Psychophysiology, ■■■ (2010), 1–9. Wiley Periodicals, Inc. Printed in the USA.

DOI: 10.1111/j.1469-8986.2010.01152.x

On the number of trials needed for a stable feedback-related negativity

JOSEP MARCO-PALLARES,[a,b] DAVID CUCURELL,[b] THOMAS F. MÜNTE,[c,d] NADINE STRIEN,[e] AND ANTONI RODRIGUEZ-FORNELLS[a,b,f]

[a]Department of Basic Psychology, University of Barcelona, Barcelona, Spain
[b]Institut d'Investigació Biomèdica de Bellvitge (IDIBELL), Barcelona, Spain
[c]Department of Neuropsychology, Otto von Guericke University, Magdeburg, Germany
[d]Department of Neurology, University of Lübeck, Lübeck, Germany
[e]Deutsches Zentrum für Neurodegenerative Erkrankungen (DZNE), Partnerinstitut Magdeburg, Magdeburg, Germany
[f]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

Abstract

Feedback-related negativity is an event-related brain potential elicited by negative feedback. Its properties make it a valuable tool for the assessment of cognitive-affective processes that are involved in feedback and reward processing. The present study sought to determine the minimum number of trials that are required to obtain a reliable FRN component using a simple gambling paradigm. Three independent groups of young participants and one group of old participants were used. In the experimental conditions with healthy young controls, 20 trials were sufficient to measure the optimal FRN amplitude. In older participants, 50 trials were needed to obtain a reliable FRN. Whereas 20 trials would be enough to ensure a reliable FRN component in studies with nonclinical samples, the number of trials needed in clinical and cognitively impaired populations has to be determined based on the signal-to-noise ratios and the characteristics of the signals recorded.

Descriptors: Reliability, ERP, Medial frontal negativity, FRN, Reward, Gambling

The feedback-related negativity (FRN), also known as the mediofrontal negativity, is an event-related potential (ERP) component elicited after negative feedback that signals an incorrect response (Miltner, Braun, & Coles, 1997; Müller, Möller, Rodriguez-Fornells, & Münte, 2005) or elicited after feedback that signals a monetary loss in gambling paradigms (Gehring & Willoughby, 2002; Marco-Pallares et al., 2008). The FRN shows a mid-frontal distribution (slightly right-lateralized in some studies) and peaks at about 250 to 300 ms after the appearance of the feedback stimulus. Research on this component has provided fundamental information for the development of theories (Frank, Moustafa, Haughey, Curran, & Hutchison, 2007; Holroyd & Coles, 2002) with regard to reinforcement-learning models and reward processing. The amplitude of the difference waveform between the win and loss trials can be representative of an error signal that is used to adjust the cognitive control system in the presence of negative action outcomes (see Holroyd, Pakzad-Vaezi, & Krigolson, 2008).

These properties make the FRN component valuable for the assessment of cognitive–affective processes that are involved in feedback and reward processing in normal populations and in clinical samples. Because of its potential use in clinical settings, short paradigms with relatively few trials that do not induce fatigue, drowsiness, or decreases in motivation in patients are highly desirable. Clinical evaluations are often carried out under nonoptimal conditions; however, these evaluations lead to low signal-to-noise ratios (SNRs) due to the lack of proper electromagnetic and acoustic isolation, involuntary movements, and other sources of noise. The present study was therefore designed to determine the minimum number of trials required to obtain a stable FRN component using a simple gambling paradigm. We first compared the data from two large samples that were obtained in different settings and with different SNRs using the same gambling paradigm. These data have been published previously with a focus on the basic electrophysiological effects in a gambling paradigm (Marco-Pallares et al., 2008) and their modulation by genetic factors (Marco-Pallarés et al., 2009). Also, because of the field's interest in studying the FRN component in clinical populations, we investigated the number of trials needed to obtain a reliable FRN component in two additional groups of old and young participants that were recorded using the same experimental paradigm. In doing so, we aimed to

Support from the Ramon y Cajal research program was awarded to J.M.-P. Funding was also provided by Spanish Government grants (MICINN, PSI2008-03901 to A.R.-F. and PSI2009-09101 to J.M.-P.) and a grant from the Catalan Government (2009 SGR 93). T.F.M. was supported by the DFG (SFB 779, TP A5).

Address correspondence to: Josep Marco-Pallares, Ph.D., Department of Basic Psychology, Campus Bellvitge, University of Barcelona, Feixa Llarga s/n, 08907 L'Hospitalet, Barcelona, Spain. E-mail: josep marco@gmail.com

PSYCHOPHYSIOLOGY

determine the basic requirements for standard experiments and the requirements for clinical studies that would lead to SNR reductions.

A number of previous studies have been published that address the question of the minimum number of trials required for the reliable delineation of certain ERP components. For example, Polich (1986) showed that 20 trials were enough for a reliable recording of the P300 in an oddball paradigm and that no differences were observed if further trials were added (Cohen & Polich, 1997). With regard to the error-related negativity (ERN) component, Olvet and Hajcak (2009) showed that a minimum number of six to eight trials suffice. This result was similar to that reported in a study by Pontifex et al. (2010) that was devoted to establishing the minimum number of trials needed to characterize the ERN across the life span. The functional similarity and the possible common neural sources of the FRN and the ERN components (Yeung, Botvinick, & Cohen, 2004) suggest that a similar number of trials might be necessary to record a reliable FRN component. However, the ERN is a response-locked ERP component that is elicited in a tightly time-locked manner shortly after an erroneous response. As a result, the SNR of this component might be higher than that of the FRN component, which is elicited after the presence of a specific negative event and later in the chain of cognitive events (at about 250–300 ms).

Taking into consideration the unexplored effects of the SNR (e.g., recording settings, paradigm used, and the population under study) and the differences between the FRN and the ERN components, we investigated the reliability of the FRN component in three different contexts, a healthy population with high SNR recording conditions, a healthy population with low SNR recording settings, and a population that was possibly cognitively impaired with a low SNR.

Material and Methods

Participants

Twenty-five right-handed healthy undergraduate psychology students attending the University of Barcelona participated in the first experiment (Study 1; 7 men, mean age $\pm$ *SD*: 23.7 $\pm$ 5.4 years). Forty-eight students (mean age: 28.1 $\pm$ 3.1 years, 34 women) participated in the second experiment (Study 2). Eight older subjects (mean age: 60.5 $\pm$ 5.1 years, 2 women) and 8 young volunteers (mean age: 24.3 $\pm$ 2.5 years, 3 women) participated in the third experiment. The local ethics committee approved all of the procedures, and written informed consent was obtained from all of the participants.

