Social Cognitive and Affective Neuroscience Advance Access published May 1, 2014

"Why should I care?" Challenging free will attenuates neural reaction to errors

Running title: Determinism alters error processing

Davide Rigoni[1*], Gilles Pourtois[2], Marcel Brass[1,3]

[1]Department of Experimental Psychology, Ghent University, Henri Dunantlaan 2, 9000 Gent, Belgium

[2]Department of Experimental Clinical and Health Psychology, Ghent University, Henri Dunantlaan 2, 9000 Gent, Belgium

[3]Behavioural Science Institute, Radboud University Nijmegen, Montessorilaan 3, A.08.29

6525 HR Nijmegen, The Netherlands

*Corresponding author:

Dr Davide Rigoni

Department of Experimental Psychology, University of Gent

H. Dunantlaan 2, 9000 - Gent, Belgium

davide.rigoni@ugent.be

Abstract

Whether human beings have free will or not has been a philosophical question for centuries. The debate about free will has recently entered the public arena through mass media and newspaper articles commenting on scientific findings that leave little to no room for free will. Previous research has shown that encouraging such a deterministic perspective influences behavior, namely by promoting cursory and antisocial behavior. Here we propose that such behavioral changes may, at least partly, stem from a more basic neurocognitive process related to response monitoring, namely a reduced error detection mechanism. Our results show that the Error-Related Negativity, a neural marker of error detection, was reduced in individuals led to disbelieve in free will. This finding shows that reducing the belief in free will has a specific impact on error detection mechanisms. More generally, it suggests that abstract beliefs about intentional control can influence basic and automatic processes related to action control.

Keywords: free will, belief, response monitoring, error detection, error-related negativity

1. Introduction

Whether human beings have free will or not has been a core philosophical question for centuries. Many scientists today are quite skeptical about the existence of free will, at least in its traditional conception, as it would imply a "*ghost in the machine*" (Ryle, 1949) and hence entail the homunculus fallacy. Already more than 40 years ago, the famous behaviorist B.F. Skinner argued against the "*autonomous man – the inner man, the homunculus, the possessing demon, the man defended by the literatures of freedom and dignity*" (Skinner, 2002, p. 200). More recently, the debate about free will has entered the public domain through mass media and newspaper articles commenting on scientific findings that seemingly leave little to no room for free will. This raises a fundamental question, if it really matters whether people believe in free will or not.

Recent social psychological research has started to address this question. Empirical evidence suggests that believing in free will has important consequences for our lives, health and psychological well-being. For instance, Stillman and colleagues could demonstrate that individuals with strong beliefs in free will have better actual job performance and career attitudes than individuals with low beliefs in free will (Stillman, Baumeister, Vohs, Lambert, Fincham, & Brewer, 2010). Recent studies in social psychology have shown that being exposed to scientific information weakening the belief in free will can lead to negative changes in social behavior, such as cheating (Vohs & Schooler, 2008) and aggressiveness, and reduces prosocial and altruistic behavior (Baumeister, Masicampo, & DeWall, 2009). Further, it has been shown that deterministic primes threatening free will lead to reduced autonomous and independent thoughts (Alquist, Ainsworth, & Baumeister, 2012). These results suggest that believing in free will is crucial for motivating people to implement effort and deliberation in order to overcome automatic thinking and behavior.

Why do people that are artificially encouraged to disbelieve in free will show cursory and antisocial tendencies? It has been argued that weakening the belief in free will reduces individuals' will to exert self-control (Vohs & Schooler, 2008; Baumeister et al., 2009). Self-control is demanding and energy-consuming (Baumeister, 2008), and a disbelief in free will would prevent individuals to invest that energy. Here we focused on much more basic neurocognitive processes that may underlie the cursory pattern displayed by individuals who are artificially primed to disbelieve in free will. More specifically, we propose that such behavioral changes arise from altered response monitoring processes, namely a reduced error detection. Interestingly, recent experiments in cognitive neuroscience showed that challenging free will can alter basic cognitive and neural processes underlying voluntary actions. People induced to disbelieve in free will show altered brain activity underlying motor readiness (Rigoni, Kühn, Sartori, & Brass, 2011), reduced intentional inhibition and perceived control (Rigoni, Kühn, Gaudino, Sartori, & Brass, 2012), and altered post-error adaptation (Rigoni, Wuilquin, Brass, & Burle, 2013).

The aim of the present study is to test whether weakening people's belief in free will can alter a basic brain process considered to be the hallmark of efficient action control, that is, response monitoring. This cognitive capacity is particularly important when errors can be committed, as it is crucial to identify whether our ongoing response is correct or erroneous in order to adjust behavior accordingly. Our specific hypothesis is that reducing the belief in free will can alter the neurophysiological correlates of error monitoring. We recorded response-locked Event-Related Potentials (ERPs) while healthy adult participants performed a previously validated speeded go/nogo task (Vocat, Pourtois, & Vuilleumier, 2008; Pourtois, 2011). In a typical go/nogo task, participants are asked to respond as quickly as possible to dominant *go* stimuli that are often presented on a computer screen. In a minority

of trials, deviant *nogo* stimuli indicate that the participant has to withhold responding. The use of time pressure in this type of tasks typically results in people committing a considerable number of response errors. Earlier psychophysiological studies showed that when a response error is produced during similar interference tasks, a negative wave appears over fronto-central electrodes starting just before the onset of the mechanical response and peaking within ~0-100 ms following it (Falkenstein, Hohnsbein, Hoormann, & Blanke, 1990; 1991; Gehring, Coles, Meyer, & Donchin, 1990; Holroyd & Coles, 2002; Yeung, Botvinick, & Cohen, 2004). This error-related deflection is often referred to as Error-Related Negativity (ERN) and empirical evidence suggests that it is generated within the medial prefrontal cortex, most likely the anterior cingulate cortex (Ridderinkhof, Ullsperger, Crone, & Nieuwenhuis, 2004) or the supplementary motor area (Bonini, Burle, Liégeois-Chauvel, Régis, Chauvel, & Vidal, 2014). Different cognitive interpretations have been put forward to account for the functional significance of the ERN (e.g. "pure" error detection, conflict monitoring, "generic" mismatch detection, negative reinforcement learning), but at present there is no general agreement on what is the specific function reflected by the ERN. However, there is little doubt that the ERN is involved in the early stages of response monitoring during or following action execution (e.g. Weinberg, Riesel, & Hajcak, 2012).

