Psychophysiology, 53 (2016), 1870–1877. Wiley Periodicals, Inc. Printed in the USA.

DOI: 10.1111/psyp.12771

The electrophysiological signature of deliberate rule violations

ROLAND PFISTER,[a] ROBERT WIRTH,[a] KATHARINA A. SCHWARZ,[a] ANNA FOERSTER,[a] MARCO STEINHAUSER,[b] AND WILFRIED KUNDE[a]

[a]Department of Psychology, Julius Maximilians University of Würzburg, Würzburg, Germany
[b]Department of Psychology, Catholic University of Eichstätt-Ingolstadt, Eichstätt, Germany

Abstract

Humans follow rules by default, and violating even simple rules induces cognitive conflict for the rule breaker. Previous studies revealed this conflict in various behavioral measures, including response times and movement trajectories. Based on these experiments, we investigated the electrophysiological signature of deliberately violating a simple stimulus-response mapping rule. Such rule violations were characterized by a delayed and attenuated P300 component when evaluating a rule-relevant stimulus, most likely reflecting increased response complexity. This parietal attenuation was followed by a frontal positivity for rule violations relative to correct response trials. Together, these results reinforce previous findings on the need to inhibit automatic S-R translation when committing a rule violation, and they point toward additional factors involved in rule violation. Candidate processes such as negative emotional responses and increased monitoring should be targeted by future investigations.

Descriptors: Rule violation, Nonconformity, P300

Deliberately violating a rule does not come easily: Not only do humans tend to conform to rules by default (Asch, 1956; Cialdini & Goldstein, 2004), but empirical studies are also beginning to uncover peculiarities of rule violation behavior from a perspective of the individual who violates a rule.

A key finding of these studies is that rule violations induce cognitive conflict for the rule breaker. This conflict has been shown in terms of prolonged response times for rule violations as compared to rule-based responses and in a robust signature of rule violations on movement trajectories (Pfister, Wirth, Schwarz, Steinhauser, & Kunde, 2016; Wirth, Pfister, Foerster, Huestegge, & Kunde, 2016). Participants in these studies classified stimuli according to an arbitrary S-R mapping rule that specified a movement either to the upper left or to the upper right of a screen. Moreover, participants indicated before each trial whether they intended to perform according to the mapping rule or whether they intended to violate the mapping rule and deliberately aim for the "wrong" target (Pfister et al., 2016). In case of correct, rule-based responses, movements followed a relatively direct path to their target destinations. In case of rule violations, however, movement trajectories were markedly attracted toward the opposite target that would correspond to a rule-based response. Similar results emerged when participants were prompted whether to follow or whether to violate the rule (Wirth, Pfister et al., 2016). These rule-based action tendencies indicate that rule representations cannot be fully suppressed, even after the deliberate decision for a rule violation has already been made.

With the present experiment, we aimed at gathering converging evidence for this interpretation by examining the electrophysiological signature of deliberate rule violations. This approach fills a gap in the current literature, because, to the best of our knowledge, thorough electrophysiological investigations of deliberate rule violations have not yet been reported. This stands in stark contrast to investigations of unintended failures to follow a rule (i.e., unintended errors). The electrophysiological signature of such errors has been subject to empirical investigation for decades, and this research has established the error-related negativity (ERN) or error negativity (N_E) as a robust marker of error processing (Falkenstein, Hohnsbein, & Hoormann, 1990; Gehring, Goss, Coles, Meyer, & Donchin, 1993; Renault, Ragot, & Lesevre, 1980; for reviews, see Falkenstein, Hoormann, Christ, & Hohnsbein, 2000; Gehring, Liu, Orr, & Carp, 2012).[1] A prominent theoretical account for this ERP component links the ERN to the concept of prediction errors (the reinforcement learning approach; Holroyd & Coles, 2002; Holroyd, Yeung, Coles, & Cohen, 2005). In this view, the ERN signals that events such as the actual response deviate from the actor's expectations, possibly involving the assessment of an error's significance (Hajcak, Moser, Yeung, & Simons, 2005; Maier & Steinhauser, 2013; Maier, Steinhauser, & Hübner, 2008). Because the mentioned deviation from the actor's expectations is not necessarily present for rule violations (i.e., when deliberately behaving counter to a rule), rule violations should not be accompanied by ERN-like

The experiment was conducted as partial fulfillment of the first author's Ph.D. thesis.

Address correspondence to: roland.pfister@psychologie.uni-wuerzburg.de

1. A second ERP component that is reliably associated with error processing is the error positivity (P_E; Overbeek, Nieuwenhuis & Ridderinkhof, 2005; Steinhauser & Yeung, 2010). The P_E has been related to subjective confidence that an error has just occurred, with higher P_E amplitudes signalling higher certainty (Boldt & Yeung, 2015).

PSYCHOPHYSIOLOGY

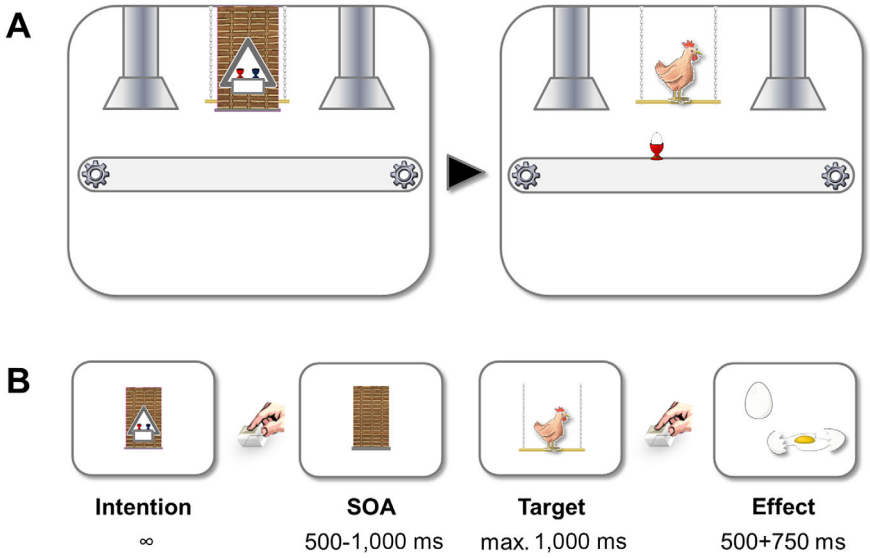

Figure 1. Experimental design and procedure. A: Participants first announced whether they intended to follow the factory rules and perform a correct, rule-based response, or whether they intended to violate these rules and commit an error by intention. Following the factory rules implied placing an egg cup in a position to catch a falling egg by pressing a left or a right response button, whereas violating the rules implied placing the cup in a position where the egg would not be caught and destroyed. B: The sequence of trial events was as follows. Participants announced their intention for the upcoming trial at leisure, and this announcement was followed by a variable SOA. The critical events for all analyses were target onset and the corresponding response that was registered for up to 1,000 ms after target onset. Responses as well as response omissions were followed by an animated effect that illustrated the egg's fate.

responses (see Stemmer, Witzke, & Schönle, 2001, for a preliminary test of this hypothesis).