Experimental Paradigm

We used an established gambling task (Gehring & Willoughby, 2002) in which the numbers 5 and 25 were presented in white on a black background in one of the two possible orders: [5 25] or [25 5] (Marco-Pallares et al., 2008, 2009). Participants selected one of the numbers by pressing a spatially corresponding button with the left or right index finger. One second after the choice, one of the numbers turned green, whereas the other changed to red. If the number selected by the participant changed to red (or green), this signaled a loss (or gain) of the corresponding amount of money (in Euro cents). Two seconds later, the next trial began with the presentation of a warning signal (an asterisk for 1000 ms duration in the first experiment and 500 ms duration in the second experiment) followed by a new pair of numbers. Participants were provided with an initial sum of 10€, were encouraged to gain as much as possible, and were familiarized with the task during a brief practice block. The first experiment comprised 16 blocks of 48 trials each, whereas the second experiment comprised 17 blocks of 40 trials each. In the first experiment, each trial presented a .5 probability of winning and a .5 probability of losing. In the second experiment, the mean expected value of the monetary outcome was zero on each block.

In the third experiment, there was a slight modification of the previous designs, with the inclusion of boost wins and losses. The experiment comprised 21 blocks of 40 trials each in which feedback was either a standard trial in 90% of cases (as in the previous experiments) or a boost trial. The latter occurred only in 10% of all of the trials and consisted of a doubling of the quantity of the monetary gain or loss. These trials were not analyzed in the present article. As in the first experiment, each trial presented a .5 probability of winning and a .5 probability of losing. Finally, in the present study, we only analyzed monetary gains and losses of 25 cents on standard trials (except in those cases that are mentioned explicitly in the Results section).

An EEG was recorded using tin electrodes that were mounted in an elastic cap and located at 29 standard positions (Fp1/2, Fz, F7/8, F3/4, Fc1/2 Fc5/6, Cz, C3/4, T7/8, Cp1/2, Cp5/6, Pz, P3/4, P7/P8, Po1/2, and O1/2). Biosignals were referenced off-line to the mean of the activity at the two mastoid processes. Vertical eye movements were monitored with an electrode that was placed at the infraorbital ridge of the right eye. Electrode impedances were kept below 5 kΩ. The EEG-recording devices for the experiments were a Neuroscan Synamp system for the first experiment, a BrainAmp system for the second experiment, and a PL-351-system (Walter Graphtek GmbH, Bad Oldesloe, Germany) for the third experiment. The first experiment was performed inside a Faraday Cage, whereas the second and third experiments were performed in non-isolated laboratory environments. In both non-isolated environments, EEGs were recorded in standard psychological laboratories but without Faraday Cage isolation.

EEG was lowpass filtered off-line to 40 Hz and ERPs time-locked to the color change of the number displays were averaged for epochs starting 100 ms prior to the stimulus (baseline) to 600 ms after the presentation of the feedback. Epochs exceeding $\pm$ 50 μV in the electrooculogram (EOG) or the EEG were removed from further analyses. For the purposes of the present study, only ERPs associated to gains and losses at Fz (maximum activity electrode for FRN) were studied. Difference waveforms for the loss minus the gain were extracted by averaging 1 to 60 trials in 1 trial steps. The mean amplitude of the FRN component in the difference ERP was quantified in a 100-ms time window around the 60-trials FRN peak.

To study the reliability of the data that were computed with different numbers of trials, we used the Cronbach's alpha of the amplitude of the FRN component that was averaged over *n* trials (n = 1–59) with a full average of 60 trials. We also computed the SNR dividing the mean root square of the signal by the variance of the baseline (– 100 to 0 ms) as in a previously published work (Maidhof, Rieger, Prinz, & Koelsch, 2009). To assess the differences between the SNRs using different numbers of trials, we used the nonparametric test proposed by Maris and Oostenveld (2007) that was corrected for multiple comparisons arising from the large number of comparisons performed (one comparison for each time point).

Results

Twenty-five subjects in the first study, 43 subjects in the second study, and all of the participants in the third study had more than 60 artifact-free trials in both the maximum gain and loss conditions and were used for the rest of the computations.

Figure 1 shows the FRN component that was elicited by averaging 5, 10, 20, 30, 45, and 60 trials in Experiments 1 and 2. As shown in Figure 1, for the first experiment, a clear FRN component was present in the 10-trial average, and from 20 trials onward there were virtually no differences from the full average of 60 trials. A similar result was found in Experiment 2; however, the averages based on 10 trials were not yet stable (see also the increase in the variability reflected in the standard error of the mean; Figure 2).

Figure 2A shows the mean amplitude of the difference between the maximum gain and the maximum loss (the time window measured was 250–350 ms after the presentation of the feedback). As shown in Figure 2A, the FRN component for Experiment 1 rapidly stabilizes at 10 trials, whereas for Experiment 2, the amplitude is stable after 20 trials. Figure 2B shows the Cronbach's alpha of the FRN amplitude computed with different numbers of trials and the FRN amplitude obtained with the maximum number of 60 trials. In Experiment 1 Cronbach's alpha was higher than .7 from 5 averaged trials onward. In contrast, Cronbach's alpha reached a value of .7 only after 22 trials in Experiment 2. These differences may be explained by the different recording conditions: The first experiment was conducted in a shielded recording chamber, which should lead to a better SNR and, hence, to a smaller number of trials required for the computation of a stable FRN component. The SNR differs greatly between Experiments 1 and 2 as illustrated in Figure 2C,D. In the first experiment, an increase in the SNR of the FRN time range was present at 10 trials, and the SNR was 5 for 30 trials and nearly 7 for 60 trials. In contrast, the first identifiable increase in the SNR for the second experiment was at 20 trials. In this experiment, the SNR for 30 and 45 trials was 3.5 and 4.5, respectively, and thus was considerably less than in the first experiment. However, both experiments presented similarly significant differences when we directly compared the mean amplitude in 60 trials versus 5, 10, and 20 trials (the selected time window was from 200 to 350 ms; see color bars in Figure 2C,D). In contrast, no significant differences were found between 60 trials and 30 or 45 trials in both experiments (Figure 2C,D).