Recent evidence suggests that the ERN does not reflect a purely cognitive response monitoring mechanism, but rather it is modulated by motivational and affective factors and can be influenced by current individuals' motivational states. In this view, the ERN would entail a rapid motivational reaction to a deviation of one's own performance, and would thus represent a kind of "*cortical alarm bell*" (Inzlicht & Tullett, 2010), a neural marker of how much a person is concerned about performance outcomes. Empirical support for a motivational/affective account of the ERN comes from studies that

examined the effect of distress levels on people's reaction to errors. A few studies found larger ERN in patients with anxiety disorders than in healthy controls (Gehring, Himle, & Nisenson, 2000; Aarts & Pourtois, 2010). Similarly, anxiolytic and antidepressant drugs reduce the amplitude of the ERN (Johannes, Wieringa, Nager, Dengler, & Munte 2001; de Bruijn, Sabbe, Hulstijn, Ruigt, & Verkes, 2006). Other studies have described how the amplitude of the ERN changes according to contextual motivational factors, such as monetary incentives that are offered for accuracy (Gehring, Goss, Coles, Meyer, & Donchin, 1993; Gehring & Willoughby, 2002; Pailing & Segalowitz, 2004). It has also been proposed that very abstract beliefs, such as religious beliefs, can mitigate distress reactions to errors (Inzlicht, NcGregor, Hirsh, & Nash, 2009; Inzlicht & Tullett, 2010). For instance, Inzlicht and Tullett (2010) showed that religious primes reduce the amplitude of the ERN in individuals who strongly believe in God.

Consistent with this framework, we propose that the cursory inclination displayed by individuals encouraged to disbelieve in free will (Vohs & Schooler, 2008; Baumeister et al., 2009) is associated to traceable changes in brain systems that underlie response monitoring mechanisms, and in particular the detection of response errors. More specifically, we reckon that artificially weakening the belief in free will (through a standard induction technique, see Vohs & Schooler, 2008; Rigoni et al., 2011; 2013) could alter early stages of response monitoring, and thus blunt the ERN.

Another ERP component that is observed when people commit a response error is the error positivity (Pe; Falkenstein et al., 1991). The Pe is a slow positive wave that follows the ERN and that peaks between 200 and 500 ms following error commission. The Pe is not a uniform deflection as it appears to consist of an initial component (early-Pe) that immediately follows the ERN and has a fronto-central distribution, and a slower component (late-Pe) that shows a more posterior scalp distribution

(O'Connell, Dockree, Bellgrove, Turin, Ward, Foxe, & Robertson, 2009; Endrass, Klawohn, Preuss, & Kathmann, 2012). Empirical evidence suggests that the Pe reflects aspects of error-related processing that are independent from those manifested by the ERN (for a review, see Overbeek, Nieuwenhuis, & Ridderinkhof, 2005). More precisely, the Pe appears to reflect error awareness and is linked to strategic and remedial processes, such as post-error adaptation (Endrass, Franke, & Kathmann, 2005; Nieuwenhuis, Ridderinkhof, Blom, Band, & Kok, 2001; Ridderinkhof, Ramautar, & Wijnen, 2009). However, given that the specificity of the Pe for error monitoring processes remains debated (Ridderinkhof et al., 2009), we did not formulate clear predictions regarding possible modulatory effects of free will depletion on this component in the current study.

2. Method

2.1. Participants

Thirty-three volunteers (26 females, 7 males; mean age = 23.14, SD = 3.95) with no psychiatric or neurological history participated to the experiment, provided informed consent, and were compensated 20 euros for their participation. Data from 3 participants were excluded from statistical analyses because there were less than 6 artifact-free error trials in the baseline session (see section 2.2.), a number that prevents to compute reliable error-related ERP waveforms (Olvet & Hajcak, 2009). The study was conducted in accordance with the Declaration of Helsinki and was approved by the ethics committee of the Faculty of Psychological and Educational Sciences, Ghent University.

2.2. Experimental design and procedure

The experimental design was divided into a baseline and a post-manipulation session.

2.2.1. Baseline session and belief manipulation

In the *baseline session*, after installing the EEG cap and the electrodes, participants performed the speeded go/nogo task (see section 2.3.). Participants were then randomly allocated to one out of two different conditions. The *no-free will* group (n=15) read an excerpt from the book entitled *The Astonishing Hypothesis*, by Francis Crick (1994), that challenges in several ways the existence of free will – e.g. that scientists now recognize that free will is an illusion, that there is no soul, that our behavior is totally determined by chemical processes. Conversely the *control* group (n=15) read a passage from the same book that did not mention free will (Vohs & Schooler, 2008). Both texts were translated into Dutch by a native speaker and were about one A4 page long. The anti free will text counted 644 words, while the control text counted 529 words. Participants were given up to 5 minutes and were encouraged to read the text carefully, and they were told that a comprehension/memory test would be administered at the end of the experiment.

2.2.1. Post-manipulation session

During the *post-manipulation session* (i.e. after reading the text), each participant carried out again the same speeded go/nogo task. At the end of the experimental session, participants completed the Positive and Negative Affective Schedule (PANAS; Watson, Clark and Tellegen 1988), and the Free Will and Determinism-Plus scale (FAD-Plus; Paulhus and Carey, 2011). The FAD-Plus is composed by 27 Likert-type items (scores ranging from 1 = totally disagree, to 5 = totally agree) measuring beliefs about free will and related constructs (i.e. free will, scientific determinism, fatalistic determinism, and unpredictability). A global FAD-Plus score was calculated separately for each individual (Lynn, Van Dessel and Brass, 2013). In order to obtain a positive score of the belief in free will, individuals scores

on the Scientific Determinism (7 items), the Fatalistic Determinism (5 items), and Unpredictability (8 items) subscales were inversed and aggregated to the score on the Free Will subscale (7 items).