But, if rule violations cannot be expected to differ from correct responses in terms of ERN-like potentials, is there still reason to expect a notable electrophysiological signature? This seems to be the case: Because rule violations entail an inhibition of automatic response tendencies as described above, they should delay and attenuate the P300 component of the stimulus-locked ERP. Current theories on the P300 suggest that, among other processes, P300 reflects the translation of stimuli to associated responses (Nieuwenhuis, Aston-Jones, & Cohen, 2005; Verleger, 1997; Verleger, Jaśkowski, & Wascher, 2005).[2] The P300 amplitude, according to this view, depends on how strongly stimuli and following responses are coupled. Evidence for this claim comes from a study in which participants performed a simple classification task (Roche & O'Mara, 2003). Participants first trained the mapping of a particular stimulus to the corresponding response. In a following test block, this particular stimulus triggered an enlarged P300 response as compared to the remaining stimuli. This view also yields direct predictions regarding the electrophysiological signature of rule violations, because rule violations can be construed as the very opposite of responding with a canonical response to a stimulus. Accordingly, rule violations should be characterized by an attenuated amplitude of the P300 component, as compared to normal, correct responses. Preliminary evidence for this claim comes from studies on active lying that is indeed associated with attenuated P300 responses as compared to truthful responding (Johnson, Barnhardt, & Zhu, 2003, 2005; Pfister, Foerster, & Kunde, 2014). In the present study, we aim to test this hypothesis for rule violations.

We tested the P300 attenuation hypothesis in a variant of previous experimental designs (based on Experiment 1 in Pfister et al., 2016). Participants were instructed with a simple stimulus-response mapping rule that mapped two stimuli to either a left or a right key press response. They were again asked at the beginning of each trial whether they intended to follow the mapping rule for the upcoming task or whether they intended to violate the rule. We further introduced several changes to the previous setup to optimize the design for electrophysiological recordings. As noted above, we had participants perform simple key press responses instead of the mouse or finger movement tasks used in previous studies to minimize noise in the ERP data due to overt movements. Second, we used a larger number of trials to ensure sufficient statistical power for the ERP analyses. Third, we adapted the framing of the experiment to ensure high motivation of the participants throughout the session. The present experiment was designed as a computer game in which participants operated an "egg factory" (Figure 1). They placed egg cups under a chicken that was going to lay an egg at either a left or a right location and, to ensure smooth operation of the factory, factory rules held that the egg cup had to be placed under the chicken's rear because wrong placements would destroy the egg. P300 amplitudes were examined stimulus-locked to the onset of the chicken stimulus, and we expected attenuated P300 amplitudes for rule violations as compared to rule-based responses.

Method

Participants

Sixteen volunteers were recruited and gave informed consent for participation. The data of two participants was replaced due to technical difficulties. The participants of the final sample (mean age: 22.1 years, 14 female, 1 left-handed, 1 ambidextrous) reported normal or corrected-to-normal vision and received either course credit or monetary compensation. The experiment was conducted according to the Declaration of Helsinki and the guidelines of the local ethics committee.

2. An alternative term for the ERP component described here is P3b, which is distinct from the stimulus-driven processes related to detection and attention as indexed by P3a (e.g., Polich, 2007).

Apparatus and Stimuli

Participants sat in front of a 17" monitor at an eye-screen distance of about 60 cm, and they responded with their left and right index finger on the *A* and the *hash* (#) key of a standard German QWERTZ keyboard. Each key was marked with a small sticker.

The task was embedded in a gamelike setting that made the participants operate an egg factory as sketched in Figure 1. The screen showed a conveyor belt spanning the horizontal midline right below the screen center (19.8 cm × 2.0 cm). Two tubes were displayed to the left and right of the screen, extending from the virtual ceiling (6.8 cm × 9.1 cm each). The upper center of the screen either featured a shutter (6.0 cm × 9.4 cm) or the image of a chicken looking to the left or to the right (approximately 4.7 cm × 6.3 cm). Further stimuli that appeared on screen during a trial were a warning sign (4.4 cm × 4.7 cm) that was superimposed on the shutter, an egg (1.3 cm × 1.5 cm), as well as a red and a blue egg cup (1.5 cm × 1.7 cm).

EEG Setup

We recorded EEG data throughout the session by means of a BrainVision QuickAmp amplifier with 32 active Ag/AgCl electrodes (actiCAP; Brain Products, Germany) that were placed at the following positions according to the extended 10-20 system: Fp1, Fp2, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, TP9, CP5, CP1, CP2, CP6, TP10, P7, P3, Pz, P4, P8, PO9, O1, Oz, O2, PO10. We used average reference and recorded the EEG signal at a sampling rate of 500 Hz, low-pass filtered at 100 Hz (low cutoff: DC). Impedances were below 10 kΩ at the start of the experiment.

Ocular movements were recorded with passive bipolar electrodes on the outer canthi of both eyes and above and below the left eye (electrooculogram, EOG). Participants were encouraged to reduce eye movements and blinks, especially between target onset and response, to minimize artifacts in the EEG data.

Instructions

Participants received verbal instructions, supported by exemplar stimuli on screen. They were first informed that, to ensure smooth operation of the factory, they would have to place a cup to the left or to the right of the chicken in each trial by pressing the left or the right key. Correct cup placement ensured that an egg could fall down into the cup, be carried away on the belt, and absorbed in the tubes. Wrong cup placement resulted in the egg hitting the conveyor belt and shattering to pieces. Both effects were demonstrated as on-screen animations.

Afterward, participants were introduced to the crucial element of the task: a compliance prompt asking whether participants intended to comply with the factory rules in a given trial. This compliance prompt consisted of the warning sign that showed two cups, a red and a blue one standing next to each other. Below these cups appeared the letters *R* (for German "richtig," correct) and *F* (for German "falsch," wrong). Participants pressed the left or the right key to indicate whether they wanted to stick to the factory regulations and perform correctly, or whether they wanted to violate them and commit an error by intention. Each intention was indicated by a constant cup color (red vs. blue), and this color-intention mapping was counterbalanced across participants.