Figure 3A shows the evolution of the kurtosis (left) and skewness (right) of the FRN component across the number of trials. As shown in this figure, from 10 trials on, the distribution of the FRN amplitudes across subjects shows a moderate leptokurtosis in Experiments 1 and 2. Also, the skewness of the data shows values close to 0 in both experiments at 20 trials and beyond. Figure 3B shows the distribution of the FRN components for Experiment 1 for $N = 1$ (left) and $N = 10$ (right). As expected, increasing the number of trials results in a more normal distribution. A similar pattern is observed in Experiment 2 (Figure 3C) for $N = 1$ (left) and $N = 20$ (right), which is the proposed cutoff for the number of trials.

Although the results that have been previously described showed a clear effect of the number of trials on obtaining a stable FRN component in a healthy population, we wanted to investigate these effects in different averaging conditions. First, the previous results were based on the average of sequential trials. We also wanted to study the effect of averaging trials in a randomized order, instead of in a consecutive order. Therefore, we generated 5000 data sets that randomized the order of the 60 trials. Then, we repeated the analysis of the mean of the maximum gain minus the maximum loss between 250 and 350 ms. Figure 4 shows the result of 20 randomizations (in colors) and the original data set that was computed with consecutive trials (in black) for Experiment 1 (Figure 4A) and Experiment 2 (Figure 4B). Also, in Figure 4C, the histogram of the amplitudes from the 5,000 randomized data sets for 1 (left) and 10 (right) trials is shown. As shown by this figure, with a low number of trials, differences among the different sets are very large; however, from 10 trials on, these differences are very small, suggesting that the sequential approach is not biased. Along the same lines, Figure 4D shows the histogram of the 5,000 data sets for Experiment 2 for 1 (left) and 20 (right) trials, yielding results similar to those in Experiment 1. Therefore, these results clearly show that the mean amplitude obtained using a very low number of randomly sampled trials is different when compared to those averages calculated using sequential trials; however, these differences tend to disappear when the number of averaged trials increases (see Figure 4A,B).

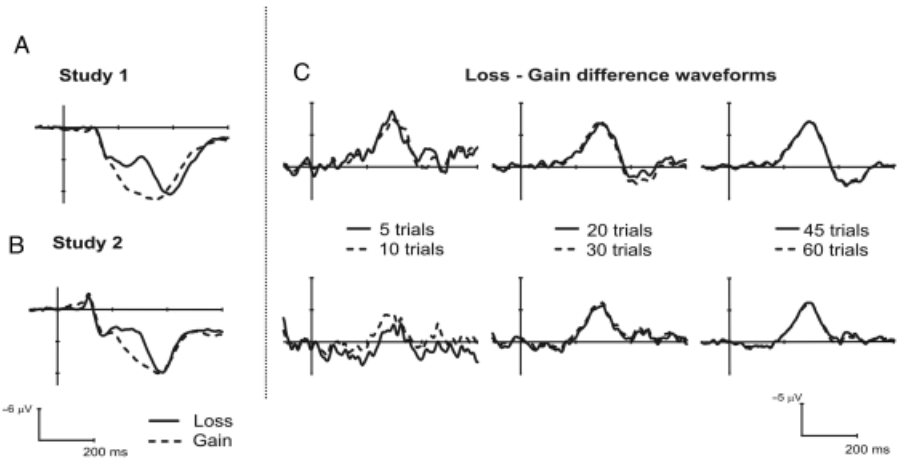

Figure 1. Event-related potentials for maximum gain (dashed line) and maximum loss trials (solid) based on 60 trials for Experiments 1 (A) and 2 (B). C: Difference waves for the maximum loss minus the maximum gain based on different numbers of trials (Experiment 1, top, and Experiment 2, bottom).

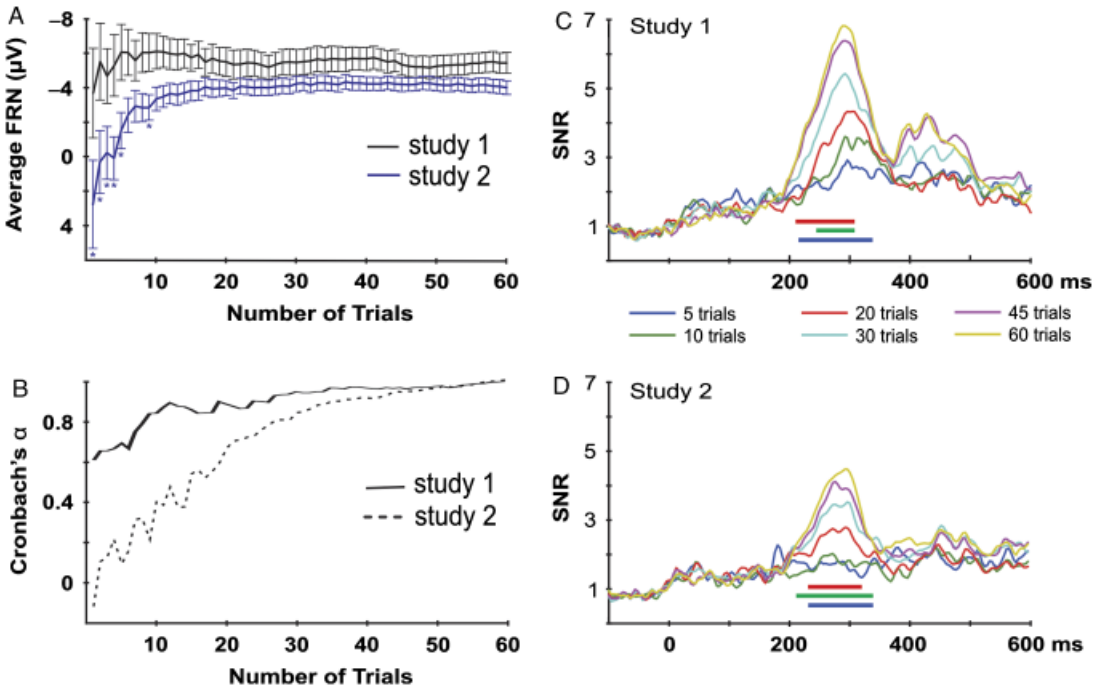

Figure 2. A: Average (± SEM) of the FRN component (mean amplitude for the 250–350-ms window) based on different numbers of trials for Experiments 1 (black) and 2 (blue). B: Cronbach's alpha of the FRN amplitude that was determined from averages of *n* trials and the full average of 60 trials for Experiments 1 (solid) and 2 (dashed). C: The SNR of Experiment 1 as a function of the number of trials per average. Horizontal lines indicate the time ranges in which the significant differences between the average based on 60 trials and the averages based on 5 (blue), 10 (green), or 20 (red) trials were significant. D: Same as C for Experiment 2.