The whole experiment lasted about 2 hours.

2.3. Stimuli and task

The go/nogo task used in the present study has been used and described extensively elsewhere (Vocat, Pourtois, Vuilleumier 2008; Koban, Pourtois, Vocat, Vuilleumier 2010; Koban, Brass, Lynn, Pourtois 2012; Pourtois, 2011; Aarts & Pourtois, 2010; 2012). Each trial started with a fixation cross, displayed for 1 s in the center of the screen, followed by a black arrow that changed color after a randomly jittered interval (ranging from 1s to 2 s). In 2/3 of the trials – i.e. go trials – the arrow turned green, indicating that participants should respond as quickly as possible by pressing the space bar. In the remaining 1/3 of the trials – i.e. nogo trials – the arrow turned either turquoise, or turned green but changed in-plane orientation (relative to the black arrow), indicating that participants had to withhold their response. One second after the response, a performance feedback was given as a green or red circle for correct versus incorrect responses, respectively. Participants performed six blocks in the baseline session and six blocks in the post-manipulation session. Each of the six blocks consisted of 60 trials shown in random order (40 go and 20 nogo), resulting in 360 trials in total (~30 min) per session. In order to obtain an adequate number of "unwanted" response error in both groups, an individually calibrated and updated speed pressure was imposed (yielding many false alarms on the nogo trials throughout the experimental session; see Vocat et al., 2008). Using this procedure, only fast responses (i.e., hits made below a pre-defined RT cutoff; "fast hits") were eventually associated with a positive feedback. RTs for hits above this limit ("slow hits") were associated with a negative feedback. Fast vs.

slow responses were determined by calibrating an individual RT limit in three interspersed calibration blocks. Each calibration block was similar to the experimental blocks, but consisted of only 12 trials (8 go, 4 nogo), and was presented before two consecutive experimental blocks. Calibration of RT was never disclosed to participants. This speed pressure promoted the use of an impulsive response mode and hence the occurrence of many responses errors, allowing for reliable error-related ERP waveforms in each condition.

The experiment took place in an electromagnetically shielded room. Stimulus presentations and all stimulus and response timing were controlled using E-Prime 2 (Schneider, Eshman, & Zuccolotto, 2002). Stimuli were displayed on a 17 inch Dell 1708FPb flat panel monitor positioned at eye height approximately 60 cm from participants.

2.4. EEG recording and signal processing

EEG was recorded (sampling rate 2048 Hz) from 64 active Ag-AgCl electrodes (BioSemi Active-Two, BioSemi, Amsterdam) mounted in an elastic cap. Raw EEG traces were down-sampled offline to 512 Hz, re-referenced to averaged mastoids and filtered (0.1–30 Hz). Ocular movements were corrected following the standard procedure developed earlier by Gratton and Coles (1983) as implemented in Brain Vision Analyzer 2.0 software. The epochs were segmented from −500 to 500 ms time-locked to the response, separately for fast correct hits and commission errors. Epochs containing global and local (i.e. at one electrode only) artifacts were rejected prior to baseline correction (from -500 ms to -300 ms relative to the button press) on the basis of visual inspection and automatic artifact detection (i.e. epochs exceeding ±100 μV). Individual epochs were averaged separately for each condition and participant.

Difference waves were then computed for each participant by subtracting the averaged waveform for fast hits from the averaged waveform for response errors (Falkenstein et al., 1991). The ΔERN was calculated at electrodes Fz, FCz, and Cz as the mean amplitude in the time window from -25 ms to 25 ms around the response. While the early-PE and the late-Pe usually have distinctive scalp distributions – i.e. frontal-central and centro-parietal, respectively – the latency of these two components depends on the specific task demands (Overbeek et al., 2005). In light of the topographical properties of the current data set and based on previous ERP studies using a similar go/nogo task (Ruchsow, Grön, Reuter, Spitzer, Hermle, & Kiefer, 2005; Falkenstein, Hoormann, Christ, & Hohnsbein, 2000), the early-Pe was computed at fronto-central electrodes Fz, FCz, and Cz as the mean amplitude spanning from 100 ms to 250 ms following response onset. The late-Pe was measured at posterior parietal electrodes CPz and Pz as the mean amplitude spanning from 250 ms to 400 ms after response onset.

3. Results

3.1. Go/nogo task

Error rate and RTs were submitted to separate mixed ANOVAs with Session (baseline, post-manipulation) as within-subjects factor and Group (no-free will, control) as between-subjects factor. Since slow hits are typically characterized by a different RT distribution and are more frequent than fast hits and errors (Pourtois, 2001), they were analyzed separately in the RTs analysis. In addition, while fast hits and errors are associated to a clear positive and negative valence, respectively, slow hits are typically associated to a mixed and undifferentiated valence (i.e. neither clearly positive nor negative) (Aarts, Houwer, & Pourtois, 2012). For these reasons, slow hits trials were not included in the EEG analysis.

3.1.1. Error rate

Overall, participants committed more errors in the post-manipulation session as compared to the baseline (13.42 % ± 7.02 vs. 8.8 % ± 4.68, respectively), as revealed by the main effect of Session ($F(1,28) = 26.47$, $p < .001$, $\eta_p^2 = .49$). Error rate was slightly larger in the no-free will group than in the control group (12.79 % ± 6.51 vs. 9.39 % ± 3.57, respectively), but neither the main Group effect ($F(1,28) = 3.14$, $p = .09$) nor the Group × Session interaction ($F(1,28) = .21$, $p = .65$) reached significance level ($\alpha = .05$).