Procedure

Participants completed a training block of 48 trials and 10 experimental blocks of 48 trials each (see Figure 1 for an overview of the trial procedure). Individual sessions lasted between 1.5 h and 2 h including preparation of the electrodes.

Trials started with the compliance prompt that stayed on screen until the participant gave the corresponding response to announce their intention for the upcoming stimulus. Afterward, the warning sign disappeared, and the empty shutter was displayed for a variable stimulus onset asynchrony (SOA; 500 ms vs. 750 ms vs. 1,000 ms). Then, the shutter disappeared and gave way to a chicken looking to the left or right (the target stimulus). The chicken waited for 1,000 ms, then laid an egg that appeared below the chicken's rear (i.e., the egg appeared to the left when the chicken looked to the right, and vice versa). This interval served as response deadline for the participant who had to place an egg cup to the left or right by pressing the corresponding key.

The cup appeared immediately after the response, and its color depended on the participant's announced intention. Whereas the events up to this point were static and discrete to allow for undisturbed EEG recordings, the remaining procedure after the chicken had laid its egg was animated and showed the egg falling down and either landing safely in the cup (if the cup stood below the chicken's rear) or shattering on the conveyor belt if this was not the case (duration of the animation: 500 ms). Then, the belt started moving and transported the egg to the nearest tube, which started to absorb the contents of the belt (750 ms). Finally, the shutter went down again, and the empty factory was displayed for an intertrial interval of 1,000 ms.

Data Treatment

Unintended errors in terms of wrong key presses occurred only rarely (2.0%), as did accidentally correct responses (after having announced they would violate the factory rules; 2.0%). Procedural errors—double key presses for intention or target response, responses during the SOA or during feedback—occurred in an additional 3.9% of the trials. These data were excluded from all reaction time (RT) and ERP analyses, as were trials following such errors. Furthermore, we corrected for outliers by excluding trials with RTs that deviated by more than 2.5 standard deviations from the corresponding cell mean, calculated separately for each participant and condition (2.8%).

EEG data were preprocessed via FieldTrip (Oostenveld, Fries, Maris, & Schoffelen, 2011) and custom MATLAB scripts (for a different but complementary approach to the data, see Experiment 7 in Pfister, 2013). We filtered the signal with a 0.1 Hz high-pass filter, a 70 Hz low-pass filter, and a [47.5 Hz; 52.5 Hz] band-stop filter, and segmented the data into target-locked epochs around the onset of the chicken stimulus (200 ms prestimulus to 1,000 ms poststimulus). For additional exploratory analyses, we also created response-locked epochs around the corresponding response (600 ms preresponse to 600 ms postresponse). Trials with artifacts were eliminated by using the FieldTrip outlier detection mechanism based on z scores ($z = 20$). Ocular artifacts were addressed via independent component analysis (ICA; Makeig, Bell, Jung, & Sejnowski, 1996), and we removed components that correlated with at least one EOG channel ($r > .40$; Flexer, Bauer, Pripfl, & Dorffner, 2005). Following this ICA, we performed a baseline correction with a baseline period of 200 ms before the event of interest until event onset.

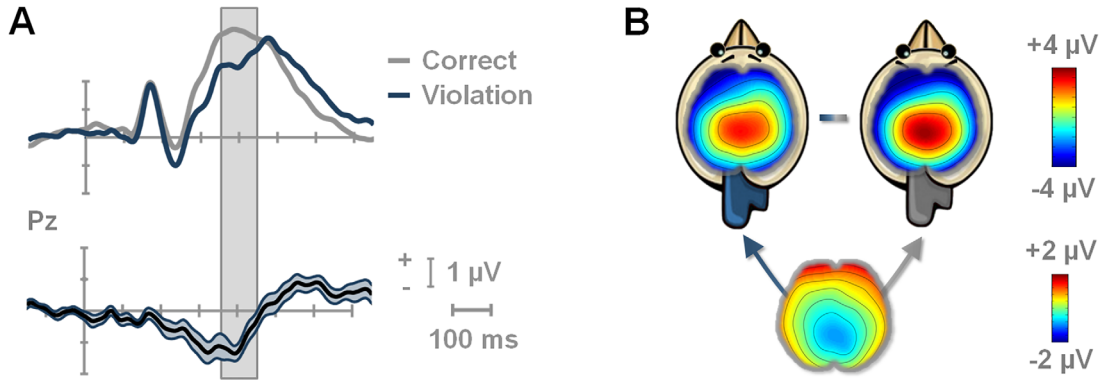

Figure 2. Results of the stimulus-locked ERP analysis at Pz prior to RT matching of both conditions. A: ERP waveforms for correct response trials (light gray line) and for violation trials (dark blue line) and the corresponding difference wave ($voltage_{Violation} - voltage_{Correct}$) ± 1 standard error of this paired difference, computed separately for each data point (shaded area; see Pfister & Janczyk, 2013). B: Mean voltage distributions across the scalp in a 100-ms time window centered on the mean time to peak amplitude across all participants and conditions (412 ms). Correct response trials are plotted to the right (gray mock heads), whereas violation trials are plotted to the left (blue mock heads). The distribution of the difference wave is plotted in between the heads ($voltage_{Violation} - voltage_{Correct}$).

Results

Behavioral Data

Participants showed a consistent preference for adhering to the factory rules (63.4%), and this preference deviated from chance, $t(15) = 3.85$, $p = .001$, $d = 0.96$ (with $d = \frac{t}{\sqrt{n}}$). Correct responses were also faster than rule violations (461 ms vs. 492 ms), $t(15) = 3.44$, $p = .004$, $d = 0.86$. We further computed the intraindividual standard deviations of the participants' RT for each condition to ensure that potential effects on the P300 amplitude did not derive from differences in variability (Ramchurn, de Fockert, Mason, Darling, & Bunce, 2014). Mean standard deviations only showed a small numeric difference (correct responses: $S\bar{D} = 94$ ms; rule violations: $S\bar{D} = 97$ ms) and did not differ significantly, $t(15) = 0.54$, $p = .600$, $d = 0.14$.

Target-Locked ERPs

For analyzing target-locked ERPs, we had initially planned to focus on the electrode site Pz to evaluate the P300 response that typically comes with a centroparietal scalp distribution (Polich, 2007). Figure 2 shows the grand-averaged waveform for correct trials and violation trials at this electrode, accompanied by the corresponding difference wave ($voltage_{Violation} - voltage_{Correct}$) and voltage distributions across the scalp.

