universität
wien

Diplomarbeit

Titel der Arbeit

The impact of different positive and negative feedback stimuli on the FRN and P300: A synopsis of three event-related potential studies

Verfasser

Michael Zeiler

Angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag. rer. nat.)

Wien, im Mai 2012

Studienkennzahl: 298

Studienrichtung: Psychologie

Betreuer: Dr. Daniela Pfabigan

Danksagung

An dieser Stelle möchte ich mich bei einigen Personen bedanken, die zum Gelingen dieser Arbeit beigetragen haben.

Allen voran bedanke ich mich bei meinen Betreuerinnen Dr. Uta Sailer für das Angebot, die Diplomarbeit im Arbeitsbereich „Biologischen Psychologie“ zu schreiben und Dr. Daniela Pfabigan für die Unterstützung bei der Datenauswertung und die hilfreichen Kommentare und Korrekturvorschläge.

Des Weiteren bedanke ich mich bei den StudienassistentInnen im Arbeitsbereich „Biologische Psychologie“ für die Unterstützung bei der Datenauswertung.

Ich bedanke mich auch bei meinen KollegInnen aus dem Arbeitsbereich Methodenlehre für die gute Arbeitsgemeinschaft, die v.a. motivational zum Gelingen dieser Arbeit beigetragen hat.

Schließlich bedanke ich mich auch bei meinem Bruder Thomas für das Korrekturlesen, bei meinen Eltern und meiner Freundin Mirjam für die motivationale Unterstützung in jeder Hinsicht.

Tu erst das Notwendige, dann das Mögliche,
und plötzlich schaffst du das Unmögliche.

(Hl. Franz von Assisi)

Table of contents

ABSTRACT (ENGLISH)

Although the impact of different feedback stimuli on event-related brain potentials (ERPs) was frequently investigated, a systematic comparison of different feedback types within one study was rarely done in previous studies. The Feedback-related Negativity (FRN), a negative deflection approx. 250ms following feedback presentation, was associated with the feedback valence and the P300, a positive deflection approx. 300-600ms after stimulus presentation was primarily associated with feedback magnitude and salience. The Reinforcement Learning Theory (Holroyd, & Coles, 2002) suggests that the FRN reflects a binary evaluation of good vs. bad outcomes. It was further suggested that feedback valence and magnitude / salience are processed separately in the brain, the FRN reflecting the valence dimension and the P300 reflecting the magnitude / salience dimension. More recent studies demonstrated that this simplified classification of feedback valence and salience to FRN and P300 cannot be maintained. Affective and motivational aspects have been recently identified as moderating variables in the context of electrophysiological error processing.

Three studies were conducted in order to compare the influence of two feedback types (symbols and pictures of emotional facial expressions) on the FRN and P300. Participants received negative and positive feedback as a function of their performance in a time-estimation task. In Study 1 feedback in the form of symbols (plus sign, minus sign) whereas in Study 2 and 3 feedback in the form of symbols and faces (happy face, angry face) was presented. In Study 2 the feedback type changed randomly from trial to trial, whereas in Study 3 the feedback type remained stable within blocks of 50 trials. Different methods were used to quantify FRN and P300 amplitudes.

Statistical analysis revealed that the FRN amplitude was larger in the negative feedback conditions and - only in Study 2 - after symbols compared to faces. The P300 amplitude was larger in the positive and facial feedback conditions. The FRN effect sizes significantly differed between Study 2 and 3 which indicates that there is a meaningful impact of the feedback presentation mode on the FRN. Different methods of ERP quantification were differentially suitable for obtaining the FRN amplitude. They further differed in respect of whether a statistical result reached statistical significance or not, but the used methods didn't differ significantly in respect of the elicited effect sizes.

The FRN is interpreted as an error signal that reflects a classification of good vs. bad outcomes. This classification seems to be complicated when the feedback type changes frequently within an experiment. The P300 results provide evidence that feedback valence and salience are not processed by strictly separate neural mechanism in the brain. A blocked presentation mode of different feedback types is recommended for further studies. A synopsis of different ERP quantification methods can avoid possible misinterpretations and enhance the validity of conclusions drawn by statistical results.

ABSTRACT (DEUTSCH)

Obwohl der Einfluss verschiedener Feedback-Stimuli auf ereigniskorrelierte Potentiale (EKPs) bereits vielfach untersucht wurde, sind systematische Vergleiche mehrerer Feedback-Arten innerhalb einer Studie selten. Die Feedback-related Negativity (FRN), ein EKP mit negativem Amplitudenmaximum ca. 250ms nach Feedback Präsentation, wurde mit der Valenz des Feedbacks in Zusammenhang gebracht, während die P300, ein EKP mit positivem Amplitudenmaximum ca. 300-600ms nach Vorgabe des Feedbacks, mit der Feedback Größe bzw. Salienz assoziiert wurde. Laut der Reinforcement Learning Theory (Holroyd, & Coles, 2002) reflektiert die FRN eine binäre Bewertung von Ergebnisses in gut vs. schlecht. Weiters wurde vermutet, dass Feedback Valenz und Größe / Salienz separat im Gehirn verarbeitet werden, wobei die FRN die Valenz-Dimension und die P300 die Salienz-Dimension reflektiert. Kürzlich publizierte Studien lassen allerdings an einer vereinfachten Zuordnung von Feedback Valenz und Salienz zu FRN und P300 zweifeln. Affektive und motivationale Aspekte haben sich kürzlich als Moderatorvariablen im Kontext elektrophysiologischer Fehlerverarbeitung erwiesen.

Es wurden drei Studien durchgeführt, in denen der Einfluss zweier Feedback Arten (Symbole und Bilder von emotionalen Gesichtsausdrücken) auf FRN und P300 verglichen wurde. Die Versuchspersonen erhielten negatives und positives Feedback in Abhängigkeit von deren Leistung in einer Zeitschätzungsaufgabe. In Studie 1 wurde Feedback in Form von Symbolen (Plus-Symbol, Minus-Symbol) präsentiert, während in Studie 2 und 3 Feedback in Form von Symbolen und Gesichtern (glücklicher Gesichtsausdruck, verärgerter Gesichtsausdruck) gegeben wurde. In Studie 2 wechselte die Feedback Art zufällig von Durchgang zu Durchgang, während in Studie 3 die Feedback Art innerhalb von Blöcken bestehend aus 50 Durchgängen konstant blieb. Mehrere unterschiedliche Methoden wurden für die Quantifizierung der FRN und P300 angewandt.

Die Ergebnisse zeigten größere FRN Amplituden nach negativem Feedback, sowie – nur in Studie 2 - nach Symbol-Feedback verglichen mit Gesichter-Feedback. Die P300 Amplituden waren größer nach positivem, sowie nach Gesichter-Feedback. Die

Effektstärken der FRN-Haupteffekte unterschieden sich signifikant zwischen Studie 2 und 3, was darauf schließen lässt, dass die Art der Feedbackvorgabe einen bedeutsamen Einfluss auf die FRN hat. Verschiedene Methoden zur EKP-Quantifizierung haben sich als unterschiedlich geeignet für die FRN-Bestimmung erwiesen. Auf Ebene der einzelnen Studien wurden verschiedene Ergebnisse in Abhängigkeit von der verwendeten Methode erzielt, jedoch unterschieden sich die verwendeten Methoden nicht signifikant im Hinblick auf die hervorgerufenen Effektstärken.

Die FRN wird als Fehler-Signal interpretiert, welches eine Klassifikation von Ereignissen in gut vs. schlecht reflektiert. Jedoch scheint diese Klassifikation erschwert, wenn die Feedback-Art häufig innerhalb eines Experiments wechselt. Vor allem die Ergebnisse bezüglich der P300 weisen darauf hin, dass Feedback Valenz und Salienz nicht durch separate neuronale Mechanismen verarbeitet werden. Für zukünftige Studien kann ein geblockter Vorgabemodus bei dem Vergleich mehrerer Feedback Arten empfohlen werden. Die Anwendung mehrerer Methoden der EKP-Quantifizierung kann möglichen Fehlinterpretationen vorbeugen, sowie die Validität von Schlussfolgerungen, die aufgrund statistischer Ergebnisse getroffen werden, erhöhen.