We also analyzed to what extent the measure employed in the determination of the FRN component could affect the results. Given that in some studies the FRN component is assessed as the peak of the component for individual subjects, we repeated the FRN measures by computing the peak of the FRN and computing the mean value using 10 ms around it for each individual subject and number of trial. Figure 5 shows how this measure changes with the number of trials for Experiment 1 (Figure 5A, left) and Experiment 2 (Figure 5A, right). As expected, when using a small number of trials, the amplitudes were larger and the amount of variability increased, especially compared to the measures obtained when we computed the mean amplitude between 250 and 350 ms. However, this amplitude was reduced with the number of trials and stabilized for number of trials similar to those in the mean value results. Therefore, Cronbach's alpha for Experiment 1 reached values greater than .7 from 5 trials on (Figure 5B), whereas for Experiment 2, the number of trials needed to reach an alpha value of greater than .7 was 22 trials (Figure 5C). Therefore, the number of trials needed to acquire a stable FRN component was very similar when we used a global mean amplitude measure from a selected time window or an individual peak value.

The measures used to obtain results described in previous paragraphs use maximum gains and losses. We also wanted to test whether minimum gains and losses followed the same pattern. Figure 6 shows the results for the 250–350-ms mean of the maximum loss minus the maximum gain (black) and the minimum loss minus the minimum gain (green) for Experiment 1 (Figure 6A) and Experiment 2 (Figure 6C). In the latter study, only 39 of the original 43 registers presented more than the 60 minimum gain and loss trials. Although the minimum conditions presented smaller amplitudes in both studies, the stabilization pattern was similar. Figure 6B,D also shows the Cronbach's alpha for the two experiments and shows results similar to the ones described using the maximum conditions.

Finally, to test whether the previous results could be extended to settings in clinical studies, we applied a similar analysis to two small groups ($N = 8$) of young and old participants (see maximum gain-loss averages in Figure 7A,B). Data were obtained in a non-isolated environment. Figure 7C shows that for the young and old groups, in spite of the low number of subjects, the FRN component changed very little after 20 trials (see also Figure 8A). As in Experiments 1 and 2, Figure 8A clearly shows a decrease in the amplitude of the FRN component for the older participants after it stabilized (after approximately 20 trials; Eppinger & Kray, 2011; Eppinger, Kray, Mock, & Mecklinger, 2008; Nieuwenhuis et al., 2002). To test how changes in FRN with the number of trials might affect a possible difference between the FRN amplitude of both groups, we used an independent two-sample *t* test for each FRN amplitude resulting from a different number of trials. Group differences ($p < .05$) appeared systematically after 33 trials and remained stable until 60 trials. Therefore, more than 30 trials were required to observe a reliable group effect.

Importantly, the old group but not the young group exhibited a larger variability than the samples in Experiments 1 and 2. This fact became apparent when the Cronbach's alpha was computed using the amplitude of the FRN component that was averaged from *n* trials and using the full average (Figure 8b). Whereas the young group showed a Cronbach's alpha greater than .7 after 21 trials, the older group reached this value at 47 trials. Also, the SNR for the young group was similar to that in Experiments 1 and 2, whereas the SNR for the older group presents smaller values (Figure 8C,D). However, the SNR at 60 trials in the

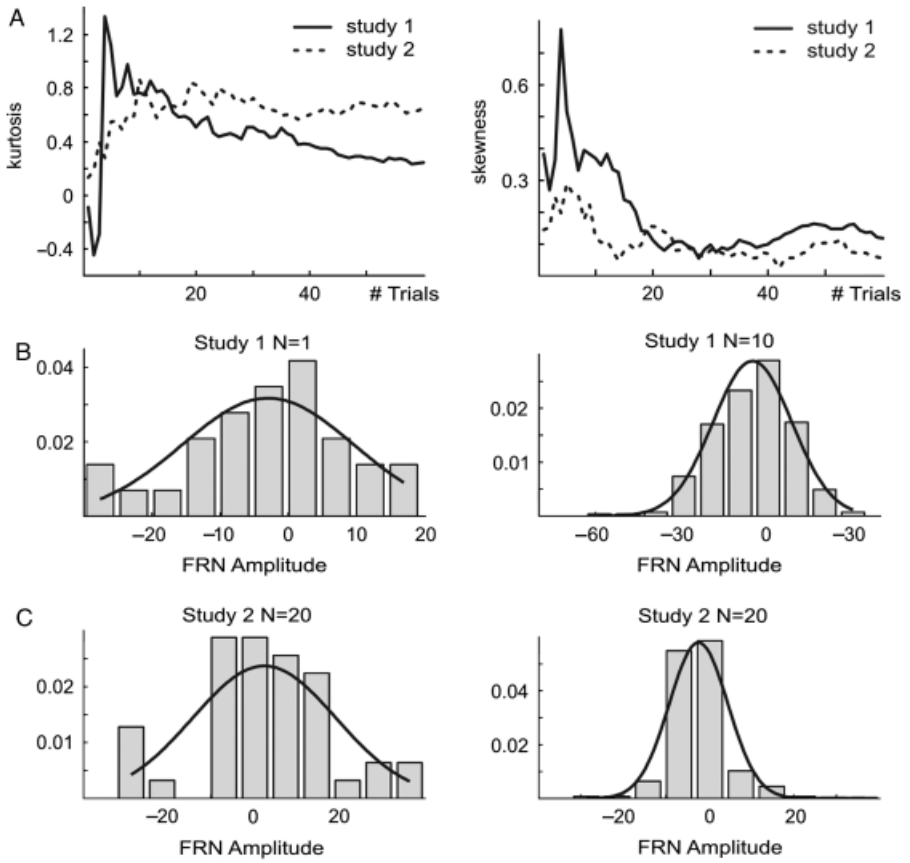

Figure 3. A: The kurtosis (left) and skewness (right) of the FRN component (mean amplitude for the 250–350-ms window) for the different numbers of trials for Experiment 1 (solid line) and Experiment 2 (dashed line). B: Distribution of the FRN amplitudes for $N = 1$ (left) and $N = 10$ (right) trials for Experiment 1. The Gaussian distribution is shown as a solid line. C: Distribution of the FRN amplitudes for $N = 1$ (left) and $N = 20$ (right) trials for Experiment 2.