3.1.2. RTs

Responses faster than 50 ms were considered as anticipations and were therefore excluded from the analyses. The Type of trial (fast hits, slow hits, errors) was entered into a mixed ANOVA with Session as within-subjects factors, and Group as between-subjects factor. The main effect of Session showed that overall RTs were faster in the post-manipulation session than in the baseline session (315.44 ms ± 44.23 vs. 345.46 ms ± 37.72, respectively; $F(1,28) = 39.41$, $p < .001$, $\eta_p^2 = .59$). RTs were slower for slow hits (395.67 ms ± 50.51) as compared to fast hits (292.73 ms ± 37.51) and errors (302.95 ms ± 37.14), as revealed by a significant effect of the Type of trial ($F(2,56) = 244.49$, $p < .001$, $\eta_p^2 = .89$). The Session × Type of trial interaction was also found to be significant ($F(2,56) = 3.33$, $p = .04$, $\eta_p^2 = .11$), suggesting a differential effect of Session on the different types of trials (fast hits: 277 ms ± 47 vs. 307 ms ± 31, post-manipulation session and baseline session, respectively; slow hits: 376 ms ± 55 vs. 414 ms ± 54; errors: 291 ms ± 40 vs. 314 ms ± 37). Neither other effects nor interaction reached or approached significance ($.14 < p_s < .79$).

We also wanted to test whether the anti-free will manipulation led to reduced post-error slowing in the no-free will group, as reported previously (Rigoni et al., 2013). RTs were submitted to ANOVA with Session and Previous trial (correct, error) as within-subjects factors and Group as between-subjects factors. The analysis revealed a main effect of Previous trial ($F(1, 28) = 38.21$, $p < .001$, $\eta_p^2 = .58$), with slower RTs for trials following errors than after correct trials (357.3 ms ± 49.47 vs. 339.31 ms ± 41.6, respectively). The main Session effect ($F(1, 28) = 23.57$, $p < .001$, $\eta_p^2 = .46$) revealed that participants were overall faster in the post-manipulation session as compared to the baseline session (334 ms ± 51.76 vs. 362.61 ms ± 43.35, respectively). RTs were numerically but not significantly faster in the no-free will group than in the control group (333.58 ms ± 47.95 vs. 363. 04 ms ± 37.91; $F(1,28) = 3.48$, $p = .07$). The Session × Type of trial × Group interaction was not significant ($F(1, 28) =.45$, $p = .51$), showing that the anti-free will manipulation did not lead to reduced post-error slowing in the no-free will group. No other effect reached or approached significance ($.51 < p_s < .94$).

3.2. ERPs

Grand average ERP Δ-waveforms are shown in Figure 1. Components amplitude was entered as dependent variable into a mixed ANOVAs with Session (baseline, post-manipulation) and Site (Fz, FCz, and Cz for the ERN and the early-ΔPE; CPz and Pz for the late-ΔPE) as within-subjects factors, and Group (no-free will, control) as between-subjects factor.

3.2.1. ΔERN

The main effect of Site revealed that the ΔERN was not uniform across the midline ($F(2,58) = 30.44$, $p < .001$, $\eta_p^2 = .15$). Paired-samples t-tests showed that it was larger at Cz (-12.31 μV ± 7.12) as

compared to both FCz (-11.04 μV ± 6.62; t(29) = -3.49, p = .002) and Fz (-8.47 μV ± 5.32; t(29) = -5.91, p < .001). Moreover, the ERN uniformly decreased in the post-manipulation as compared to the baseline session (-9.96 μV ± 6.73 vs. -11.26 μV ± 6.27, respectively), as revealed by a marginally significant effect of Session (F(1,28) = 3.73, p = .06, η_p^2 = .12). Crucially for our prediction, we found that the amplitude of ΔERN decreased in the post-manipulation compared to the baseline session in the no-free will group selectively (-11.38 μV ± 7.51 vs. -8.47 μV ± 6.64), with no such effect in the control group (-11.14 μV ± 5.02 vs. -11.44 μV ± 6.71), as revealed by a significant Session × Group interaction effect (F(1,29) = 5.69, p = .02, η_p^2 = .17). Paired-samples t-tests confirmed that the ΔERN significantly decreased in the post-manipulation session compared to the baseline session in the no-free will group (-8.47 μV ± 6.64 vs. -11.38 μV ± 7.51, respectively; t(14) = -3.71, p = .002), while it did not in the control group (-11.44 μV ± 6.72 vs. -11.14 μV ± 5.02, respectively; t(14) = .28, p = .78). Importantly, independent-samples t-tests showed that the no-free will group and the control group had comparable ΔERN amplitudes in the baseline session (t(28) = 28, p = .92), thereby excluding potential confounding due to a sampling bias. Neither other main effects nor interactions reached significance (.14 < p_s < .83).