1. INTRODUCTION

1.1 Feedback processing in Event-Related-Potential-Research

In the 1970s, the P300 component (Sutton, Tueting, Hammer, & Hakerem, 1978), one of the most famous event-related-potentials (ERPs), was first associated with feedback indicating a correct or incorrect response and the outcome probability. Since that time, feedback processing has grown to a tremendous research area in cognitive, social and affective neuroscience. Due to the high temporal resolution of the electroencephalogram (EEG), ERPs were often used to investigate the early electrophysiological response to presented feedback stimuli. Apart from the P300 a second ERP component, the Feedback-Related-Negativity (FRN), discovered in 1997 (Miltner, Braun, & Coles, 1997) has aroused a great deal of interest up to today. A lot of influencing variables on the FRN and P300 were discussed, theories about their neural basis were constructed and the connection between FRN and P300 effects in the context of feedback processing was investigated so far.

1.2 The Feedback-Related-Negativity (FRN)

1.2.1 Discovery and general characteristics of the FRN

The FRN is a negative ERP deflection which occurs approximately 250ms to 300ms after the presentation of unfavorable feedback stimuli and which reaches its maximum at frontocentral electrode sites (Holroyd, & Coles, 2002; Miltner et al., 1997; Yeung, & Sanfey, 2004). The FRN was first described by Miltner et al. (1997). Their participants received feedback as a function of their performance on a time estimation task. Depending on the accuracy of their time estimation, they either received positive or negative feedback in terms of visual, auditory, or somatosensory stimuli. A negative ERP deflection, the FRN, occurred and was larger following the negative feedback conditions compared to the positive feedback conditions. This effect was independent from the feedback modality. Equivalent dipole analysis indicated that the anterior cingulate cortex (ACC) was the source of the FRN.

Such a negative deflection following negative feedback stimuli was subsequently observed across many experiments (e.g. Cohen, Elger, & Ranganath, 2007; Donkers, Nieuwenhuis, & van Boxtel, 2005; Gehring, & Willoughby, 2002).

1.2.2 The FRN and its relation to the Error-Related-Negativity (ERN)

An Error-Related-Negativity (ERN), a negative-going component, is evoked when participants make an erroneous response and reaches its maximum at frontocentral electrode sites approximately 50-100ms after the incorrect response. It is absent or smaller when the response is correct (Falkenstein, Hohnsbein, Hoormann, & Blanke, 1990; Gehring, Goss, Coles, Meyer, & Donchin, 1993). A typical experimental task used for eliciting an ERN is the Erikson Flanker Task, where subjects have to make fast and accurate choice reactions (Gehring et al., 1993; Potts, 2011)[1]. Due to the speed of this task, motoric errors in some trials are presumable. An erroneous response is generally detected at the moment of the button press by the participants themselves. Gehring et al. (1993) were the first to suggest that the ERN was elicited by an underlying neural system for error detection and error compensation. As the FRN, the ERN was localized in the ACC (Dehaene, Posner, & Tucker, 1994). Due to the similarity of the response-locked ERN and the feedback-driven FRN - both components were negative-going and were localized in the ACC - it was concluded that there is one basic neural system for error detection. Thus, an ERN was elicited when errors were committed and an FRN when a feedback stimulus signaled an incorrect response. A functional interrelationship between ERN and FRN was also proposed by Heldmann, Rüsseler and Münte (2008) showing that either the ERN or the FRN were released dependent on whether the most salient information about an incorrect response came from an internal self-generated error signal or from an external feedback cue.

[1] Erikson Flanker Task: In a typical version of the Eriksan Flanker Task (see Gehring et al., 1993), arrays of letters are presented. There are compatible arrays (e.g. HHHHH and SSSSS) and incompatible arrays (e.g. SSHSS and HHSHH). Subjects have to respond with the left or the right hand dependent on the central letter (H vs. S).

1.2.3 Anterior cingulate cortex (ACC) and the Reinforcement learning theory

As mentioned above, the ACC was frequently identified as the source of the ERN (Dehaene et al., 1994) and the FRN (Miltner et al., 1997; Ruchsow, Grothe, Spitzer, & Kiefer, 2002) by dipole analysis. Studies using functional magnetic resonance imaging (fMRI) delivered further evidence for the ACC area as the most likely source of the ERN and FRN (Holroyd et al., 2004; Ullsperger, & von Cramon, 2003).

At this point, it has to be remarked that the ACC is equitable to the anterior midcingulate cortex (aMCC). It was recently found that these two areas of the cingulate cortex overlap according to their cognitive and affective functions (Shackman et al., 2011).

The Reinforcement Learning Theory (Holroyd, & Coles, 2002) links the ACC, as a system for error detection, with the mesencephalic dopamine system jointly responsible for reinforcement learning. This theory supposes that the basal ganglia increase or decrease the activation of dopaminergic neurons in the mesencephalon as a function of whether an event is evaluated as better or worse than expected. A decrease in the activation is associated with a "worse-than-expected"-signal which is sent to the ACC. According to the Reinforcement Learning Theory, this signal is reflected in the ERN and FRN. Due to the fact that the ACC was identified as an area where motor intentions are transformed into motor responses, the Reinforcement Learning Theory proposes that the reinforcement learning signal from the midbrain dopamine system is used by the ACC for behavioral adaption and modification. Nieuwenhuis, Holroyd, Mol and Coles (2004) postulated four predictions that could be made from the Reinforcement Learning Theory. First, FRN amplitude variation should reflect a simple good vs. bad evaluation of an event. Secondly, FRN amplitude variation should reflect the mismatch between actual and expected outcome. Thirdly, learning should be a moderator of the relationship between ERN and FRN amplitudes. FRN amplitudes should decrease and ERN amplitudes increase during the continuation of a learning process because a person should be more and more able to detect inappropriate behavior from the response itself instead of an explicit external feedback stimulus. Fourthly, the FRN should be generated in the ACC. Much effort was expended in order to examine this Reinforcement Learning Theory and the role of the ACC in the context of electrophysiological feedback processing. Although the majority of studies using dipole analysis for source localization of the FRN revealed that the ACC is the most likely generator of the FRN, other studies, primarily using fMRI failed to find stronger activation of the ACC in negative feedback compared to positive feedback conditions (Nieuwenhuis, Slagter, von Geusau, Heslenfeld, & Holroyd, 2005; Van der Veen, Röder, Mies, Van der Lugt, & Smits, 2011; Van Veen, Holroyd, Cohen, Stenger, & Carter, 2004).

1.2.4 Discussion of the some moderating variables concerning ERN/FRN effects

Gambling Tasks are frequently used experimental paradigms to investigate the impact of moderating variables on FRN amplitudes. In such a typical task, several stimuli are presented simultaneously and participants have to decide which of those stimuli is associated with a monetary reward (e.g. Hajcak, Moser, Holroyd, & Simons, 2006). After choosing one stimulus, the participants are informed by a feedback cue whether their choice was good or bad, and whether they earn or loose money in the current trial. A monetary reward serves as positive feedback and a monetary loss as negative feedback. The magnitude of a monetary reward or punishment is often varied and represents the salience of the feedback. The prospect theory (Kahneman, & Tversky, 1979) provides evidence for this assumption by demonstrating that higher wins and losses have a greater subjective value compared to small wins and losses.

From now on, the term “feedback magnitude” is used when the feedback salience is varied by different amounts of monetary wins and losses and the term “feedback salience” is used, when the salience of a feedback stimulus is varied by different types of feedback stimuli.