Figure 2A indicates that the P300 component was indeed attenuated in violation trials relative to correct trials in a time window of about 300 ms to 450 ms poststimulus. In this time window, the P300 responses of both conditions were characterized by a centroparietal scalp distribution, and the same applied to the difference wave as shown in Figure 2B. Peak times (T_{Peak}) were computed as the time to the most positive peak on the average ERP of each participant and condition. This analysis showed the P300 response for violation trials to peak later than in correct trials (432 ms vs. 390 ms), $t(15) = 2.38$, $p = .031$, $d = 0.60$, and individual effects on times to peak amplitude ($\Delta T_{Peak} = T_{Peak|Violation} - T_{Peak|Correct}$) were correlated with the effects on RTs ($\Delta RT = RT_{Violation} - RT_{Correct}$) across participants, $r = .55$, $p = .028$.

Because these observations suggested that the effects on P300 might mainly mirror the behavioral results in terms of prolonged response times in the violation condition, we decided to rerun the analyses with an RT-matched subset of the trials. To this end, we scanned the data for one correct trial that matched the RT of each violation trial as closely as possible (note that this procedure also equates signal-to-noise ratios across conditions). Because this procedure indicated that RT matching was not possible for the slowest violation responses for individual participants, we further trimmed the RT distribution by omitting the 10% longest RTs prior to the matching procedure. Mean RTs of the matched data were 483 ms for correct responses and 482 ms for rule violations, $t(15) = 0.20$, $p = .841$, $d = 0.05$. To arrive at a clearer picture of the effects at parietal sites, we further extended the analyses of the P300 amplitude to the electrodes P3 and P4.[3]

Figure 3A shows the resulting ERPs for the RT-matched data at the three parietal electrode sites. Peak latencies at Pz were more similar than in the unmatched data, though a numerical difference still remained (439 ms for violation trials vs. 418 ms for correct trials), $t(15) = 1.47$, $p = .162$, $d = 0.37$. Also, the correlation of ΔTAmp and ΔRT was no longer significant, $r = -.16$, $p = .559$ (ΔRT still refers to the data of the unmatched conditions because ΔRT for the RT-matched data is almost equal to zero for trivial reasons). Peak amplitudes were addressed with a 2 (Condition: correct vs. violation) × 3 (Electrode: P3 vs. Pz vs. P4) repeated measures analysis of variance (ANOVA) on the peak amplitude in a time window starting 200 ms poststimulus until the end of the trial (1,000 ms poststimulus). This analysis yielded a significant main effect of condition, $F(1,15) = 6.58$, $p = .022$, $\eta_p^2 = .30$, with higher peak amplitudes for correct trials than for violation trials (4.85 µV vs. 4.34 µV). A significant main effect of electrode, $F(1,15) = 5.03$, $p = .013$, $\eta_p^2 = .25$, reflected overall higher P300 amplitudes for Pz (5.02 µV) and P4 (4.90 µV) than for P3 (3.87 µV). The interaction was not significant, $F < 1$.

Figure 3B shows the results of additional exploratory analyses for the frontocentral midline electrodes Fz and Cz. These analyses indicated a more pronounced frontal positivity for rule violations to emerge from around 400 ms poststimulus to 650 ms poststimulus (see the difference waves ± standard errors of the difference, as plotted in Figure 3B). This impression was further supported by analyses of the time course of the entire scalp topography as shown in Figure 3C. Rule violations came with an attenuated amplitude at