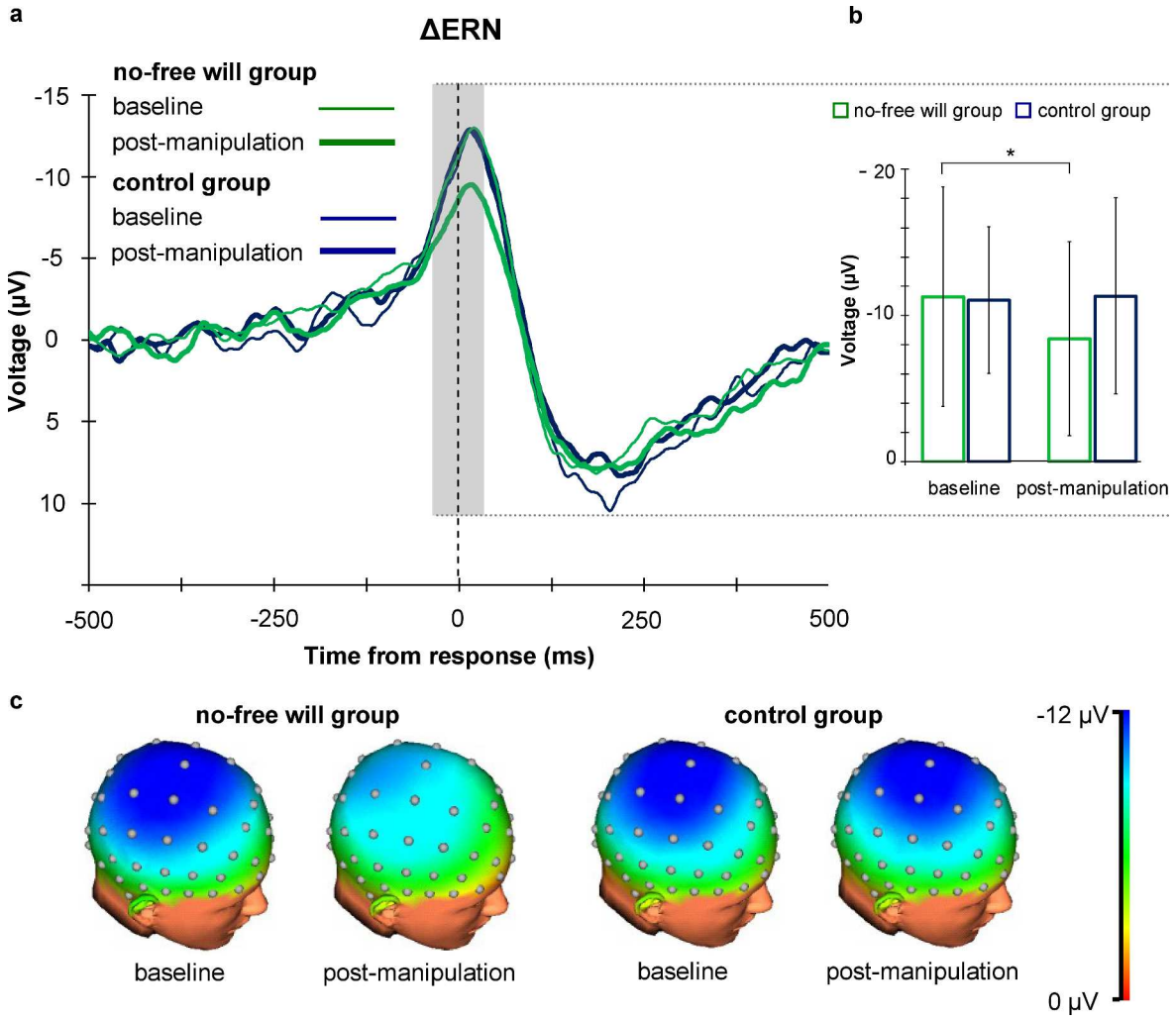

Figure 1 (a) Grand-averaged waveforms from aggregated electrodes (Fz, FCz, and Cz) showing the **Δ**ERN activity (from -25 ms to 25 ms relative to response onset, baseline correction from -500 ms to -300 ms) before and after the belief manipulation, both for the no-free will group and the control group. (b) Bar graph showing mean amplitudes of the **Δ**ERN (error bars refer to standard error of the mean). A significant reduction of the **Δ**ERN amplitude was found in the no-free will group, but not for the control group. (c) 3-D topographic maps showing the topographical distribution of the **Δ**ERN in the two groups before and after the belief manipulation.

3.2.2. ΔPe

Two distinct components were clearly expressed during the time-interval corresponding to the Pe. The early-ΔPe had a clear fronto-central distribution, and peaked between 150 and 250 ms. Conversely, the

late-ΔPe showed a parietal maximum between 300 and 400 ms. Concerning the early-ΔPe, the main effect of Site was significant ($F(2,56) = 14.76$, $p < .001$, $\eta_p^2 = .35$). Paired-comparisons showed that the early-ΔPe was larger at FCz (8.09 μV ± 5.91) than at either Fz (6.84 μV ± 5.39; $t(29) = 4.99$, $p < .001$) or Cz (6.19 μV ± 4.75; $t(29) = 4.76$, $p < .001$). No any other effect reached significance ($.15 < p_s < .87$). As for the late-ΔPe, its amplitude was not uniform across scalp (4.84 μV ± 4.66 vs. 4.15 μV ± 3.52; CPz and Pz, respectively), although the effect of Site was not significant ($F(1,28) = 3.44$, $p = .07$). Neither other main effects nor interactions reached or approached significance ($.22 < p_s < .86$).

3.3. FAD-plus

First, we tested whether the two groups differed in their FAD-plus score after the manipulation. Previous research has shown inconsistent results regarding the effect of the no-free will manipulation on FAD-plus (Rigoni et al., 2013). While numerically the no-free will group indeed showed reduced FAD-plus score after the belief manipulation compared to the control group (64.53 ± 10.03 vs. 68.73 ± 8.48) this difference was not significant ($t(28) = -1.24$, $p = .23$).

More importantly, we wanted to test whether the belief in free will was associated with the amplitude of the ERN by performing a correlation analysis between the FAD-Plus score and the amplitude of this early error-monitoring component in the post-manipulation session, separately for each group. The global FAD-Plus score correlated negatively with the ERN amplitude in the no-free will group (Spearman's $rho(15) = –.54$, $p = .04$), suggesting that individuals with low belief in free will had a blunted ERN. Such a systematic negative relationship was not found in the control group (Spearman's $rho(15) = .42$, $p = .12$).

3.4. Auxiliary measures

We wanted to undermine the possibility that other processes, unrelated to the belief in free will per se, might contaminate or conflate the belief manipulation and hence account for the present ERP results. Reading a text that challenges free will may have triggered an unspecific emotional arousal in the no-free will group, and this in turn might have affected the ERN. A mixed ANOVA with the type of PANAS scale (positive, negative) as within-subjects factor and Group (no-free will, control) as between-subjects factor showed that the two groups scored differently on the PANAS, as revealed by the significant type of PANAS scale × Group interaction ($F(1,28) = 4.39$, $p = .045$, $\eta_p^2 = .14$). The no-free will group reported higher scores to the positive scale of the PANAS as compared to the control group, but the difference was only marginally significant (30.73 ± 6.89 vs. 26.40 ± 7.54; $t(28) = 1.64$, $p = .11$). No differences were found concerning the negative scale of the PANAS (15.2 ± 5.19 vs. 17.33 ± 6.74; $t(28) = -.97$, $p = .34$). Scores to the negative and positive scales of the PANAS did not show a correlation with the ΔERN amplitude in the no-free will group (Spearman's $rho(15) = -.36$, $p = .19$; Spearman's $rho(15) = -.26$, $p = .35$; respectively), suggesting that the reduction of the ΔERN in the no-free will group was not influenced by affective changes induced by the belief manipulation.