The majority of those studies including reward / punishment magnitude (small vs. large monetary wins and losses) as a separate factor demonstrated that FRN amplitudes were only sensitive to the feedback valence but insensitive to the magnitude of a loss or win. It was concluded that the FRN amplitude reflected an early and basic classification of good vs. bad outcomes (Hajcak, Moser, Holroyd, & Simons, 2006; Yeung, & Sanfey, 2004). Bellebaum, Polezzi and Daum (2010) reported larger FRN amplitudes after big losses compared to small losses. The authors interpreted this result as a violation of reward expectation. Some recent studies also reported a significant effect of reward magnitude on FRN amplitudes towards larger amplitudes after big losses (Kreussel et al., 2012; Santesso, Dzyundzyak, & Segalowitz, 2011). These results suggest that the FRN might reflect more than a simple good vs. bad distinction of events. As predicted by the Reinforcement Learning Theory (Holroyd, & Coles, 2002), the subjective expectation whether an outcome will be positive or negative could be an important moderating process on the FRN effect. This assumption is supported by studies ascertaining that even neutral feedback, neither a win nor a loss, could evoke an FRN if the neutral feedback is perceived as negative (Hajcak et al., 2006; Holroyd, Hajcak, & Larsen, 2006). This implies that not the objective but the subjective perception of a feedback stimulus is crucial and that the FRN amplitude variation classifies an event as satisfying or not satisfying. Outcome expectations can also be formed by the manipulation of the frequency of negative and positive feedback. The majority of studies reported a larger FRN amplitude for infrequent negative feedback compared to frequently given negative feedback (Gibson, Krigolson, & Holroyd, 2006; Holroyd, Nieuwenhuis, Yeung, & Cohen, 2003; Liao, Gramann, Feng, Deak, & Li, 2011; Pfabigan, Alexopoulos, Bauer, & Sailer, 2011).

Another frequently discussed question in the literature is whether an active choice behavior is absolutely required for eliciting an FRN. It was recently shown that an FRN was also elicited when subjects observed a virtual person making choices and receiving feedback and no action was performed by the subjects themselves (Kobza, Thoma, Daum, & Bellebaum, 2011). Yu and Zhou (2006) found that an FRN-like peak was elicited after cues predicting a possible loss in the following trial. However, this negative deflection was smaller compared to the FRN amplitude following negative feedback after a response choice that indicated an actual loss. Similar results were presented by Yeung, Holroyd and Cohen (2005). A reduced FRN appeared in conditions where no active choices and overt actions had to be made. The authors explained the reduced FRN amplitudes in no-choice tasks compared to response choice tasks by the higher motivational significance of tasks with an active response. These results indicate that an active behavior is not essential for the appearance of an FRN component. Yet, the personal task involvement seems to be an important moderating variable.

Recently, the influence of personality traits and affective state and trait aspects on error-related brain potentials is increasingly discussed. Several findings indicate that personality traits should get more attention in FRN research. Individuals who scored highly on trait neuroticism showed an enhanced FRN amplitude in negative feedback conditions (Compton et al., 2010), for uncertain feedback (Hirsh, & Inzlicht, 2008) and for ambiguous outcomes (Gu, Ge, Jiang, & Luo, 2010). FRN amplitude variations were also observed in respect of the individual level of extraversion towards larger amplitudes following unpredicted non-rewards compared to unpredicted rewards (Smillie, Cooper, & Pickering, 2011). Furthermore, personal responsibility for task outcomes was positively correlated with the FRN amplitude (Li, Han, Lei, Holroyd, & Li, 2011), the self-reported sensitivity towards punishment was identified as an influencing factor (Santesso et al., 2011), and the FRN amplitude was partly associated with antisocial traits (Pfabigan, Alexopoulos, Bauer, Lamm, & Sailer, 2011). The assumption that emotional aspects could have an impact on error-related ERPs is supported by studies showing that the ACC is a part of a neuronal circuit regulating cognitive and affective processes (Bush, Luu, & Posner, 2000). A negative affect trait, meaning a general individual tendency towards negative emotions which was ascertained with the trait version of the PANAS-scales (Watson, Clark, & Tellegen, 1988), was significantly correlated with a larger ERN amplitude with $r = .46$ (Luu, Collins, & Tucker, 2000) and with a larger FRN amplitude with $r = .49$ (Sato et al., 2005). Larger ERN amplitudes were also observed when subjects' negative affective state was induced by derogatory verbal feedback, whereas encouraging verbal feedback had no influence on the ERN (Wiswede, Münte, & Rüsseler, 2009). In another study, a negative or positive affect was induced by the presentation of IAPS pictures with negative or positive valance ratings (Lang, Bradley, & Cuthbert, 1999) preceding the presentation of each flanker stimulus (Wiswede, Münte, Goschke, & Rüsseler, 2009). The presentation of negatively valenced IAPS pictures evoked larger ERN amplitudes in incorrect trials. These studies suggest that negative affective personality traits as well as an induced negative affect enhance the salience of an erroneous response or feedback stimulus which is reflected in a larger ERN or FRN amplitude.

1.3 The P300 in context of feedback processing

The P300 component is a positive ERP deflection which occurs approximately 300 to 600ms after stimulus presentation and which reaches its maximum at parietal-central electrode sites.

Since the discovery of the P300 by Sutton, Braren, Zubin and John (1965) being more enhanced after uncertain compared to certain stimuli, the dimension of stimulus uncertainty, stimulus expectancy respectively, remained one of the most discussed questions; also in connection with feedback processing until today. A larger P300 amplitude was found for unexpected compared to expected feedback (Hajcak, Holroyd, Moser, & Simons, 2005; Hajcak, Moser, Holroyd, & Simons, 2007; Pfabigan et al., 2011b). However, another study failed to observe a P300 difference in regard to feedback expectancy (Wu, & Zhou, 2009). These authors proposed that the expectancy dimension might have already been coded by the earlier FRN.

An aspect that was consistently associated with the P300 amplitude variation was the magnitude of a reward or non-reward. In experiments where the participants were able to earn or loose different amounts of money, a main effect of feedback magnitude towards a larger P300 amplitude for larger amounts of money was consistently observed (Bellebaum et al., 2010; Moser, Hajcak, Holroyd, & Simons, 2004; Sato et al., 2005; Wu, & Zhou, 2009). It could be argued that a larger amount of money is equivalent with a higher salience of the feedback, no matter whether the outcome is positive or negative. This point of view is supported by Polezzi, Sartori, Rumiati, Vidotto and Daum (2010) who argued that outcomes with a higher magnitude have a greater motivational significance.

Findings concerning the influence of feedback valence on the P300 are much more inconsistent. On the one hand, some studies reported larger P300 amplitudes after positive compared to negative feedback (Bellebaum et al., 2010; Hajcak et al., 2007). On the other hand, no P300 differences between positive and negative feedback was found in other experiments (Sato et al., 2005; Yeung, & Sanfey, 2004). Moser et al.

(2004) gave the hint that the valence of feedback had only an effect when there were more than two possible outcomes. The authors observed that the feedback valence had no influence on P300 amplitudes whenever there were just one positive and one negative outcome. Only after introducing more categories (i.e. two types of positive and two types of negative feedback stimuli), the feedback valence affected the P300 amplitudes towards larger amplitudes in the positive compared to the negative feedback conditions. Although the question remains whether feedback valence impacts P300 amplitude, the authors showed that the experimental context should receive more attention concerning this question.

Motivational and affective issues have to be discussed in context of P300 amplitude variations as well. As discussed for the FRN, subjects' involvement in the task, their motivation respectively, also enhances the P300 amplitude (Yeung et al., 2005). Experiments where emotional pictures were presented indicate that the P300 amplitude reflects the affective significance of stimuli. A larger P300 amplitude was observed when emotionally charged pictures were presented in contrast to emotional neutral pictures, whereby there was no difference in the P300 amplitude between positive and negative pictures (Briggs, & Martin, 2009; Keil et al., 2002). Although Briggs and Martin (2009) found that unpleasant pictures were associated with a larger P300 amplitude, the P300 effect seemed not to be caused by simple arousal effects. As Wu and Zhou (2009) discussed, negative feedback stimuli tend to elicit a higher level of arousal, If P300 amplitude variation reflects arousal, negative feedback stimuli, should be associated with a greater P300 response. The tendency of larger P300 amplitudes after positive feedback observed in several studies (Bellebaum et al., 2010; Hajcak et al., 2007) is an argument against this arousal theory. Due to these inconsistent findings which could not be resolved by considerations about affect and arousal, the question concerning the influence of feedback valence on the P300 amplitude is still unresolved.