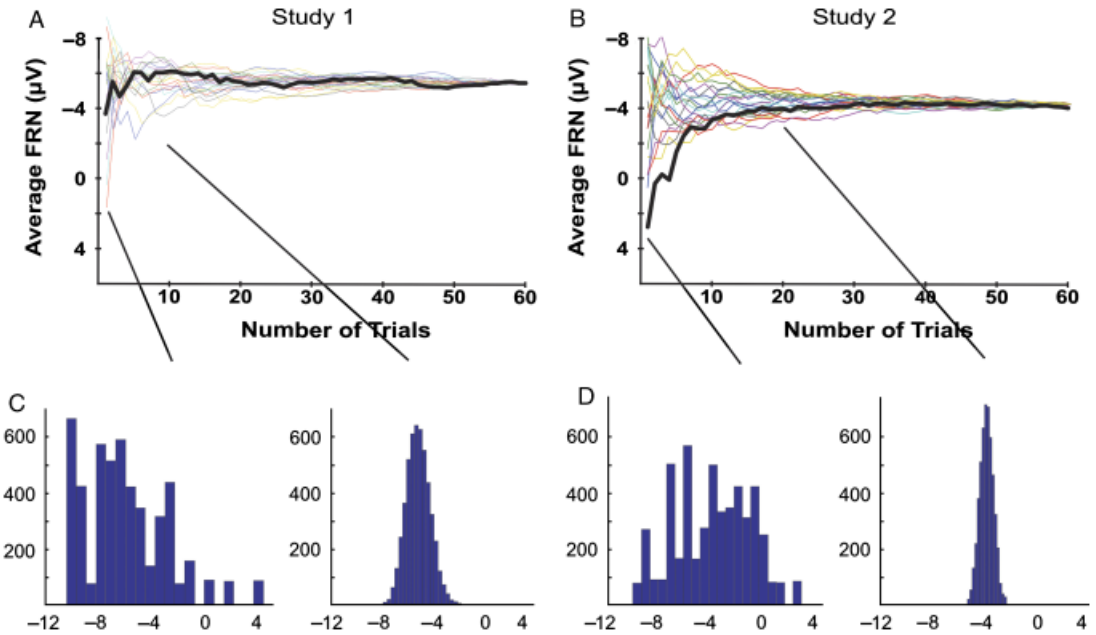

Figure 4. Changes in the of FRN amplitude with the number of trials for 20 data sets of trials in a randomized order (color) and sequential order (black) for Experiment 1 (A) and Experiment 2 (B). The distribution of 5,000 data sets of trials in a randomized order for Experiment 1 is shown in C ($N = 1$ left, $N = 10$ right). This distribution for Experiment 2 is shown in D ($N = 1$ left, $N = 20$ right).

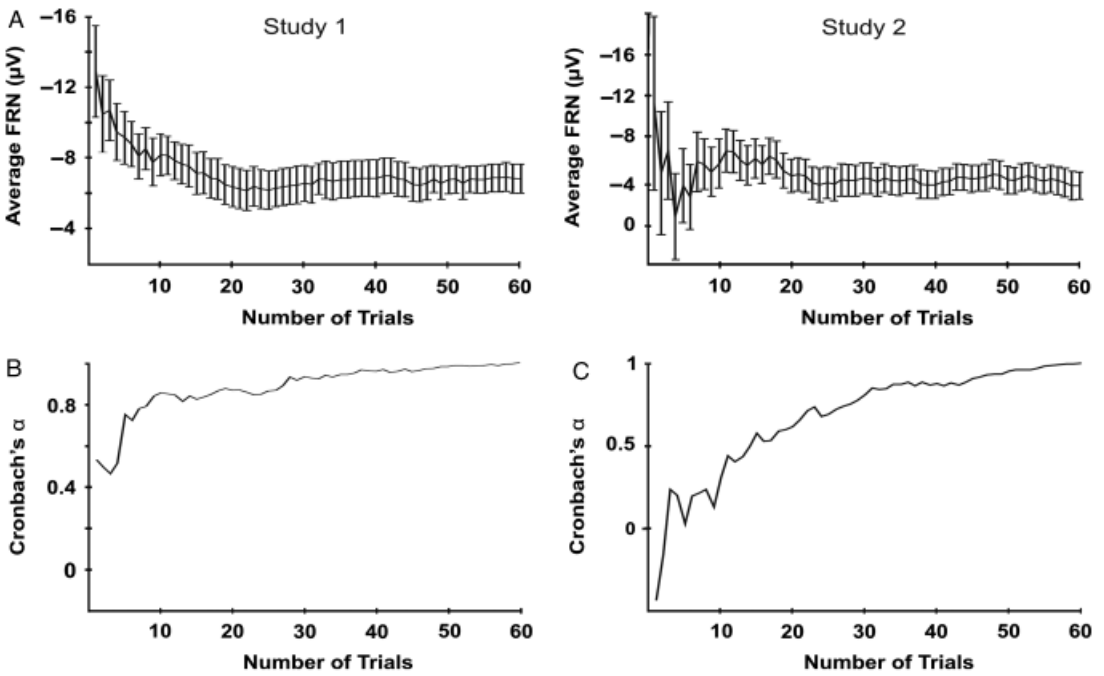

Figure 5. A: Average (± SEM) of the FRN components computed for 10 ms around the individual peak and based on the different numbers of trials for Experiments 1 (left) and 2 (right). B: Cronbach's alpha of the FRN amplitude determined from the averages of *n* trials and the full average of 60 trials for Experiment 1. C: The same value as B for Experiment 2 is shown.

young group was significantly different only at 5 and 10 trials in the 275–300-ms time range, whereas no significant differences were found in the SNR in the older group among the different numbers of trials. Finally, using randomized mean trials instead of sequential trials (see above) showed that, for both groups, the use of only 1 trial presented a sparse distribution, whereas for 20 trials, the width of the distribution was reduced (Figure 9A for the young group and Figure 9B for the old group).