4. Discussion

Here we could demonstrate that "artificially" weakening belief in free will through a standard implicit induction technique can alter basic error monitoring processes during a simple and fully unrelated speeded go/nogo task. Participants who were led to disbelieve in free will showed reduced ΔERN amplitude, while those who were not induced to disbelieve in free will did not. This finding dovetails

the hypothesis that down-playing free will (by means of reading a scientific excerpt) can have profound repercussions at the neurocognitive level during the earliest stage of response monitoring.

The reduction of the ΔERN could not be due to asymmetries at the behavioral level between the two groups. We used a stringent RTs calibration procedure during the speeded go/nogo task enabling us to match RT speed and accuracy (and hence the frequency of response errors) between the two groups. While participants were overall less accurate during the post-manipulation session relative to the baseline session across the two groups, this unspecific effect did not interact with our main experimental manipulation.

The specific interpretation of the current findings depends on the functional significance of such error detection system – i.e. the ERN/Pe complex. Strikingly, the belief manipulation effect was specific to the ΔERN, leaving the subsequent ΔPe unaffected by the belief manipulation. This finding corroborates the notion that these two consecutive error-related activities reflect dissociable processes during performance monitoring and action control (Nieuwenhuis et al., 2001; Ridderinkhof, Ramautar, & Wijnen,2009). The different interpretations of the functional significance of the ERN share in common that this error-related component somehow reflects the early (online) evaluation of the motor performance during action execution (e.g., Weinberg, Riesel, & Hajcak, 2012). It has been proposed that at this early stage, the error detection process reflected by the ERN operates independently of conscious error perception, and is not directly involved in the activation of remedial processes – i.e. the correction of the ongoing erroneous response. Conversely, a later error monitoring process, reflected by the Pe, is associated with the awareness of the occurrence of the incorrect response and is related to post-error behavioral adaptation (Nieuwenhuis et al., 2001; Endrass, Reuter, & Kathmann, 2007; Steinhauser, & Yeung, 2010). Thus, while the ERN presumably indexes an automatic detection of

response errors, the Pe would translate more strategic regulatory processes during action monitoring, likely requiring conscious awareness. Our ERP results therefore indicate that the belief manipulation influences selectively action monitoring processes deemed rapid and automatic (ΔERN), while leaving untouched more controlled and conscious processes (ΔPe). However, earlier studies provided evidence that the ERN reflects brain processes that are linked to remedial action processes alike (Burle, Roger, Allain, Vidal & Hasbroucq, 2008; Bonini et al., 2014; Debener, Ullsperger, Siegel, Fiehler, Von Cramon, & Engel, 2005). For instance, Burle and colleagues (2008) analyzed partial error trials – i.e. correct trials showing an early activation of the incorrect response – and demonstrated that the amplitude of the ERN correlates to the onset of remediation processes (i.e. the activation of the correct response). The ERN was suppressed once the remedial process has started, indicating that the ERN reflects the activity of an "alarm signal" which lasts until a remediation process takes place. According to this view, the ERN would thus represent a dynamic monitoring process signaling that the correction of the ongoing (wrong) response is needed. Therefore, our new findings showing a reduced ΔERN amplitude in the no-free will group suggest that encouraging people to disbelieve in free will actually influences such an early and dynamic control process during action monitoring.

As described in the introduction, several studies already demonstrated that the ERN is also influenced by the motivational significance of response errors. For instance larger ERNs have been found for response errors associated with a high motivational value, as compared to response errors with a low motivational value (Gehring et al., 1993; Gehring & Willoughby, 2002; Pailing & Segalowitz, 2004; Hajcak, Moser, Yeung, & Simons, 2005). According to this framework one can thus argue that people reading a deterministic message denying free will might reason that their behavior is determined or caused by something else than their mere intentions or will – i.e. their genetic make-up, environmental

factors, or both – and that it is therefore useless to be concerned about the outcome of their own actions, as "it does not matter anyway". This alternative interpretation entails an attenuation of the motivational significance of response outcome – i.e. whether the response is correct or not – and might thus explain why a reduced ΔERN was observed in the no-free will group. It should be noted, however, that a decrease in the motivational significance of response outcomes should probably influence the Pe to a greater extent than the ERN component (Ridderinkhof et al., 2009). Since the ΔPe remained unchanged by the free will manipulation in the current study, an account in terms of reduced motivational significance of the response outcome appears unlikely.

The decrease of the ΔERN amplitude in the no-free will group during the post-manipulation session was found to be proportional to the level of disbelief in this group, further substantiating a link between the ΔERN component and the disbelief in free will. However, the absence of a baseline measurement of the belief in free will is a limitation of the current study, since it was not possible to titrate the actual change in free will belief following the implicit disbelief manipulation. Accordingly, this correlation result should be taken cautiously.

It must be noted that, in contrast to our previous study showing a reduced post-error slowing effect after the anti-free will manipulation (Rigoni et al., 2013), the results of this study do no show a similar post-error adaptation effect. However, this discrepancy may be explained by the fact that calibration blocks meant to match RTs and error rates between the two groups were used in this study, thereby strongly attenuating potential group differences at the behavioral level.

In this study, we show for the first time that undermining the idea that one can determine one's own behavior has direct measurable effects on a very basic and automatic mechanism involved in the online monitoring of (self-generated) actions. The current study does not provide direct evidence for an effect

of this early neurophysiological change (at the level of the ERN) on more complex (social) behaviors. However, independent evidence suggests that complex behaviors may be altered by similar depletion of the belief in free will (for a review, see Rigoni & Brass, 2014), and we reasoned that those behavioral changes might stem, at least in part, from a change in early response monitoring brain processes. The cursory behavior described in literature (Vohs & Schooler, 2008; Baumeister et al., 2009; Alquist et al., 2013), may therefore be the consequence of malfunctioning action control processes, namely reduced error monitoring. As such, our new neurophysiological results add to the existing literature showing that believing or disbelieving in free will is not entirely innocuous, but instead, it may have real life consequences (Stillman et al., 2010).