Some studies found an effect of feedback valence on the P300 latency. The latency was prolonged in negative feedback conditions compared to the positive feedback conditions (Pfabigan et al., 2011a; Yeung et al., 2005). Due to the fact that the P300 latency is, among other things, a measure of stimulus evaluation speed (Polich, 2007), it can be argued that negative feedback stimuli are processed more thoroughly in the brain.

1.4 Separate vs. common neural processing of feedback valence and feedback magnitude / salience

Considering both the FRN and P300, there arises the question whether these two components code the same or different aspects of an outcome. This point was discussed in respect of the valence and magnitude or salience of feedback. Yeung and Sanfey (2004) proposed that feedback valence and magnitude are processed separately in the brain. According to this, the FRN should reflect the valence but not the magnitude of the feedback; and on the contrary the P300 should reflect the feedback magnitude but should be independent from the feedback valence. This theory is called the “independent coding model” and has become popular in event-related potential research. This model is supported by studies reporting that the FRN was only sensitive to the feedback valence dimension, but not to the magnitude dimension (e.g. Hajcak et al., 2006). Other studies point against a simplified classification of feedback valence to the FRN, and feedback magnitude to the P300. As already discussed, the P300 was influenced by the feedback valence in some cases (Bellebaum et al., 2010; Hajcak et al., 2007), and there were also a few studies showing an effect of feedback magnitude on the FRN (Santesso et al., 2011). Goyer, Woldorff and Huettel (2008) demonstrated that the FRN effect was not only a function of feedback valence but also a function of the combination of multiple factors including valence, magnitude, outcome alternatives, and the prior outcome history. An interesting study was recently conducted by Gu et al. (2011). Valence and magnitude information of the feedback stimuli were presented sequentially and the respective impact on FRN and P300 amplitudes was examined. The results revealed that the FRN and the P300 could be sensitive to both, to the valence as well as to the magnitude information of the feedback stimuli. As the authors discussed, feedback valence and magnitude are coded in a flexible way depending on which information is provided. Studies including affective or motivational cues (Wiswede et al., 2009a) which are expected to enhance the salience of feedback stimuli, provide further evidence for a flexible coding of feedback valence and magnitude, and feedback salience respectively. However, further investigations to address this question are needed.

1.5 Feedback stimuli used in previous ERP studies

A view on usually presented visual feedback stimuli in previous ERP studies reveals that most stimuli were simple, non-complex symbols, like two letters; whereby one letter represents positive and the other one negative feedback, like a plus and a minus symbol or like a number indicating the amount of money that is won or lost. Only a few studies presented other, more complex, feedback stimuli like words or pictures of facial expressions. A short list of frequently used stimuli in feedback research is presented in Table 1.

Table 1: Examples of used feedback stimuli in previous ERP studies

Feedback stimuli	Examples of publications
x, o	Miltner et al. (1997), Sato et al. (2005)
x, o, -	Holroyd et al. (2006)
+, -	Hajcak et al. (2007)
++, +, -, --	Hajcak et al. (2006)
$$$$, XXXX	Holroyd et al. (2006)
+10, -5, 0	Yu, & Zhou (2005)
+25, +5, -25, -5	Wu, & Zhou (2009)
picture of a 5 cent coin, no coin (empty box)	Bellebaum et al. (2008)
“correct”, “incorrect”	Kobza et al. (2011)
smiling face, crying face	Li et al. (2011)
happy face, angry face	Pfabigan et al. (2011a)

Considering that the great majority of results concerning feedback-related brain potentials were achieved by the usage of non-complex symbols as feedback stimuli, little is yet known about the effects of other types of feedback stimuli, like words or faces. The quantity of different feedback types raises various questions we have to consider: It has to be considered whether the valence of a feedback stimulus can be assessed directly from the stimulus itself or whether the meaning has to be learned during the experiment. Furthermore, the emotional and motivational significance of a feedback stimulus has to be discussed. In the case of single letters (e.g., “x”, “o”) the participants have to learn which letter represents a positive feedback and which represents a negative feedback. These stimuli do not provide any valence information themselves. In contrast, the meaning of a plus- and a minus-symbol, of words like “correct” and “incorrect”, and of the presented amount of money that is won or lost, can be assessed directly from the stimulus. As described by Ekman and Friesen (1971), basic facial expression can be recognized directly and universally. Streit, Wölwer, Brinkmeyer, Ihl and Gaebel (2000) found that the decoding of facial expressions starts very early (approximately 120-170ms following stimulus presentation) in the brain. Further, it could be argued that the presentation of a facial expression has a greater emotional significance than the presentation of simple letters or symbols. Van der Veen et al. (2011) suggested that the presentation of faces should elicit a greater activation in brain areas associated with feedback due to their greater emotional value compared to verbal feedback. Furthermore, more attentional resources are allocated to negative compared to positive facial expression which is reflected in prolonged P300 latencies (Yamamoto et al., 2000).

So far, little is known about the impact of different types of feedback stimuli on feedback-related brain potentials. However, the varying visual features and emotional significance of different feedback types give reason to expect differences in FRN and P300 amplitudes and latencies.

1.6 Commonly used methods for the quantification of FRN and P300

There are many ways to quantify ERP amplitudes and several methods were used so far to obtain FRN and P300 amplitude values. In the recent literature two, respectively three general groups of methods of ERP quantification are described: Peak amplitudes, mean amplitudes, and difference waves. While peak amplitudes are point measures, mean amplitudes use the average of all time points within a time range to quantify ERPs. For example, the *base-to-peak* method measures the amplitude at the point when a component reaches its maximum within a defined time window in comparison to a pre-defined baseline interval. For example, the FRN was defined as the most negative deflection between 200ms and 400ms by Gu et al. (2011). As the amplitude is measured relative to its baseline, a neutral baseline which is stable across the conditions is required for a valid measure (Seifert, 2005). Furthermore, the pre-stimulus baseline is often influenced by the CNV (Contingent Negative Variation)[2] which reflects the preparation for a stimulus (Hoormann, Falkenstein, Schwarzenau, & Hohnsbein, 1998). This method is at a disadvantage when peaks cannot be identified clearly because of noise. Thus, it should only be used for well-defined peaks (Handy, 2005). It is not surprising that the P300, as one of the largest components, was mostly quantified base-to-peak in previous feedback-related ERP studies (e.g. Hajcak et al., 2005, 2007; Jia et al., 2007). Only a few studies quantified the FRN, a smaller component which is steeply rising and that is sometimes absent in the positive feedback conditions, with the base-to-peak method (Donkers et al., 2005; Gu et al., 2011). A way to overcome possible confounding artifacts in the pre-stimulus baseline is the *peak-to-peak* method where the interesting peak is compared to an adjacent peak and the amplitude difference is calculated (Handy, 2005). However, the reference peak should be uninfluenced by the experimental manipulation, or should reflect the same underlying process. To the contrary, if the reference peak varies between the conditions, the calculated difference value would be confounded. Therefore, the choice of an appropriate reference is important. In several studies the FRN was related to the preceding positive peak (e.g. Bellebaum et al., 2010, Pfabigan et al., 2011a; 2011b; Sato et al., 2005). Using this method, there is the possibility to set the amplitude value at zero in the case of absent peaks (Hajcak et al., 2006). Yeung and Sanfey (2004) and Sailer, Fischmeister and Bauer (2011) used a modified peak-to-peak method where the difference between the FRN amplitude and the average of the preceding and following positive peak was calculated ("peak-to-peak-to-peak"). *Mean amplitudes* calculate the arithmetic average of all time points within a defined time window. This method has an advantage over peak amplitudes when the peaks cannot be identified clearly (Handy, 2005). They generally have a greater reliability as single peak measures, but should only be used for rather flat potentials (Seifert, 2005). Mean amplitudes were frequently used for the FRN quantification (e.g. Heldemann et al., 2008; Jia et al., 2007; Yeung et al., 2004; Yu, &

[2] The Contigent Negative Variation (CNV) is a negative-going component which occurs between a warning signal and a target stimulus. It reflects the preparation for the target, respectively the preparation for a motoric action that is required as a response to the target stimulus. The CNV is a long-lasting component which reaches its maximum at frontal electrode sites (Luck, 2005).