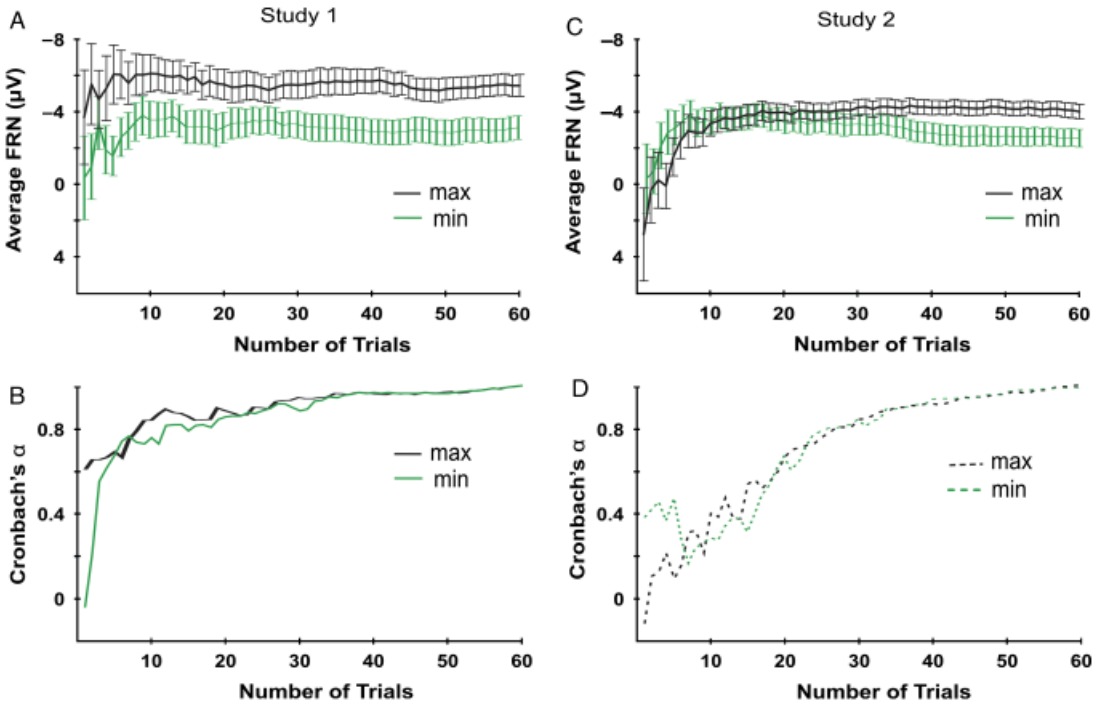

Figure 6. A: Average (± SEM) of the minimum loss minus the minimum gain of the FRN component (mean amplitude for the 250–350-ms window) based on different numbers of trials for Experiment 1. These averages were computed using the minimum loss minus the minimum gain (green) and the maximum loss minus the maximum gain (black). B: Cronbach's alpha of the FRN amplitude for the minimum loss minus the minimum gain determined from the averages of *n* trials and the full average of 60 trials for Experiment 1. Calculation of this value used minimum trials (green) and maximum trials (black). C,D: Same as A and B for Experiment 2.

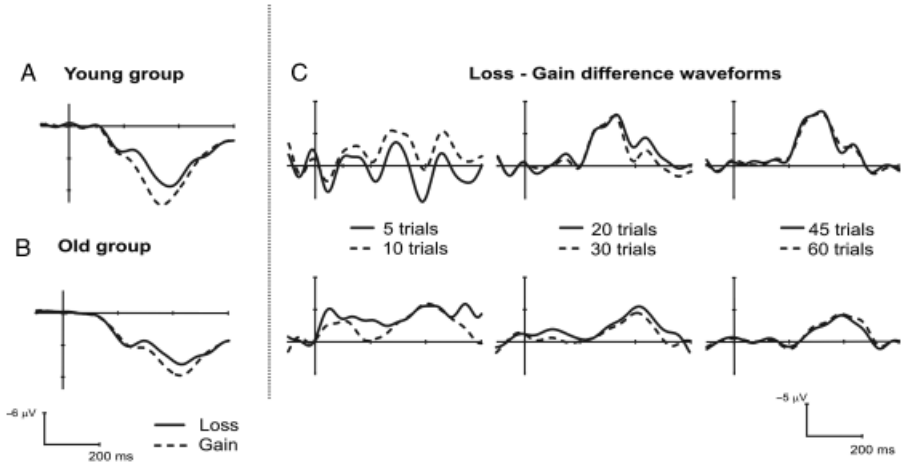

Figure 7. Event-related potentials for maximum gain (dashed line) and maximum loss trials (solid) based on 60 trials for the young (A) and older group (B) from Experiment 3. Note that the FRN of the older group peaks approximately 100 ms later than that seen in the young group. C: Difference waves for the maximum loss minus the maximum gain for different numbers of trials (young group, top, and old group, bottom).

Discussion

The goal of the present study was to determine how many trials were needed to obtain a reliable FRN component. To this end, four different samples and recording environments with different SNRs were assessed. When the SNR was large, as in Experiment 1, and the recordings were carried out in an isolated laboratory, as few as 10 trials were enough to obtain an optimal FRN. Note that this number is similar to the one reported for the ERN component by Olvet and Hajcak (2009) and Pontifex et al. (2010). For lower SNRs, as in the second experiment and in the recordings from the young group that were obtained in non-isolated environments, the number of trials required for a reliable FRN component was considerably larger (approximately 20 trials). Thus, the number of trials needed depends on the SNR of the component in a specific sample.

