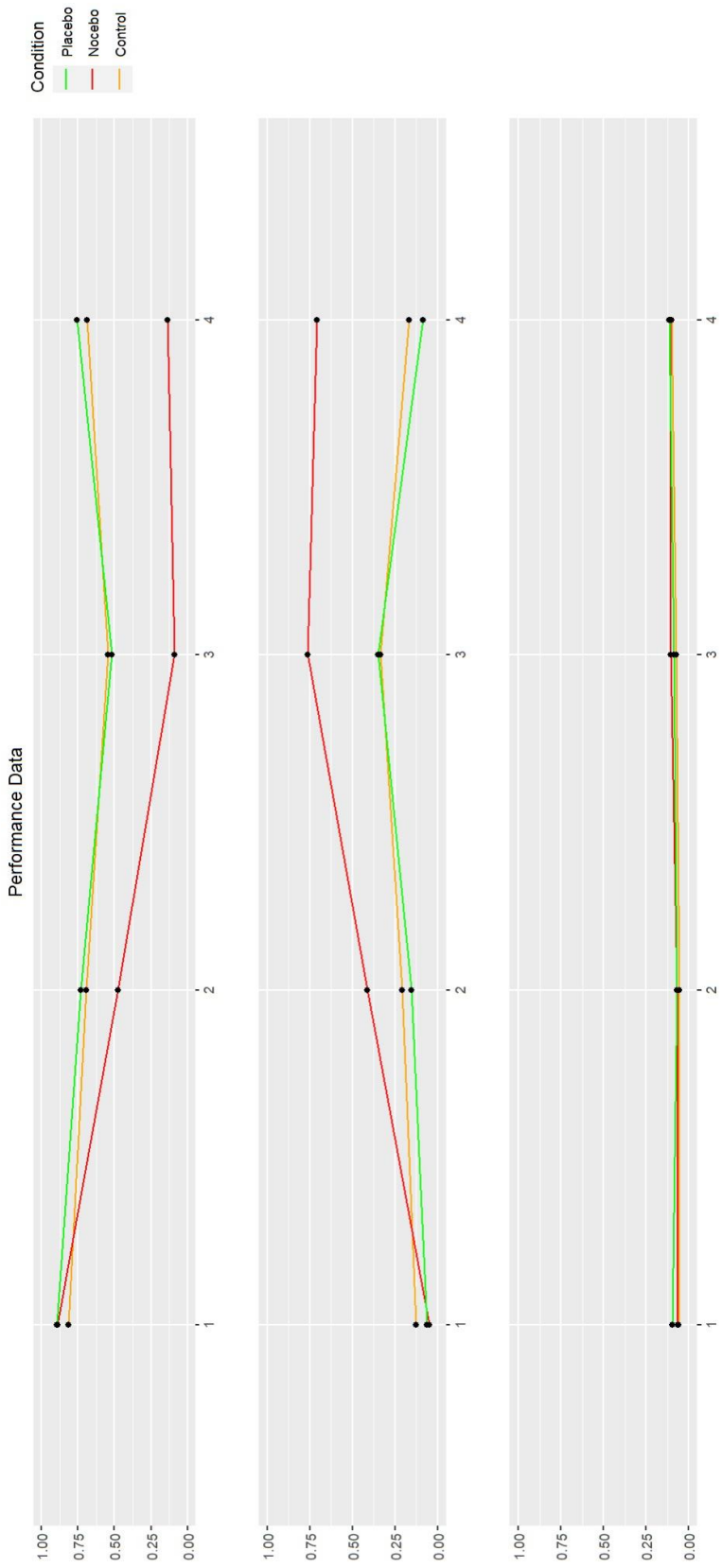


*Figure Appendix B.* Performance data during the conditioning phase. The y-axes represent proportions, and the x-axes represents time (in blocks). The separate lines represent the different conditions. The top panel depicts the proportion correct trials, the middle panel depicts the proportion missed trials, and the bottom panel depicts the proportion incorrect

## Appendix C- Correlation tables

Table 3 -5 display correlations between all outcome measures across conditions. We see correlations between subjective performance and accuracy in all conditions. This means that the higher objective performance is associated with higher subjective performance. In addition, we see that accuracy is not associated with sense of agency, which indicates that differences in sense of agency ratings are probably not associated with differences in accuracy. We also see that the susceptibility for neuromyths scale is correlated with all EEG outcome measures. The amount participants were convinced about the fact that we used brain stimulation is correlated with sense of agency ratings and FRN amplitude.