Zhou, 2006) and rarely used for the P300 (e.g. Polezzi et al., 2010). *Difference waves* are an appropriate alternative when two conditions should be compared. Using this method, the ERPs of two conditions are subtracted from each other and the resulting difference waveshape is analyzed by defining peaks or calculation mean amplitudes. As Hoormann et al. (1998) described, difference waves "suppress the ERP activity that is constant across two conditions" (p.106). In a certain manner, difference waves could also overcome the problem of component overlap (Seifert, 2005). Some studies used this method for the FRN quantification (Hajcak et al., 2007; Holroyd, & Krigolson, 2007; Nieuwenhuis et al., 2004).

To sum up, while the P300 was quantified with the base-to-peak method in the great majority of feedback-related ERP studies, various methods were used for the FRN amplitude quantification. The variability of the methods used raises the question whether or not they all are equally suitable for quantifying the FRN. Interestingly, the majority of authors neglected to provide any explicit explanation regarding why the respective method was applied.

1.7 Own research outline and hypothesis

Many ERP studies were conducted with several different feedback stimuli so far, but there were no ERP studies performed so far directly comparing the impact of different feedback types within one study. One aim of the present study is to compare two completely different visual feedback types in respect of their impact on FRN and P300 amplitudes. Simple symbols, a plus- and a minus sign, will be contrasted to facial feedback stimuli, a happy looking face and an angry looking face. The feedback will be provided as a function of the participants' accuracy in a time-estimation task (Miltner et al., 1997). This study will be performed in three steps, sub-studies respectively. In the first study, only feedback in terms of symbols is presented; in Studies 2 and 3 both feedback types, symbols and faces, are used. Since there is no experience in comparing two feedback types within one experiment up until now, two feedback presentation modes are tested, a random (Study 2) vs. a blocked presentation mode (Study 3).

Referring to the consistent findings that negative feedback is associated with a larger FRN, the first hypothesis is: *The FRN amplitude is larger in the negative feedback conditions compared to the positive feedback conditions regardless of the feedback type.*

The Reinforcement Learning Theory (Holroyd, & Coles, 2002) predicts that the FRN should reflect a general classification of good vs. bad outcomes. Consequently the FRN effect should be unaffected by the feedback type. Nevertheless, recent studies demonstrated that the salience of feedback, a negative affective trait (Sato et al., 2005; Lu et al., 2000), as well as experimental induced affect (Wiswede et al., 2009a, 2009b) had an impact on error-related brain potentials. Since it is assumed that emotional facial expressions enhance the affective salience of the feedback, the second hypothesis is: *The FRN amplitude is larger in the facial feedback conditions compared to the feedback symbols regardless of the feedback valence.*

Previous studies consistently showed that the feedback salience had a significant impact on the P300 amplitude. The third hypothesis is: *The P300 amplitude is larger following facial feedback compared to symbols regardless of the feedback valence*.

Concerning the influence of the feedback valence on the P300, there are controversial results in literature, but there seem to be a tendency for a larger P300 amplitude after positive feedback. Therefore the forth hypothesis is: *The P300 amplitude is larger in the positive feedback conditions compared to the negative feedback conditions.*

Preliminary findings indicate that negative feedback (Pfabigan et al., 2011b; Yeung et al., 2005) and facial stimuli (Yamamoto et al., 2000) attract larger attention and are processed more thoroughly in the brain which is reflected in a longer P300 latency. Thus, the fifth hypothesis is: *The P300 latency is longer in the negative and facial feedback conditions compared to the positive and symbol feedback conditions.*

A synopsis of the FRN and P300 effects can make an important contribution to the question whether feedback valence and salience are processed separately in the brain (Yeung, & Sanfey, 2004), or whether the feedback valence and salience are reflected by both, the FRN and P300 (Goyer et al. 2008; Gu et al., 2011). The former would be confirmed if the FRN is only affected by the valence but not by the feedback type, and the P300 is only affected by the feedback type but not by the feedback valence.

Another aim of this study is to assess the FRN and P300 amplitudes by different methods, peak measures and mean amplitudes, in order to enhance the reliability and validity of the ERP analysis and in order to compare the results obtained by the different quantification methods. As described above, the frequently used methods to assess the FRN differ from each other in respect of which data points they include (just one single peak amplitude, two or more peak amplitudes, all data points within a time range, etc.). Therefore the sixth hypothesis is: *Different methods for FRN and P300 quantification produce significantly different results*.

Studies 2 and 3 will be compared in order to investigate if there is a difference between a random or blocked presentation mode of the two types of feedback with regard to the FRN and P300 effect sizes. The seventh hypothesis - of an explorative nature - is: *There is a difference between Study 2 and 3 in the FRN and P300 effect sizes.*

Additionally, the behavioral performance of the participants in the time estimation task will be analyzed. Previous studies applying the same experimental task found that subjects used the feedback to guide their subsequent time estimation. Miltner et al. (1997), Van Veen et al. (2004), and Van der Veen et al. (2011) demonstrated that the change in time estimation was larger after negative feedback than after positive feedback. The 9th hypothesis is: *The absolute change in time estimation is larger following negative feedback than following positive feedback, and there are more correct adjustments in time estimation after negative than after positive feedback.*

An additional hypothesis was created after statistical calculations were done because of the variability of FRN effect sizes and behavioral performance concerning Study 2 and 3. Since our findings suggest that the behavioral performance is associated with the FRN amplitude, the additional hypothesis *“The individual FRN amplitude is correlated with the behavioral performance.”* is constructed. The ACC as a brain area responsible for behavioral modifications (Holroyd, & Coles, 2002) emphasizes the validity of this hypothesis.

2. SHORT OVERVIEW OF THE CONDUCTED STUDIES

The three studies were conducted at the brain research laboratory of the SCAN-Unit (Social Cognitive and Affective Neuroscience - Unit) at the Faculty of Psychology (University of Vienna) in 2008, 2009 and 2010 during the course “Forschungspraktikum II”. Table 2 provides a summary of the most important studies’ characteristics.

Table 2: Characteristics of Study 1, 2 and 3

	Study 1	Study 2	Study 3
Year of implementation	2008	2009	2010
Sample			
- male / female	4 / 11	9 / 11	6 / 11
- mean age	25.31 ± 3.13	25.80 ± 3.91	26.00 ± 6.43
- age range	20 - 31	21 - 35	20 - 48
- handedness	all right-handed	all right-handed	all right-handed
Feedback-stimuli	symbols	symbols & faces	symbols & faces
Number of trials	8 x 50 trials aprrox. 50% “+” sign, 50 % “-“ sign	8 x 50 trials approx. 25% “+” sign, 25% “-“ sign, 25% happy face, 25% angry face	8 x 50 trials approx. 25% “+” sign, 25% “-“ sign, 25% happy face, 25% angry face
Feedback presentation mode		symbols and faces changed randomly from trial to trial	blocked (in each block of 50 trials either symbols or faces)
Number of electrodes used for EEG-recording	62 whole skull	6 only central electrodes: Fz, FCz, Cz, CPz, Pz, Oz	6 only central electrodes: Fz, FCz, Cz, CPz, Pz, Oz
Statistical analysis	Repeated measures ANOVA Main effect of feedback valence	Repeated measures ANOVA Main effect of feedback valence, feedback type and interaction effect	Repeated measures ANOVA Main effect of feedback valence, feedback type and interaction effect

3. STUDY 1

3.1 METHODS

3.1.1 Participants

Fifteen right-handed participants (four male, eleven female) with a mean age of 25.31 (SD = 3.13) took part in this experiment. They were recruited by students taking part in an University course (diploma level) concerning EEG-measurement. No monetary remuneration was provided for their participation in this experiment. Written informed consent was obtained from all participants and handedness was assessed by the German version of the Edinburgh Handedness Inventory (Oldfield, 1971). All participants reported no history of any neurological or cardiovascular diseases and severe head injuries. No participant was on regular medication. The study was conducted in accordance with the Declaration of Helsinki and local ethical guidelines of the University of Vienna. One participant had to be excluded from further ERP-analysis due to large noise in the EEG signal. Since the task performance of this participant was in a normal range, the behavioral data were analyzed for the whole sample.