More generally, the current results lend support to the proposal that abstract beliefs such as free will have measurable consequences onto behavior and cognition, which are however not immediately bound to these beliefs but instead fully orthogonal to them (Vohs & Schooler, 2008; Baumeister et al., 2009; Rigoni & Brass, 2014). As such, our results may also have important implications for a better understanding of ubiquitous break-downs in performance monitoring occurring in real-life situations. Some of these apparently maladaptive error-prone actions or behaviors might very well stem from dynamic changes in the abstract belief system of the participant induced "implicitly" by situational factors (e.g., verbal or non-verbal information present in the immediate environment and somehow challenging the belief in free will), as opposed to lapses of attention, task involvement, effort or motivation per se. While earlier work already clearly showed a reliable influence of the belief in free will on complex social behavior (Vohs & Schooler, 2008; Baumeister et al., 2009), here we show that the personal belief system of an adult participant regarding free will, although abstract, may be liable to

transient changes that can later impact basic cognitive functions during performance monitoring, in a similar manner to what religious beliefs for example may also provoke (Inzlicht & Tullett, 2010).

Acknowledgements

This work was supported by the BELSPO "Interuniversity Poles of Attraction" Program (Grant P7/33) and the Bijzonder Onderzoeksfonds of Ghent University, Belgium.

Author contributions

D. Rigoni and M. Brass developed the study concept. All authors contributed to the study design. Testing and data collection were performed by D. Rigoni. Data analysis was performed by D. Rigoni and G. Pourtois. All authors contributed to the interpretation of the results. D. Rigoni drafted the manuscript and M. Brass and G. Pourtois provided critical revisions. All authors approved the final version of the manuscript for submission.

References

Aarts, K., & Pourtois, G. (2010). Anxiety not only increases, but also alters early error-monitoring functions. *Cognitive, Affective, & Behavioral Neuroscience*, *10*(4), 479-492.

Aarts, K., & Pourtois, G. (2012). Anxiety disrupts the evaluative component of performance monitoring: An ERP study. *Neuropsychologia*, *50*(7), 1286-1296.

Aarts, K., Houwer, J. D., & Pourtois, G. (2012). Evidence for the automatic evaluation of self-generated actions. *Cognition*, *124*(2), 117-127.

Alquist, J. L., Ainsworth, S. E., & Baumeister, R. F. (2013). Determined to conform: disbelief in free will increases conformity. *Journal of Experimental Social Psychology, 49*(1), 80-86.

Baumeister, R. F. (2008). Free will in scientific psychology. *Perspectives on Psychological Science*, *3*(1), 14-19.

Baumeister, R.F., Masicampo, E.J., & DeWall, C.N. (2009). Prosocial benefits of feeling free: Disbelief in free will increases aggression and reduces helpfulness. *Personality and Social Psychology Bulletin, 35*, 260–268.

Bonini, F., Burle, B., Liégeois-Chauvel, C., Régis, J., Chauvel, P., & Vidal, F. (2014). *Science, 343*(6173), 888-891.

Burle, B., Roger, C., Allain, S., Vidal, F., & Hasbroucq, T. (2008). Error negativity does not reflect conflict: a reappraisal of conflict monitoring and anterior cingulate cortex activity. *Journal of Cognitive Neuroscience*, 20(9), 1637-1655.

Crick, F. (1994). *The astonishing hypothesis: The scientific search for the soul*. New York, NY: Touchstone.

de Bruijn, E. R., Sabbe, B. G., Hulstijn, W., Ruigt, G. S., & Verkes, R. J. (2006). Effects of antipsychotic and antidepressant drugs on action monitoring in healthy volunteers. *Brain Research*, *1105*(1), 122-129.

Debener, S., Ullsperger, M., Siegel, M., Fiehler, K., Von Cramon, D. Y., & Engel, A. K. (2005). Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. *Journal of Neuroscience, 25*(50), 11730-11737.

Endrass, T., Klawohn, J., Preuss, J., & Kathmann, N. (2012). Temporospatial dissociation of Pe subcomponents for perceived and unperceived errors. *Frontiers in Human Neuroscience, 6*.

Endrass, T., Reuter, B., & Kathmann, N. (2007). ERP correlates of conscious error recognition: aware and unaware errors in an antisaccade task. *European Journal of Neuroscience, 26*(6), 1714-1720.

Falkenstein, M., Hohnsbein, J., Hoormann, J., & Blanke, L. (1990). Effects of errors in choice reaction tasks on the ERP under focused and divided attention. *Psychophysiological Brain Research*, *1*, 192-195.

Falkenstein, M., Hohnsbein, J., Hoormann, J., & Blanke, L. (1991). Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. *Electroencephalography and Clinical Neurophysiology*, *78*(6), 447-455.

Falkenstein, M., Hoormann, J., Christ, S., & Hohnsbein, J. (2000). ERP components on reaction errors and their functional significance: A tutorial. *Biological Psychology, 51*, 87–107.

Gehring, W. J., & Willoughby, A. R. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. *Science, 295*(5563), 2279-2282.

Gehring, W. J., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1990). The error-related negativity: an event-related brain potential accompanying errors. *Psychophysiology*, *27*(4), S34.

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.

Gehring, W. J., Himle, J., & Nisenson, L. G. (2000). Action-monitoring dysfunction in obsessive-compulsive disorder. *Psychological Science*, *11*(1), 1-6.

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.

Hajcak, G., Moser, J. S., Yeung, N., & Simons, R. F. (2005). On the ERN and the significance of errors. *Psychophysiology, 42*(2), 151-160.

Holroyd, C. B., & Coles, M. G. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*(4), 679.

Inzlicht, M., & Tullett, A. M. (2010). Reflecting on God Religious Primes Can Reduce Neurophysiological Response to Errors. *Psychological Science*, *21*(8), 1184-1190.