Table 3. Correlations between outcome measures in the placebo condition

| | Accuracy | SoA | SP | AR | BiN | ERN | FRN | N2c | BiE |
|---|---|---|---|---|---|---|---|---|---|
| Sense of Agency (SoA) | .028 | | | | | | | | |
| Subjective Performance (SP) | .501** | -.103 | | | | | | | |
| Absorption Rating (AR) | .071 | -.110 | -.051 | | | | | | |
| Susceptibility for Neuromyths (BiN) | .033 | -.305** | .227** | .271** | | | | | |
| ERN | .049 | -.206* | .087 | .146 | -.231** | | | | |
| FRN | .031 | -.196* | .104 | -.258** | .256** | -.084 | | | |
| N2c | .035 | .079 | -.066 | .367** | -.326** | .546** | .064 | | |
| Believe in Efficacy (BiE) | .167* | -.396** | .126 | .180* | .422** | -.063 | .407** | -.004 | |
| Convincingness of Experiment (CoE) | .110 | -.345** | .176* | .116 | .325** | .112 | .332** | -.065 | .522** |

* Significant at $p < .05$ level
** Significant at $p < .001$ level

Table 4. Correlations between outcome measures in the nocebo condition

| | Accuracy | SoA | SP | AR | BiN | ERN | FRN | N2c | BiE |
|---|---|---|---|---|---|---|---|---|---|
| Sense of Agency (SoA) | .278** | | | | | | | | |
| Subjective Performance (SP) | .447** | .228** | | | | | | | |
| Absorption Rating (AR) | -.022 | .048 | .020 | | | | | | |
| Susceptibility for Neuromyths (BiN) | -.021 | -.240** | .394** | .271** | | | | | |
| ERN | -.047 | .142 | -.133 | .269** | -.273** | | | | |
| FRN | -.012 | -.068 | .268** | .117 | .230** | .221* | | | |
| N2c | -.089 | .145 | -.137 | .237** | -.331** | .668** | .339** | | |
| Believe in Efficacy (BiE) | -.079 | -.269** | -.126 | .187* | .415** | .019 | .008 | -.227* | |
| Convincingness of Experiment (CoE) | -.116 | -.401** | .100 | .125 | .317** | .032 | .188* | -.233** | .520** |

* Significant at $p < .05$ level
** Significant at $p < .001$ level

Table 5. Correlations between outcome measures in the control condition

| | Accuracy | SoA | SP | AR | BiN | ERN | FRN | N2c | BiE |
|---|---|---|---|---|---|---|---|---|---|
| Sense of Agency (SoA) | .282** | | | | | | | | |
| Subjective Performance (SP) | .438** | .108 | | | | | | | |
| Absorption Rating (AR) | -.082 | -.004 | .074 | | | | | | |
| Susceptibility for Neuromyths (BiN) | -.002 | -.205** | .098 | .271** | | | | | |
| ERN | -.049 | .190 | -.056 | .494** | -.841** | | | | |
| FRN | -.014 | -.176 | .260* | -.014 | .514** | -.334* | | | |
| N2c | -.030 | .167 | -.128 | .167 | -.380** | .380** | -.881** | | |
| Believe in Efficacy (BiE) | .027 | -.138 | -.122 | .179* | .412** | -.603** | -.085 | .041 | |
| Convincingness of Experiment (CoE) | .012 | -.353** | .207** | .125 | .317** | -.650** | .059 | -.080 | .523** |

* Significant at $p < .05$ level
** Significant at $p < .001$ level