3.1.2 Task

A time-estimation task as used by Miltner et al. (1997) was applied. The participants had to estimate the duration of a one-second-interval after the presentation of a cue ("x") and to press a button when they guessed the one-second-interval was over. After the button press the participants were informed about the accuracy of their time-estimation by a feedback cue which was presented at the screen for one second. Each feedback cue was followed by a fixation cross before the next trial started. Figure 1 presents the timeline of one trial. A plus-sign served as positive feedback informing that the estimation was accurate enough whereas a minus-sign represented negative feedback meaning that the estimation was too inaccurate. The feedback stimuli were presented in the middle of the screen on grey background (Figure 2). The valence of feedback (positive or negative) was a function of whether the time-estimation fell into a defined time-window or not. The initial time-window for positive feedback was 900ms – 1100ms. This time-window was dynamically adjusted dependent on the participants' performance on the preceding trial. After negative feedback, the time-window was increased by ± 10ms and after positive feedback it was decreased by ± 10ms. Participants received approximately 50% positive and 50% negative feedback using this procedure.

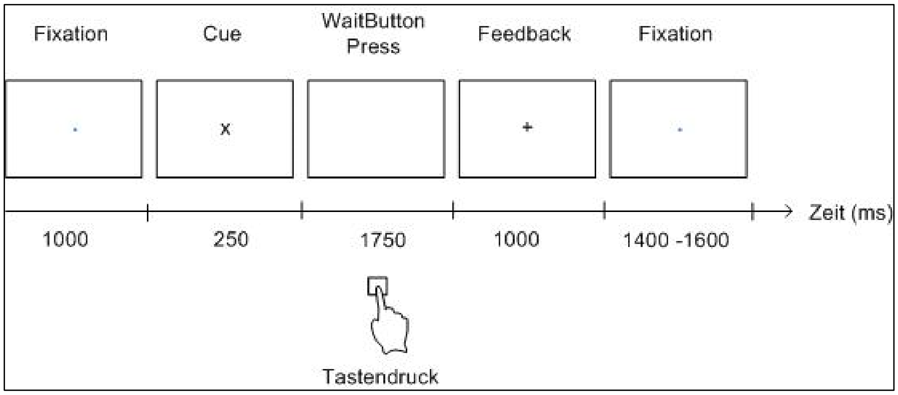

Figure 1: Schematic timeline of the used time-estimation task

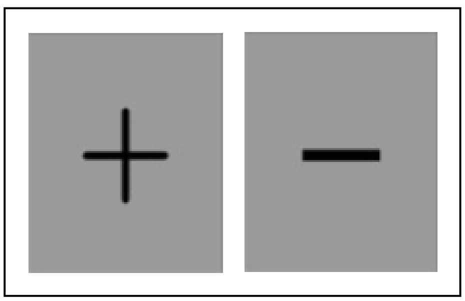

Figure 2: Used feedback stimuli in Experiment 1. The plus- and minus symbol was presented on grey background (100 x 122 pixels in size).

3.1.3 Procedure

After written informed consent was obtained and the handedness questionnaire was administered, the EEG-cap was put on, the participant was sat down in a shielded room in front of a cathode ray tube (CRT) monitor and EEG recording was started. Stimulus presentation was controlled by E-Prime 2.0 software (Psychology Software Tools, Inc., Pittsburgh, PA, USA) First, a HEOG (horizontal electrooculogram) and VEOG (vertical electrooculogram) task was administered in order to control EEG data for horizontal and vertical eye movements subsequently. The experiment started with an instruction in order to explain the task (see appendix A.1 for the whole instruction in German). The instruction was followed by 20 practice trials to become acquainted with the time-estimation task. Afterwards the participants had the possibility to ask in case anything was unclear. The experimental task consisted of 400 trials separated into four blocks of

100 trials. There was a break after each block giving the opportunity to drink and communicate with scientific staff. The length of the breaks was controlled by the participants themselves. The whole task took between 45 and 60 minutes depending on the self-chosen length of the breaks.

3.1.4 Electrophysiological Recording

EEG was recorded via 62 equidistantly positioned (according to the International 10 / 20 system; Bocker, Van Avermaete, & Van Den Berg-Lenssen, 1994) Ag/AgCl ring electrodes embedded in an elastic electrode cap (EASYCAP GmBH, Herrsching, Germany, model M10). Two reference electrodes were used, one placed at the clavicle and the other one at the vertebra prominens. Electroocular electrodes were positioned above and below the left eye and on the outer canthi of both eyes to control off-line for horizontal and vertical eye movement artifacts. The mass electrode was placed on the forehead. The two electrodes placed at the clavicle and the vertebra promines were further used to record the electrocardiogram (ECG). At each electrode position, the hair was shoved aside, the skin was scratched with a sterile needle and each electrode site was filled with a degassed conductance gel in order to reduce the impedance of each electrode to no more than 2 kΩ. Impedances were checked by an impedance meter (Ing. Kurt Zickler GmbH, model: 32-Kanal-Elektroden-Impedanzmessgerät, Pfaffstätten, Austria). The EEG signal was amplified using an AC amplifier set-up with a time constant of 10ms (Ing. Kurt Zickler GmBH, Pfaffstätten, Austria). The recorded EEG signals were within a frequency range of 0.1 – 125Hz and were stored at a sampling rate of 250Hz.

3.1.5 Data Analysis

3.1.5.1 Behavioral data

First, the percentage of received positive and negative feedback stimuli was compared subject-wise in order ensure that the experimental manipulation (ratio positive : negative feedback ≈ 50 : 50) has worked. Furthermore, the individual response times were analyzed in the following manner, based on the procedure used by Miltner et al. (1997), and Van der Veen et al. (2011):

(1) The mean response times were calculated across all subjects and trials in order to investigate if the one-second-interval was over- or underestimated in general.

(2) Mean differences between the response times and 1000ms during positive feedback trials were used to calculate the mean time window that participants had to hit for receiving positive feedback.

(3) In order to investigate if the time estimation changed more after negative than after positive feedback trials, differences between the response time on the actual trial and the preceding trial ($t_n - t_{n-1}$) were calculated. These difference scores were separated into two groups depending on whether the preceding feedback was positive or negative. Thereafter, a mean difference score for both conditions (positive vs. negative) was calculated separately for each participant and a dependent t-test was applied to test for differences between the positive and negative feedback condition. In order to make the results of Study 1 comparable to Study 2 and 3, the t-value was additionally converted into a F-value and partial η^2 was calculated.

(4) It was further coded for each trial whether the change of the response time from one trial to another was directed towards 1000ms or not. As above, a dependent t-test was used to examine whether or not the participants were able to adjust their time-estimations after negative feedback in an appropriate way (towards 1000ms). For this reason, the absolute and relative frequencies of correct and incorrect adjustments after positive and negative feedback were calculated separately for each participant.

3.1.5.2 EEG data

EEG data were analyzed offline with EEGLAB 6.0.3b (Delorme, & Makeig, 2004), an open source MATLAB 7.9.0 (The MathWorks, Inc., Natick, USA) toolbox.

Pre-Processing:

Ocular artifacts, ocular movements and eye blinks, were removed off-line using a procedure described by Bauer and Lauber (1979) and Lamm, Fischmeister and Bauer (2005). The continuous EEG signal was segmented into epochs consisting of a baseline of 200ms preceding the feedback stimulus and 1000ms following the stimulus onset.

Subsequently, the EEG date were converted to the EEGLAB format, the standardized Easy-Cap coordinates were read in according to the international 10/20 system. Additionally, the EEG signal was filtered using a 30Hz low pass filter (roll-off 6 dB per octave).

Artifact correction:

A semi-automatic approach was chosen to identify and reject trials containing artifacts. At the first step, trials were screened for extreme values and abnormal trends. If the EEG signal of a trial exceeded the boundary of ± 75μV and / or showed a linear drift, the trial was rejected. A linear drift was assumed if a linear function, determined by the EEG signal from the beginning to the end of a trial, showed a slope of more than ± 50μV. At the next step, trials were visually inspected and noisy trials, which were not identified by the automatic approach, were rejected. By applying this semi-automatic procedure for artifact correction, no more than 1/3 of the trials were rejected per subject.