If all of the results of the experiments in the present study are considered together, it seems that the number of trials needed for

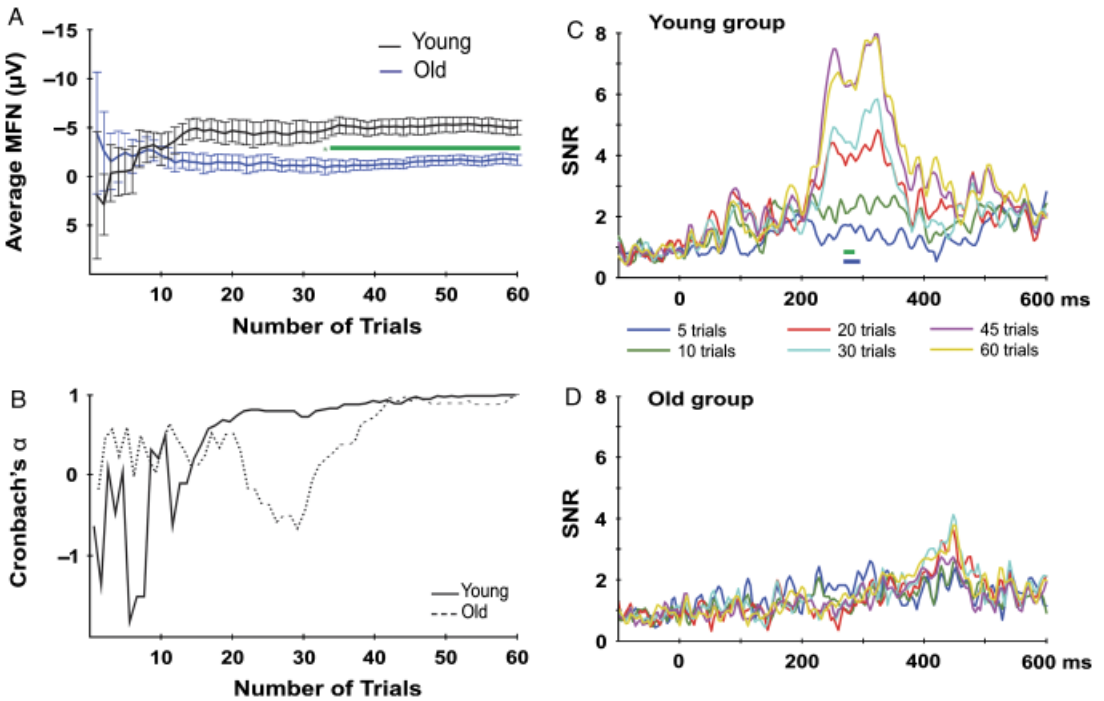

Figure 8. A: Average (± SEM) of the FRN component based on different numbers of trials for young (mean amplitude in the 250–350-ms window, black) and older (mean amplitude in the 350–450-ms window, blue) participants from Experiment 3. The green asterisk indicates the number of trials ($N = 33$) at which the difference between the old and young groups became significant (the horizontal green line indicates that this difference was significant for all subsequent comparisons based on more trials). B: Correlation of the FRN amplitude determined from averages of *n* trials and the full average of 60 trials for young (solid) and older (dashed) participants of Experiment 3. C: SNR for young participants from Experiment 3 as a function of the number of trials per average. Horizontal lines indicate the time ranges in which the significant differences between the average based on 60 trials and averages based on 5 (blue), 10 (green), or 20 (red) trials were significant. D: Same as C for the older participants.

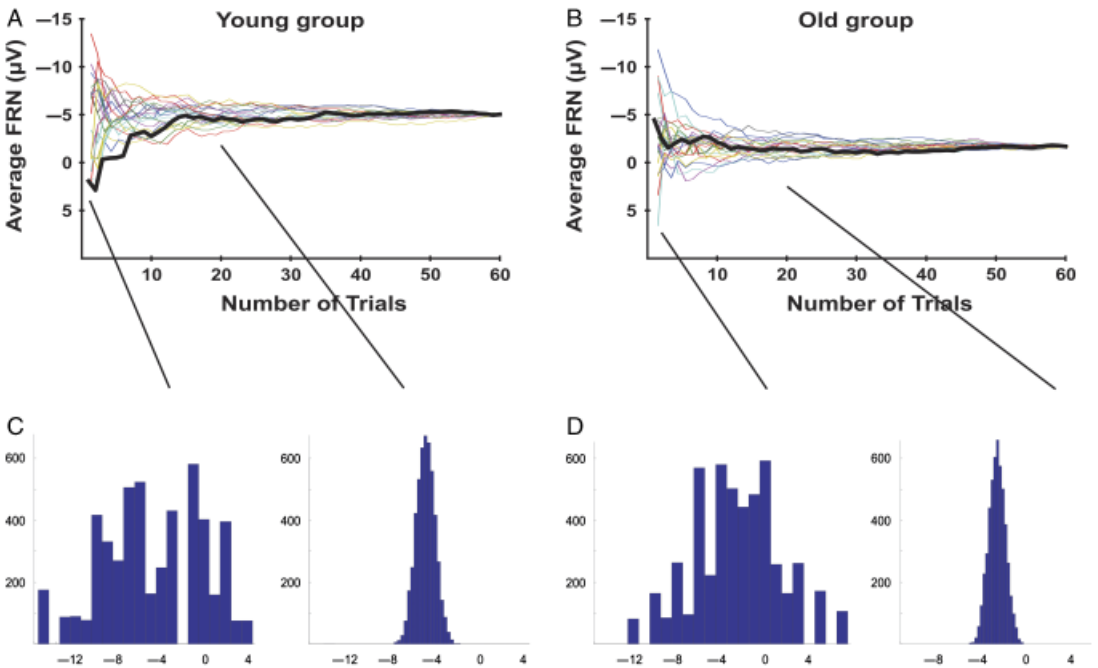

Figure 9. Changes in the FRN amplitude over the number of trials for 20 sets of trials in a randomized order (color) and in a sequential order (black) for the young group (A) and the old group (B). The distribution of 5,000 sets in a randomized order is shown in C for the young group and in D for the old group ($N = 1$ left, $N = 20$ right).

the computation of a reliable FRN component is higher than the number of trials needed for a reliable ERN component (cf. Olvet & Hajcak, 2009). This finding might be because the ERN component is a response-related component and is therefore more tightly time-locked. This leads to a higher SNR and to less trial-to-trial variation than for stimulus-locked components.