Inzlicht, M., McGregor, I., Hirsh, J. B., & Nash, K. (2009). Neural markers of religious conviction. *Psychological Science*, *20*(3), 385-392.

Johannes, S., Wieringa, B. M., Nager, W., Dengler, R., & Münte, T. F. (2001). Oxazepam alters action monitoring. *Psychopharmacology*, *155*(1), 100-106.

Koban, L., Brass, M., Lynn, M. T., & Pourtois, G. (2012). Placebo Analgesia Affects Brain Correlates of Error Processing. *PloS One*, *7*(11), e49784.

Koban, L., Pourtois, G., Vocat, R., & Vuilleumier, P. (2010). When your errors make me lose or win: Event-related potentials to observed errors of cooperators and competitors. *Social Neuroscience*, *5*(4), 360-374.

Lynn, M. T., Van Dessel, P., & Brass, M. (2013). The influence of high-level beliefs on self-regulatory engagement: Evidence from thermal pain stimulation. *Frontiers in Psychology*, *4*.

Nieuwenhuis, S., Ridderinkhof, K. R., Blom, J., Band, G. P., & Kok, A. (2001). Error-related brain potentials are differentially related to awareness of response errors: Evidence from an antisaccade task. *Psychophysiology*, *38*(5), 752-760.

O'Connell, R. G., Dockree, P. M., Bellgrove, M. A., Turin, A., Ward, S., Foxe, J. J., & Robertson, I. H. (2009). Two types of action error: electrophysiological evidence for separable inhibitory and sustained attention neural mechanisms producing error on go/no-go tasks. *Journal of Cognitive Neuroscience, 21*(1), 93-104.

Olvet, D. M., & Hajcak, G. (2009). The stability of error-related brain activity with increasing trials. *Psychophysiology*, *46*(5), 957-961.

Overbeek T. J., Nieuwnhuis, S., & Ridderinkhof, K. R. (2005). Dissociable components of error processing: On the functional significance of the Pe vis-à-vis the ERN/Ne. *Journal of Psychophysiology, 19*(4), 319.

Pailing, P. E., & Segalowitz, S. J. (2004). The error-related negativity as a state and trait measure: Motivation, personality, and ERPs in response to errors. *Psychophysiology, 41*(1), 84-95.

Paulhus, D. L., & Carey, J. M. (2011). The FAD–Plus: Measuring lay beliefs regarding Free Will and related constructs. *Journal of Personality Assessment*, *93*(1), 96-104.

Pourtois, G. (2010). Anxiety not only increases, but also alters early error-monitoring functions. *Cognitive, Affective, & Behavioral Neuroscience*, *10*(4), 479-492.

Pourtois, G. (2011). Early error detection predicted by reduced pre-response control process: An ERP topographic mapping study. *Brain Topography*, *23*(4), 403-422.

Ridderinkhof, K. R., Ramautar, J. R., & Wijnen, J. G. (2009). To PE or not to PE: A P3-like ERP component reflecting the processing of response errors. *Psychophysiology*, *46*(3), 531-538.

Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science*, *306*(5695), 443-447.

Rigoni, D., & Brass, M. (2014). From intentions to neurons: social and neural consequences of disbelieving in free will. *Topoi*, 1-8.

Rigoni, D., Kühn, S., Gaudino, G., Sartori, G., & Brass, M. (2012). Reducing self-control by weakening belief in free will. *Consciousness and Cognition*, *21*(3), 1482-1490.

Rigoni, D., Kühn, S., Sartori, G., & Brass, M. (2011). Inducing Disbelief in Free Will Alters Brain Correlates of Preconscious Motor Preparation The Brain Minds Whether We Believe in Free Will or Not. *Psychological Science*, *22*(5), 613-618.

Rigoni, D., Wilquin, H., Brass, M., & Burle, B. (2013). When errors do not matter: Weakening belief in intentional control impairs cognitive reaction to errors. *Cognition*, *127*(2), 264-269.

Ruchsow, M., Grön, G., Reuter, K., Spitzer, M., Hermle, L., & Kiefer, M. (2005). Error-related brain activity in patients with obsessive-compulsive disorder and in healthy controls. Journal of *Psychophysiology, 19*(4), 298.

Ryle, G. (1949). *The concept of mind*. University of Chicago Press.

Schneider, W., Eshman, A., & Zuccolotto, A. (2002). E-prime 2.0 user's guide. Psychology Software Tools Inc., Pittsburgh.

Skinner, B. F. (2002). *Beyond freedom and dignity*. Indianapolis: Hackett Publishing Company.

Steinhauser, M., & Yeung, N. (2010). Decision processes in human performance monitoring. *Journal of Neuroscience, 30*(46), 15643-15653.

Stillman, T. F., Baumeister, R. F., Vohs, K. D., Lambert, N. M., Fincham, F. D., & Brewer, L. E. (2010). Personal philosophy and personnel achievement: Belief in free will predicts better job performance. *Social Psychological and Personality Science*, *1*(1), 43-50.

Van Veen, V., & Carter, C. S. (2002). The timing of action-monitoring processes in the anterior cingulate cortex. *Journal of Cognitive Neuroscience, 14*(4), 593-602.

Vocat, R., Pourtois, G., & Vuilleumier, P. (2008). Unavoidable errors: a spatio-temporal analysis of time-course and neural sources of evoked potentials associated with error processing in a speeded task. *Neuropsychologia*, *46*(10), 2545-2555.

Vocat, R., Pourtois, G., & Vuilleumier, P. (2011). Parametric modulation of error-related ERP components by the magnitude of visuo-motor mismatch. *Neuropsychologia*, *49*(3), 360-367.

Vohs, K.D., & Schooler, J.W. (2008). The value of believing in free will: Encouraging a belief in determinism increases cheating. *Psychological Science, 19*, 49–54.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063.

Weinberg, A., Riesel, A., & Hajcak, G. (2012). Integrating multiple perspectives on error-related brain activity: the ERN as a neural indicator of trait defensive reactivity. *Motivation and Emotion*, *36*(1), 84-100.

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: conflict monitoring and the error-related negativity. *Psychological Review*, *111*(4), 931.