In case the electrocardiogram (ECG) affected EEG signal in terms of heart spikes or the automatic ocular movement correction was unsuccessful, Independent Component Analysis (ICA; Hyvärinen, & Oja, 2000), implemented in EEGLAB 6.0.3b, was used in order to remove ECG or EOG artifacts. Artifact-afflicted independent components were identified and then subtracted from the data set via back projection.

ERP-quantification:

Grand average ERPs were calculated and plotted for each condition (positive vs. negative feedback) and subject. FRN and P300 amplitudes were assessed using the “Peak finder” plug-in implemented in EEGLAB 6.0.3b and mean amplitudes were calculated. The FRN component was analyzed at electrode sites Fz and FCz (Gehring, & Willoughby, 2002; Yu, & Zhou, 2006). Due to the observation that the FRN amplitude values did not differ appreciable between the electrode sites Fz and FCz, we refrained from including the electrode position as a separate factor in our data analysis. The P300 component was analyzed at the electrode site Pz (Hajcak et al., 2005; Yeung, & Sanfey, 2004).

The following methods of ERP-quantification were used for the FRN analysis:

(1) *Base-to-Peak:* The FRN amplitude, defined as the most negative peak within the time window of 200ms and 400ms after feedback onset, was obtained. An absent FRN peak resulted in a missing value.

(2) *Peak-to-Peak (P2-FRN):* The difference between the FRN amplitude and the preceding positive wave (P2) was calculated. The difference value was set at zero in the case ob absent FRN peaks.

(3) *Peak-to-Peak (P3-FRN):* The difference between the FRN amplitude and the following positive wave (P3) was calculated and as above, an absent FRN peak resulted in a zero-difference.

(4) *Mean Amplitudes (Mean, SD*): Mean amplitudes were calculated based on the mean FRN latency and the standard deviation (SD) across all participants and conditions. For example, if the mean latency of the identified FRN peaks was 280ms and SD was 40ms, the EEG signal between 260ms and 300ms after feedback onset was averaged.

(5) *Individual Mean Amplitudes:* An individual mean amplitude was calculated separately for each subject and condition based on the individual FRN latency. The EEG signal ± 30ms around the individual FRN latency was averaged. This method was only used when a FRN peak was recognizable.

An important remark concerns the interpretation of the values which are attained by the presented ERP quantification methods: Using method (1), (4) and (5), a low value represents a large FRN amplitude, whereas a high value using method (2) and (3) represents a large FRN amplitude. The former becomes comprehensible when it is considered that most EEG activity takes place in the positive scale range and minor deviations from the baseline are associated with a large FRN amplitude. In the other case, a greater difference is associated with a larger FRN.

The P300 amplitude was quantified with the following methods:

(1) *Base-to-Peak:* The P300 was defined as the most positive peak within the time range between 300ms and 600ms.

(2) *Peak-to-Peak (P3-N2):* The difference between the P300 amplitude and the preceding negative peak (N2) was calculated.

(3) *Mean Amplitudes (Mean, SD):* As above, the mean and standard deviation of the P300 latencies across all participants and conditions were calculated and were taken to determine the time range for the mean amplitudes.

(4) *Individual Mean Amplitudes:* An individual mean amplitude was calculated separately for each participant and condition based on the individual P300 latency. The EEG signal ± 30ms around the individual P300 latency was averaged.

In case of the P300, a high value represents a large P300 amplitude regardless of the method used.

Statistical analysis:

Dependent *t*-tests were calculated to compare trial-to-trial mean changes in time-estimation and to compare FRN and P300 amplitudes and latencies following positive vs. negative feedback trials. Effect sizes are provided in terms of Cohen's d for the matched t-test (Cohen, 2007). According to Cohen (1988) effect sizes of $d = 0.2$, $d = 0.5$ and $d = 0.8$ correspond to a small, medium and large effect. Subsequently, t-values were converted into F-values for a direct comparison of the three studies.

3.2 RESULTS

Descriptive statistics concerning the behavioral data, the FRN- and P300 amplitudes and latencies are presented in the appendix (A.2-A.5).

3.2.1 Behavioral data

The participants received negative feedback on 50.8% of all trials. They slightly overestimated the 1000ms time interval (mean response time: 1060ms, SD = 228). The average time window required for receiving positive feedback was between 918ms and 1082ms. The mean change in time estimation was greater after negative than after positive feedback ($t(14) = 7.890$, $p < .001$, $d = 2.04$) and there was a more frequent adjustment of time estimations towards 1000ms after negative feedback trials than after positive feedback trials ($t(14) = 11.639$, $p < .001$, $d = 3.01$).

3.2.2 FRN

Due to the fact that a FRN peak was only recognizable in 5 of 14 participants, we refrained from analyzing FRN data with the Base-to-Peak method and with Individual Mean Amplitudes because of an insufficient sample size. The statistical analysis using Peak-to-Peak calculations revealed a non significant result (*Peak-to-Peak (P2-FRN):* Fz: $t(13) = -0.680$, $p = .508$, FCz: $t(13) = -1.139$, $p = .275$; *Peak-to-Peak (P3-FRN):* Fz: $t(13) = 1.401$, $p = .185$, FCz: $t(13) = 1.238$, $p = .238$). Using mean amplitudes, a greater negativity in the time range of 268ms and 300ms was observed in the negative compared to the positive feedback condition (Fz: $t(13) = -4.166$, $p = .001$, $d = -1.11$; FCz: $t(13) = -3.836(13)$, $p = .002$, $d = -1.03$). It has to be remarked that the time window used for the calculation of mean amplitudes was based on the data of five participants showing an FRN peak. Figure 3 presents the plotted grand average ERPs at electrode site Fz.

Means and standard deviations of the FRN latencies for the negative and positive feedback conditions are depicted in the appendix (A.5). We refrained from applying a statistical test in order to analyze latency differences between the two conditions due to a greatly reduced sample size.