A second important conclusion of the present study was that the reliability of the FRN component depends on the sample characteristics. Results from Experiments 1 and 2 and from the young group of the third experiment were typical for healthy, young, and collaborative volunteers. Although the results depended on the recording conditions (10 trials in Faraday Cage recordings or 20 in a non-electrically isolated room), they seemed to be stable between experiments with similar conditions. Therefore, the results from Experiment 2 and the young group in Experiment 3 were very similar, even though the number of subjects was higher in the former ($N = 43$ in Study 2; $N = 8$ in Study 3) and the EEG recording systems were different. Therefore, these results suggested that in nonclinical experiments, even those in non-isolated environments, 20 trials might be enough to get a reliable FRN component. In contrast, the results from the old population in our third experiment suggested that the number of trials needed in clinical or cognitively impaired samples might be much larger. The need of more trials was due to the decreased SNR in the older group of the third experiment. However, the FRN component that was computed from 20 artifact-free trials was very similar to the one from all 60 trials (see Figure 7C). Thus, one could conclude that 25 to 30 artifact-free trials could be sufficient, even in clinical populations. However, significant differences in the FRN between the old and young group emerged after 33 trials, and the intrasubject correlation of the FRN amplitude (Figure 8B) reached a value greater than .7 only after 50 trials. The lack of reliability of the FRN component in such samples might be an issue if the number of trials is small. For example, 20 trials were sufficient for measuring a reliable FRN component in healthy populations, but the number of trials needed in clinical and cognitively impaired populations has to be determined individually based on the SNR and the characteristics of the signal. A reasonable estimate for such populations is approximately 50.

One limitation of this study was the small sample sizes used in the third experiment for the old group. Therefore, the previous conclusions about the increase of number of trials in clinical or subclinical populations may only be true for small sample sizes. However, as we stated above, the young group of Experiment 3 had eight subjects and was recorded in the same conditions as the older group, and the results were comparable to the large group evaluated in Experiment 2 (the SNR was also very similar). Thus, we believe that it is unlikely that the size of the sample is responsible for the results from the older group. Another important open question about the generalization of these results is whether the number of trials estimated in this study applies only to gambling designs similar to the one used in the three experiments or can be generalized to other experimental paradigms that also elicit the FRN component. Although this issue is far from the scope of the present article, an important caveat arises if present results are used in designing experimental paradigms different from the gambling paradigms proposed above.

Interestingly, a recent comparison of the ERN component between young and older participants revealed a reduced amplitude for the latter group (Eppinger & Kray, 2011). In addition, the component was not present in some conditions, and this finding was also reported in previous studies (Eppinger et al., 2008; Nieuwenhuis et al., 2002). The authors of this work speculated that this finding may have been due to the lower SNR in this particular sample. The authors then rejected this argument because the P1 and N1 components were of similar or even larger amplitude for the older group compared to the young group. An

important caveat here is that the SNR of a particular component (e.g., the N1) cannot be used to infer the SNR of another one (e.g., ERN or FRN) because the neural sources involved in their generation might not be shared. Therefore, the reduced amplitude that was observed for the ERN component could be a consequence of a selective reduction of the SNR in this component. As a result, it is important to measure the specific SNR of the component of interest to find the number of trials necessary to obtain a reliable measure.

Because both the FRN and ERN components have a high SNR, future studies that attempt to characterize the variability of these component at the single trial level and to characterize their relationships to behavior (e.g., in reinforcement learning) are feasible and highly desirable. On a more general note, the determination of the minimal number of trials needed for reliability seems to be a very useful exercise that should be extended to other ERP components. For example, there is an ongoing debate in the ERP literature about syntactic processing and the conditions that are required to elicit the so-called left anterior negativity component. Much of the variability in these experiments might be because many of the studies that failed to elicit the LAN did not include a sufficient number of trials.

REFERENCES

Cohen, J., & Polich, J. (1997). On the number of trials needed for P300. *International Journal of Psychophysiology*, *25*, 249–255.

Eppinger, B., & Kray, J. (2011). To choose or to avoid: Age differences in learning from positive and negative feedback. *Journal of Cognitive Neuroscience*, *23*, 41–52.

Eppinger, B., Kray, J., Mock, B., & Mecklinger, A. (2008). Better or worse than expected? Aging, learning, and the ERN. *Neuropsychologia*, *46*, 521–539.

Frank, M. J., Moustafa, A. A., Haughey, H. M., Curran, T., & Hutchison, K. E. (2007). Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proceedings of the National Academy of Sciences, USA*, *104*, 16311–16316.

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

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*, 679–709.

Holroyd, C. B., Pakzad-Vaezi, K. L., & Krigolson, O. E. (2008). The feedback correct-related positivity: Sensitivity of the event-related brain potential to unexpected positive feedback. *Psychophysiology*, *45*, 688–697.

Maidhof, C., Rieger, M., Prinz, W., & Koelsch, S. (2009). Nobody is perfect: ERP effects prior to performance errors in musicians indicate fast monitoring processes. *PLoS One*, *4*, e5032.

Marco-Pallares, J., Cucurell, D., Cunillera, T., García, R., Andrés-Pueyo, A., Münte, T., et al. (2008). Human oscillatory activity associated to reward processing in a gambling task. *Neuropsychologia*, *46*, 241–248.

Marco-Pallarés, J., Cucurell, D., Cunillera, T., Krämer, U. M., Camara, E., Nager, W., et al. (2009). Genetic variability in the dopamine system (DRD4, COMT) modulates neurophysiological responses to gains and losses. *Biological Psychiatry*, *66*, 154–161.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*, 177–190.

Miltner, W. H. R., Braun, C. H., & Coles, M. G. H. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task: Evidence for a "generic" neural system for error detection. *Journal of Cognitive Neuroscience*, *9*, 788–798.

Müller, S. V., Möller, J., Rodriguez-Fornells, A., & Münte, T. F. (2005). Brain potentials related to self-generated and external information used for performance monitoring. *Clinical Neurophysiology*, *116*, 63–74.

Nieuwenhuis, S., Ridderinkhof, K. R., Talsma, D., Coles, M. G., Holroyd, C. B., Kok, A., et al. (2002). A computational account of altered error processing in older age: Dopamine and the error-related negativity. *Cognitive, Affective and Behavioral Neuroscience*, *2*, 19–36.

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

Polich, J. (1986). P300 development from auditory stimuli. *Psychophysiology*, *23*, 590–597.

Pontifex, M. B., Scudder, M. R., Brown, M. L., O'Leary, K. C., Wu, C. T., Themanson, J. R., et al. (2010). On the number of trials necessary for stabilization of error-related brain activity across the life span. *Psychophysiology*, *47*, 767–773.

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*, 931–959.

(Received March 19, 2010; Accepted September 16, 2010)

[Correction added after online publication November 23, 2010: in the second-to-last paragraph (Interestingly, a recent...), the authors named the component studied FRN, when it was ERN. In the same paragraph, "these works" was changed to "this work". The corrected text appears here.]