3. This analysis was prompted by an anonymous reviewer.

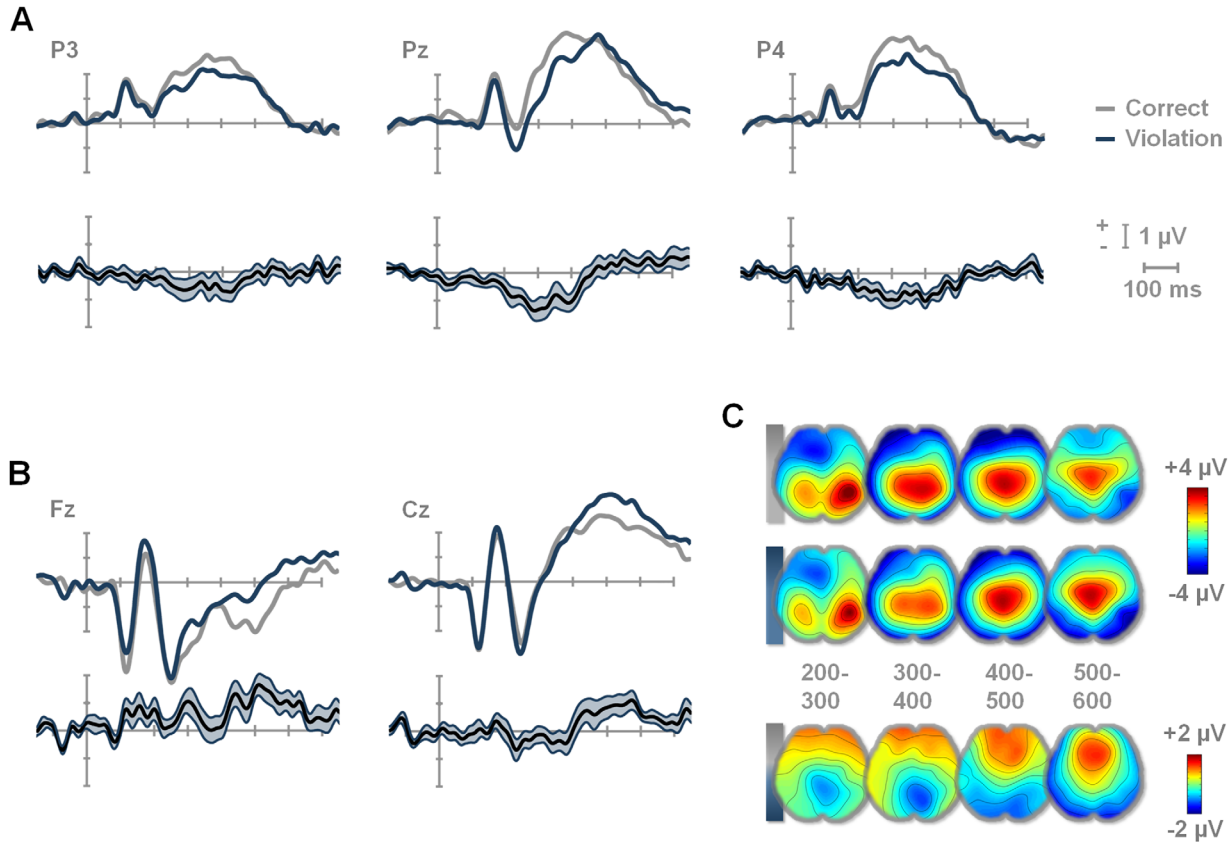

Figure 3. Results of the stimulus-locked ERP for the RT-matched data. A: ERP waveforms at parietal electrode sites for correct response trials (light gray line) and for violation trials (dark blue line) and the corresponding difference wave ($\text{voltage}_{\text{Violation}} - \text{voltage}_{\text{Correct}}$) ± 1 standard error of this paired difference, computed separately for each data point (shaded area; Pfister & Janczyk, 2013). B: ERP waveforms at frontocentral electrode sites, color-coded as in (A). C: Mean voltage distributions across the scalp in steps of 100 ms starting 200 ms poststimulus. Correct response trials are plotted in the upper row (indicated by a gray bar), whereas violation trials are plotted in the middle row (indicated by a blue bar). The distribution of the difference wave is plotted in the lower row ($\text{voltage}_{\text{Violation}} - \text{voltage}_{\text{Correct}}$).

parietal electrode sites that reached its maximum around 300–400 ms poststimulus, and this initial attenuation was followed by a frontocentral positivity around 500–600 ms poststimulus. Visual inspection of Figure 3B further indicated a slow positive drift at electrode Fz that would overshadow any P300-related effects in analyses of peak amplitudes. We therefore conducted an additional ANOVA on mean amplitudes rather than peak amplitudes across the midline electrodes Fz, Cz, and Pz. Mean amplitudes were computed for a time window of 100 ms, centered on the mean peak latencies at Pz of both conditions (428 ± 50 ms). Mirroring the above observations, this analysis did not yield a main effect of condition, $F(1,15) = 0.53$, $p = .477$, $\eta_p^2 = 0.03$. Instead, it showed a significant main effect of electrode, $F(2,30) = 25.97$, $p < .001$ ($\varepsilon = .61$), $\eta_p^2 = 0.63$, that was qualified by a significant interaction, $F(2,30) = 4.50$, $p = .044$ ($\varepsilon = .58$), $\eta_p^2 = 0.23$ (Greenhouse-Geisser corrected for violations of sphericity). Repeating the analysis of mean rather than peak amplitudes for the parietal electrodes P3, Pz, and P4 again showed significant main effects of condition, $F(1,15) = 5.87$, $p = .028$, $\eta_p^2 = 0.28$, and electrode, $F(2,30) = 5.20$, $p = .012$, $\eta_p^2 = 0.26$ ($F < 1$ for the interaction).

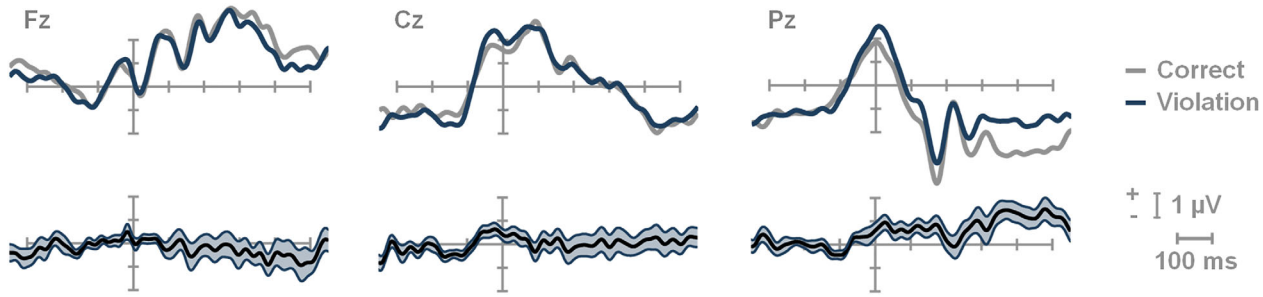

Figure 4. Response-locked ERPs for correct responses and rule violations at midline electrode sites (Fz, Cz, Pz). Upper plot: ERPs for correct responses (light gray line) and rule violations (dark blue line). Lower plot: difference waves ($\text{voltage}_{\text{Violation}} - \text{voltage}_{\text{Correct}}$) ± 1 standard error of these paired differences, computed separately for each data point (shaded area; Pfister & Janczyk, 2013).

Response-Locked ERPs

Figure 4 shows the response-locked ERP of rule violations and correct responses at the three midline electrodes Fz, Cz, and Pz. Differences between rule violations and rule-based responses were small and rather unreliable across the entire epoch, except for a slow-wave drift at Pz that emerged about 300 ms after the response.

Discussion

In this experiment, we tested the hypothesis of a delayed and attenuated P300 component for deliberate rule violations as compared to rule-based responding. Our results confirmed that P300 is delayed for rule violations and that this delay is related to the increased RTs for rule violations as compared to correct responses as observed in previous studies. The data further yielded an attenuated P300 response at parietal electrode sites as predicted.

The delayed and attenuated P300 response corroborates previous arguments that described rule violations as a two-stage process with an initial activation of rule-based response tendencies that have to be inhibited in order to successfully violate a rule (Wirth, Pfister et al., 2016). This view of rule violation behavior posits that rule representations are necessarily activated when violating a rule, and that the corresponding action has to be derived from this rule representation during each instance of rule violation. Such a two-stage process parallels findings on the cognitive mechanisms underlying dishonesty (Debey, De Houwer, & Verschuere, 2014). In this study, participants responded honestly or dishonestly to closed questions (yes/no questions) via button presses. Each question was accompanied by a distractor that could either match the honest response (e.g., *yes* if the honest answer was affirmative) or match the dishonest response (e.g., *no* if the honest answer was affirmative). This study did not only yield typical effects in terms of prolonged RTs for dishonest as compared to honest responses (e.g., Debey, Liefooghe, De Houwer, & Verschuere, 2015; Foerster, Wirth, Kunde, & Pfister, 2016; Spence et al., 2001; Walczyk, Roper, Seemann, Humphrey, 2003), but also a telling impact of the distractor stimuli: Distractors that corresponded to the honest response speeded up both honest and dishonest responses alike, as compared to distractors that corresponded to the dishonest response. These findings lend strong support to the assumption that dishonest responses are derived from an initially activated honest response rather than retrieved directly from memory. The present findings of delayed and attenuated P300 responses thus indicate that similar processes are at work for rule violation responses.

The present results further corroborate theories that relate the P300 component to stimulus-response translation (Nieuwenhuis et al., 2005; Verleger, 1997; Verleger et al., 2005). These theories have typically been discussed with regard to settings in which some stimuli are more closely related to an associated response than other stimuli, to probe for differences in P300 amplitude (Roche & O'Mara, 2003; see also Keller et al., 2006; Waszak et al., 2005). The present experiment, by contrast, yielded differential P300 responses for one and the same stimulus depending on the actor's current intentions. When a participant was going to respond based on the instructed mapping rule, P300 responses occurred earlier and with higher amplitude as compared to a situation in which he or she was going to violate the mapping rule (for corresponding findings on dishonesty, see Johnson et al., 2003, 2005; Pfister et al., 2014).

While the present experiment was inspired by stimulus-response theories of the P300 component (Verleger et al., 2005), our findings do not contradict the context updating hypothesis, the major alternative theory concerning the functional significance of the P300 component (Donchin, 1981; Donchin & Coles, 1988; Polich, 2007). According to this framework, the P300 component reflects memory processes that guide behavioral decisions (but see Verleger, 2008). In line with this assumption, P300 amplitudes were found to be reduced during memory load with task-unrelated information (e.g., Kotchoubey, Jordan, Grözinger, & Westphal, 1996; Pratt, Willoughby, & Swick, 2011) and, similarly, P300 was shown to be affected by response complexity (prolonged latencies in more complex tasks; Hoffman, Simons, & Houck, 1983; see also Kok, 2001). This view, however, suggests a different mechanism to underlie the observed effects on P300. Rather than differences in stimulus-response retrieval, this view assumes differences in resource allocation as observed in studies on dual tasking (Isreal, Chesney, Wickens, & Donchin, 1980; Kramer & Strayer, 1988; Wickens, Kramer, Vanasse, & Donchin, 1983). Diminished attentional resources during rule violation, in turn, might affect different processes from decision making to motor preparation. The present data are fully compatible with explanations in terms of both, response retrieval and resource allocation, and disentangling these explanations would be a promising goal for future studies on the electrophysiology of deliberate rule violations.

The initial attenuation of the ERP at parietal sites was followed by a more pronounced positivity for rule violations than for correct response trials, and this positivity occurred especially at frontal electrode sites. Although this effect extends into the postresponse period, it is clearly not elicited by the response. This is suggested by the absence of such an effect in the response-locked analysis, which might reflect that it starts prior to the response and is thus eliminated by baseline correction. An effect on these late positive potentials (LPPs) was not predicted beforehand, and the mentioned findings should therefore be interpreted with caution. Still, similar effects on anterior LPPs have been observed in a number of studies on socioemotional processing such as processing of in-group/outgroup information or affective states (Cunningham, Espinet, DeYoung, & Zelazo, 2005; Gable & Harmon-Jones, 2013; Hurtado, Haye, González, Manes, & Ibáñez, 2009). Even though processing of emotional stimuli has been found to affect mainly parietal sites (Gable & Harmon-Jones, 2010; Hajcak & Olvet, 2008; for a review, see Hajcak, MacNamara, & Olvet, 2010), linking this finding to affective states seems likely given the results of a previous behavioral study (Wirth, Foerster, Rendel, Kunde, & Pfister, 2016). Participants in this study either performed rule-based responses or rule violations, and each response was followed by a word classification task in which participants indicated whether a noun was either positive or negative. Rule violations facilitated the classification of negative information relative to rule-based responding, suggesting that rule violations automatically elicit slightly negative affective states (for similar findings for cognitive conflict and errors, see Aarts, De Houwer, & Pourtois, 2012; Fritz & Dreisbach, 2013; Hajcak & Foti, 2008). This study further indicated rule violations to prime authority-related concepts that could further contribute to the observed anterior LPPs (for a related model on prefrontal contributions to evaluative processing, see Cunningham & Zelazo, 2007).

Our analyses of the response-locked ERP further suggest that deliberate rule violations do not give rise to ERN-like potentials that have been observed for unintended errors. Even though rule violations, per definition, are the opposite of responding correctly

(as are errors), what counts for the ERN "is not what is correct or an error in the eyes of the experimenter, but rather what is deemed correct or an error by the brain of the subject. These are not identical, and some confusion in the literature arises from the assumption that they are" (Gehring et al., 2012, p. 241). Instead of ERN-like potentials, the response-locked analysis revealed a different pattern of slow-wave effects for rule violations and correct responses at parietal sites. Whereas this effect might resemble an error positivity for violation responses (P_E; Overbeek et al., 2005; Steinhauser & Yeung, 2010), it could also be interpreted as a stimulus-preceding negativity for correct responses that has been observed in tasks involving feedback stimuli (Brunia, 1988; Hillman, Apparies, & Hatfield, 2000; van Boxtel & Böcker, 2004). Because the chicken in the current design laid an egg precisely 1,000 ms after stimulus onset, participants were able to anticipate this event that, on average, took place about 500–550 ms after their response (1,000 ms minus mean RT). It seems plausible to us that processes such as priming of authority-related concepts and negative emotions as a consequence of rule violations would affect such anticipations, which would be consistent with the stronger slow-wave drift following correct responses as observed in the data.

Conclusion

To sum up, the present results paint a first picture of the electrophysiological signature of deliberate rule violations as compared to rule-based behavior. First, the P300 response to the stimulus that prompts the behavior is delayed for rule violations, and this delay corresponds to previous behavioral findings in terms of slower responding. Second, the P300 response is attenuated at parietal sites, indicating increased response complexity. Whether this effect of response complexity is best seen as an inhibition of stimulus-response associations or as differences in resource allocation is an open question for future research. Third, more exploratory analyses indicated the parietal attenuation of the ERP to be followed by a frontal positivity, likely indicating increased affective and evaluative processing following for rule violation behavior, as is also suggested by recent behavioral studies.

References

Aarts, K., De Houwer, J., & Pourtois, G. (2012). Evidence for the automatic evaluation of self-generated actions. *Cognition*, *124*(2), 117–127. doi: 10.1016/j.cognition.2012.05.009

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied*, *70*(9), 1–70. doi: 10.1037/h0093718

Boldt, A., & Yeung, N. (2015). Shared neural markers of decision confidence and error detection. *Journal of Neuroscience*, *35*, 3478–3484. doi: 10.1523/JNEUROSCI.0797-14.2015

Brunia, C. H. M. (1988). Movement and stimulus preceding negativity. *Biological Psychology*, *26*, 165–178.

Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, *55*, 591–621. doi: 10.1146/annurev.psych.55.090902.142015

Cunningham, W. A., Espinet, S. D., DeYoung, C. G., & Zelazo, P. D. (2005). Attitudes to the right-and left: Frontal ERP asymmetries associated with stimulus valence and processing goals. *NeuroImage*, *28*(4), 827–834. doi: 10.1016/j.neuroimage.2005.04.044

Cunningham, W. A., & Zelazo, P. D. (2007). Attitudes and evaluations: A social cognitive neuroscience perspective. *Trends in Cognitive Sciences*, *11*(3), 97–104. doi: 10.1016/j.tics.2006.12.005

Debey, E., De Houwer, J., & Verschuere, B. (2014). Lying relies on the truth. *Cognition*, *132*(3), 324–334. doi: 10.1016/j.cognition.2014.04.009

Debey, E., Liefooghe, B., De Houwer, J., & Verschuere, B. (2015). Lie, truth, lie: The role of task switching in a deception context. *Psychological Research*, *79*(3), 478–488. doi: 10.1007/s00426-014-0582-4

Donchin, E. (1981). Surprise! . . . Surprise? *Psychophysiology*, *18*(5), 493–513. doi: 10.1111/j.1469-8986.1981.tb01815.x

Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, *11*(3), 357–374. doi: 10.1017/S0140525X00058027

Falkenstein, M., Hohnsbein, J., & Hoormann, J. (1990). Effects of errors in choice reaction tasks on the ERP under focused and divided attention. In C. H. M. Brunia, A. W. K. Gaillard, & A. Kok (Eds.), *Psychophysiological brain research* (Vol. *1*, pp. 192–195). Tilburg, The Netherlands: Tilburg University Press.

Falkenstein, M., Hoormann, J., Christ, S., & Hohnsbein, J. (2000). ERP components on reaction errors and their functional significance: A tutorial. *Biological Psychology*, *51*(2–3), 87–107.

Flexer, A., Bauer, H., Pripfl, J., & Dorffner, G. (2005). Using ICA for removal of ocular artifacts in EEG recorded from blind subjects. *Neural Networks*, *18*, 998–1005. doi: 10.1016/j.neunet.2005.03.012

Foerster, A., Wirth, R., Kunde, W., & Pfister, R. (2016). The dishonest mind set in sequence. *Psychological Research*. Advance online publication. doi: 10.1007/s00426-016-0780-3

Fritz, J., & Dreisbach, G. (2013). Conflicts as aversive signals: Conflict priming increases negative judgments for neutral stimuli. *Cognitive, Affective, & Behavioral Neuroscience*, *13*(2), 311–317. doi: 10.3758/s13415-012-0147-1

Gable, P. A., & Harmon-Jones, E. (2010). Late positive potential to appetitive stimuli and local attentional bias. *Emotion*, *10*(3), 441–446. doi: 10.1037/a0018425

Gable, P. A., & Harmon-Jones, E. (2013). Does arousal per se account for the influence of appetitive stimuli on attentional scope and the late positive potential? *Psychophysiology*, *50*, 344–350. doi: 10.1111/psyp.12023

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

Gehring, W. J., Liu, Y., Orr, J. M., & Carp, J. (2012). The error-related negativity (ERN/Ne). In S. J. Luck & E. Kappenman (Eds.), *Oxford handbook of event-related potential components* (pp. 231–291). New York, NY: Oxford University Press.

Hajcak, G., & Foti, D. (2008). Errors are aversive: Defensive motivation and the error-related negativity. *Psychological Science*, *19*(2), 103–108. doi: 10.1111/j.1467-9280.2008.02053.x

Hajcak, G., MacNamara, A., & Olvet, D. M. (2010). Event-related potentials, emotion, and emotion regulation: An integrative review. *Developmental Neuropsychology*, *35*(2), 129–155. doi: 10.1080/87565640903526504

Hajcak, G., Moser, J. S., Yeung, N., & Simons, R. F. (2005). On the ERN and the significance of errors. *Psychophysiology*, *42*, 151–160. doi: 10.1111/j.1469-8986.2005.00270.x

Hajcak, G., & Olvet, D. M. (2008). The persistence of attention to emotion: Brain potential during and after picture presentation. *Emotion*, *8*(2), 250–255. doi: 10.1037/1528-3542.8.2.250

Hillman, C. H., Apparies, R. J., & Hatfield, B. D. (2000). Motor and nonmotor event-related potentials during a complex processing task. *Psychophysiology*, *37*, 731–736. doi: 10.1111/1469-8986.3760731

Hoffman, J. E., Simons, R. F., & Houck, M. R. (1983). Event-related potentials during controlled and automatic target detection. *Psychophysiology*, *20*(6), 625–632. doi: 10.1111/j.1469-8986.1983.tb00929.x

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*(4), 679–709. doi: 10.1037/0033-295X.109.4.679

Holroyd, C. B., Yeung, N., Coles, M. G. H., & Cohen, J. D. (2005). A mechanism for error detection in speeded response time tasks. *Journal of Experimental Psychology: General*, *134*(2), 163–191. doi: 10.1037/0096-3445.134.2.163

Hurtado, E., Haye, A., González, R., Manes, F., & Ibáñez, A. (2009). Contextual blending of ingroup/outgroup face stimuli and word valence: LPP modulation and convergence of measures. *BMC Neuroscience*, *10*(69). doi: 10.1186/1471-2202-10-69.

Isreal, J. B., Chesney, G. L., Wickens, C. D., & Donchin, E. (1980). P300 and tracking difficulty: Evidence for multiple resources in dual-task performance. *Psychophysiology*, *17*(3), 259–273. doi: 10.1111/j.1469-8986.1980.tb00146.x

Johnson, R., Barnhardt, J., & Zhu, J. (2003). The deceptive response: Effects of response conflict and strategic monitoring on the late positive component and episodic memory-related brain activity. *Biological Psychology*, *64*(3), 217–253. doi: 10.1016/j.biopsycho.2003.07.006.

Johnson, R., Barnhardt, J., & Zhu, J. (2005). Differential effects of practice on the executive processes used for truthful and deceptive responses: An event-related brain potential study. *Cognitive Brain Research*, *24*(3), 386–404. doi: 10.1016/j.cogbrainres.2005.02.011

Keller, P. E., Wascher, E., Prinz, W., Waszak, F., Koch, I., & Rosenbaum, D. A. (2006). Differences between intention-based and stimulus-based actions. *Journal of Psychophysiology*, *20*(1), 9–20. doi: 10.1027/0269-8803.20.1.9

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology*, *38*(3), 557–577. doi: 10.1017/S0048577201990559

Kotchoubey, B. I., Jordan, J. S., Grözinger, B., & Westphal, K. P. (1996). Event-related brain potentials in a varied-set memory search task: A reconsideration. *Psychophysiology*, *33*(5), 530–540. doi: 10.1111/j.1469-8986.1996.tb02429.x

Kramer, A. F., & Strayer, D. L. (1988). Assessing the development of automatic processing: An application of dual-task and event-related brain potential methodologies. *Biological Psychology*, *26*(1), 231–267.

Maier, M. E., & Steinhauser, M. (2013). Updating expected action outcome in the medial frontal cortex involves an evaluation of error type. *Journal of Neuroscience*, *33*, 15705–15709. doi: 10.1523/JNEUROSCI.2785-13.2013

Maier, M., Steinhauser, M., & Hübner, R. (2008). Is the error-related negativity amplitude related to error detectability? Evidence from effects of different error types. *Journal of Cognitive Neuroscience*, *20*(12), 2263–2273. doi: 10.1162/jocn.2008.20159

Makeig, S., Bell, A. J., Jung, T.-P., & Sejnowski, T. J. (1996). Independent component analysis of electroencephalographic data. In D. S. Touretzky, M. C. Mozer, & M. E. Hasselmo (Eds.), *Advances in neural information processing systems* (pp. 145–151). Cambridge, MA: Bradford Books.

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus norepinephrine system. *Psychological Bulletin*, *131*(4), 510–532. doi: 10.1037/0033-2909.131.4.510

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*. Retrieved from https://www.hindawi.com/journals/cin/2011/156869/

Overbeek, T. J. M., Nieuwenhuis, 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), 319–329. doi: 10.1027/0269-8803.19.4.319

Pfister, R. (2013). *Breaking the rules: Cognitive conflict during deliberate rule violations*. Berlin, Germany: Logos.

Pfister, R., Foerster, A., & Kunde, W. (2014). Pants on fire: The electrophysiological signature of telling a lie. *Social Neuroscience*, *9*(6), 562–572. doi: 10.1080/17470919.2014.934392

Pfister, R., & Janczyk, M. (2013). Confidence intervals for two sample means: Calculation, interpretation, and a few simple rules. *Advances in Cognitive Psychology*, *9*(2), 74–80. doi: 10.2478/v10053-008-0133-x

Pfister, R., Wirth, R., Schwarz, K., Steinhauser, M., & Kunde, W. (2016). Burdens of non-conformity: Motor execution reveals cognitive conflict during deliberate rule violations. *Cognition*, *147*, 93–99. doi: 10.1016/j.cognition.2015.11.009

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. doi: 10.1016/j.clinph.2007.04.019

Pratt, N., Willoughby, A., & Swick, D. (2011). Effects of working memory load on visual selective attention: Behavioral and electrophysiological evidence. *Frontiers in Human Neuroscience*, *5*(57). doi: 10.3389/fnhum.2011.00057.

Ramchurn, A., de Fockert, J. W., Mason, L., Darling, S., & Bunce, D. (2014). Intraindividual reaction time variability affects P300 amplitude rather than latency. *Frontiers in Human Neuroscience*, *8*, 557. doi: 10.3389/fnhum.2014.00557.

Renault, B., Ragot, R., & Lesevre, N. (1980). Correct and incorrect responses in a choice reaction time task and the endogenous components of the evoked potential. *Progress in Brain Research*, *54*, 547–554.

Roche, R. A. P., & O'Mara, S. M. (2003). Behavioral and electrophysiological correlates of visuomotor learning during a visual search task. *Cognitive Brain Research*, *15*, 127–136. doi: 10.1016/S0926-6410(02)00146-5

Spence, S. A., Farrow, T. F. D., Herford, A. E., Wilkinson, I. D., Zheng, Y., & Woodruff, P. W. R. (2001). Behavioural and functional anatomical correlates of deception in humans. *NeuroReport*, *12*(13), 2849–2853.

Steinhauser, M., & Yeung, N. (2010). Decision processes in human performance monitoring. *Journal of Neuroscience*, *30*, 15643–15653. doi: 10.1523/JNEUROSCI.1899-10.2010

Stemmer, B., Witzke, W., & Schönle, P. W. (2001). Losing the error related negativity in the EEG of human subjects: An indicator for willed action. *Neuroscience Letters*, *308*(1), 60–62. doi: 10.1016/S0304-3940(01)01974-7

van Boxtel, G. J., & Böcker, K. B. (2004). Cortical measures of anticipation. *Journal of Psychophysiology*, *18*, 61–76. doi: 10.1027/0269-8803.18.23.61

Verleger, R. (1997). On the utility of P3 latency as an index of mental chronometry. *Psychophysiology*, *34*(2), 131–156. doi: 10.1111/j.1469-8986.1997.tb02125.x

Verleger, R. (2008). P3b: Towards some decision about memory. *Clinical Neurophysiology*, *119*(4), 968–970. doi: 10.1016/j.clinph.2007.11.175

Verleger, R., Jaśkowski, P., & Wascher, E. (2005). Evidence for an integrative role of P3b in linking reaction to perception. *Journal of Psychophysiology*, *19*(3), 165–181. doi: 10.1027/0269-8803.19.3.165

Walczyk, J. J., Roper, K. S., Seemann, E., & Humphrey, A. M. (2003). Cognitive mechanisms underlying lying to questions: Response time as a cue to deception. *Applied Cognitive Psychology*, *17*(7), 755–774. doi: 10.1002/acp.914

Waszak, F., Wascher, E., Keller, P. E., Koch, I., Aschersleben, G., Rosenbaum, D. A., & Prinz, W. (2005). Intention-based and stimulus-based mechanisms in action selection. *Experimental Brain Research*, *162*(3), 346–356. doi: 10.1007/s00221-004-2183-8

Wickens, C., Kramer, A., Vanasse, L., & Donchin, E. (1983). Performance of concurrent tasks: A psychophysiological analysis of the reciprocity of information-processing resources. *Science*, *221*, 1080–1082. doi: 10.1126/science.6879207

Wirth, R., Foerster, A., Rendel, H., Kunde, W., & Pfister, R. (2016). Looking for trouble, then looking for cops: Rule-violations sensitize towards authority-related stimuli. Manuscript submitted for publication.

Wirth, R., Pfister, R., Foerster, A., Huestegge, L., & Kunde, W. (2016). Pushing the rules: Effects and aftereffects of deliberate rule violations. *Psychological Research*, *80*(5), 838–852. doi: 10.1007/s00426-015-0690-9

(Received February 7, 2016; Accepted September 12, 2016)