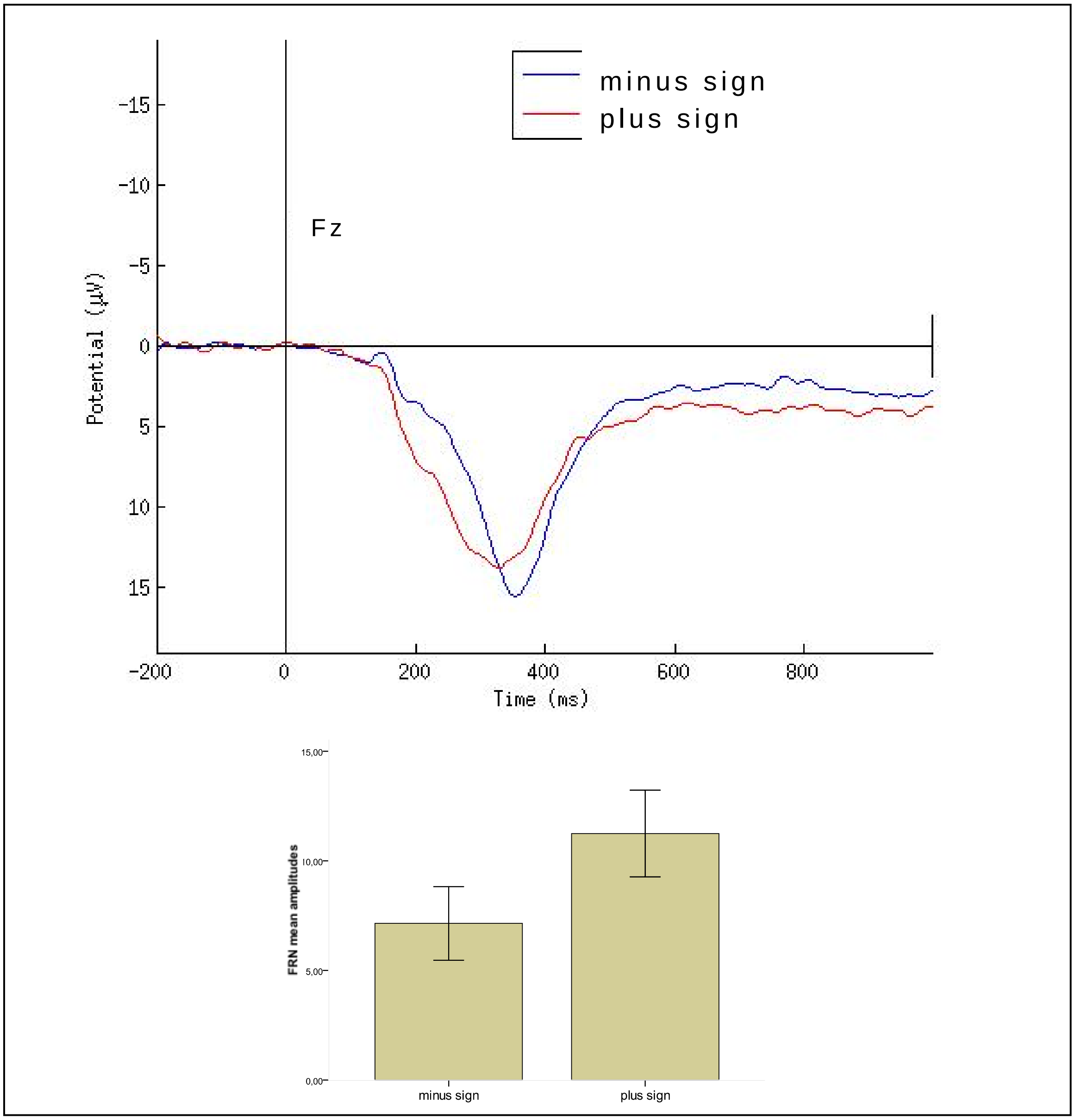

Figure 3: Grand average ERPs of Experiment 1 at electrode site Fz for negative feedback (blue line) and positive feedback (red line). Negative is drawn upward by convention. Feedback presentation started at 0ms. The bar chart depicts the respective FRN mean amplitude values. Note that a low value indicates a large FRN amplitude. Error bars indicate 1 SE.

3.2.3 P300

No significant difference between the negative and positive feedback condition concerning the P300 amplitude was found no matter which method was used: *Base-to-Peak:* t(13) = - 0.966, p = .352, *Peak-to-Peak*: t(13) = 0.354, p = .729, *Mean Amplitudes (336-364ms):* t(13) = - 1.989, p = .068, *Individual Mean Amplitudes:* t(13) = - 1.088, p =

.296 . Figure 4 presents the grand average ERPs at electrode Pz. A significant effect was observed for the P300 latency. The latency was longer in the negative feedback condition (M = 367ms, SD = 38ms) than in the positive feedback condition (M = 335ms, SD = 14ms): $t(13) = 2.860$, $p = .013$, $d = 0.77$.

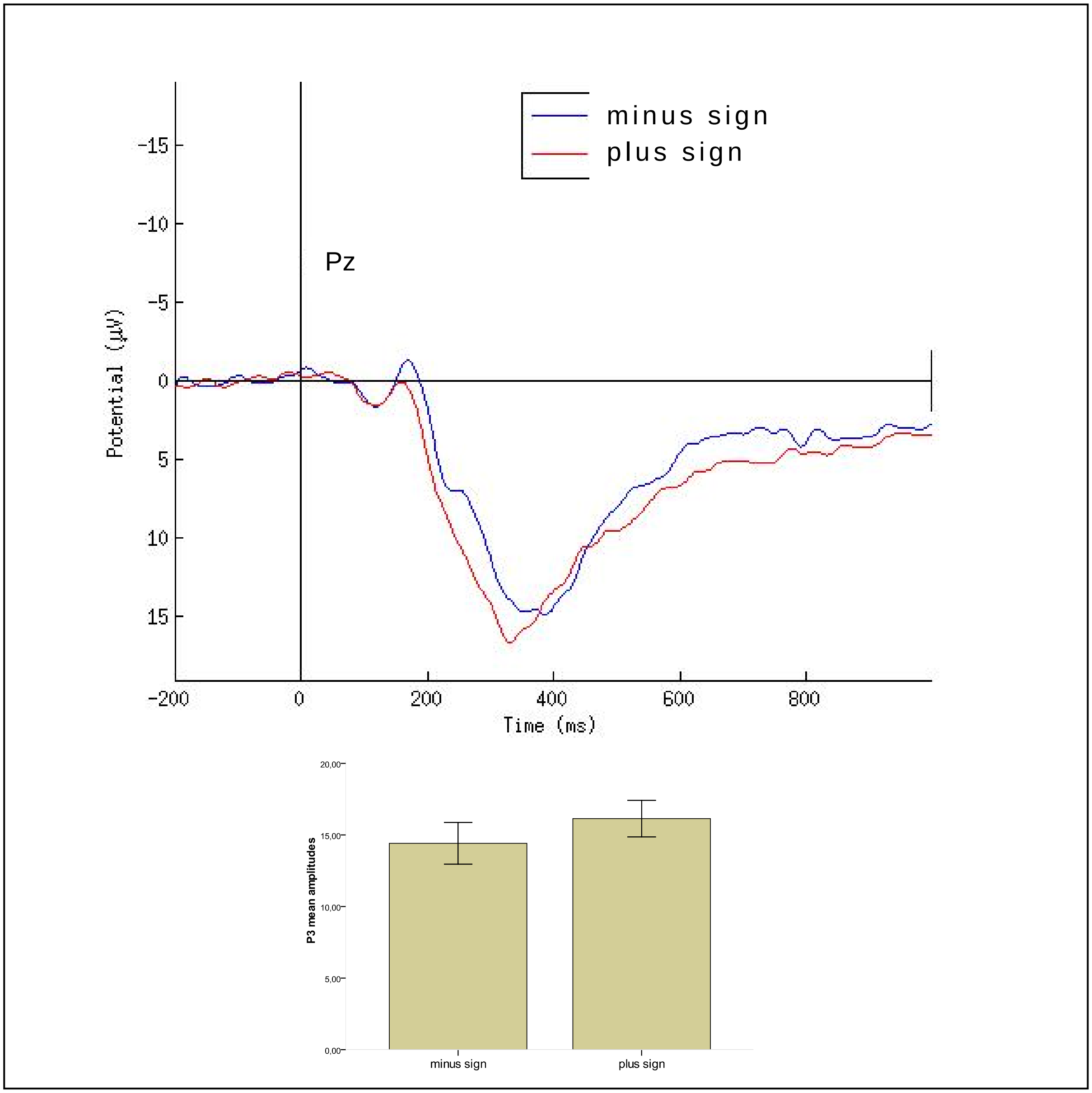

Figure 4: Grand average ERPs of Experiment 1 at electrode site Pz for negative feedback (blue line) and positive feedback (red line). Negative is drawn upward by convention. Feedback presentation started at 0ms. The bar chart depicts the respective P300 mean amplitude values and error bars indicate 1 SE.

3.2 DISCUSSION

The aim of Study 1 was to investigate the impact of two simple non-complex feedback stimuli (plus-sign, minus-sign) on feedback-related brain potentials, the FRN and P300. This study should be a replication of the experiment of Miltner et al. (1997) who first demonstrated that a negative deflection, the FRN, occurred when a negative feedback stimulus is presented. Surprisingly, the FRN peak was absent for the majority of subjects in our study and no distinct FRN peak could be identified in the grand average ERPs as well. A significant difference between the negative and positive feedback conditions was only found using mean amplitudes but not using other methods of FRN quantification. The P300 was also not sensitive for the feedback valence as well.

Therefore, we failed to replicate the findings of Miltner et al. (1997). This was unexpected because an FRN was consistently observed in previous studies using non-complex symbols as feedback stimuli (e.g. Hajcak et al., 2007; Holroyd et al., 2006). It is conceivable that negative feedback was not reflected in an obvious FRN peak, but just in a greater negativity in the FRN time range which was detected by the mean amplitude analysis. However, the present results clearly demonstrate that the choice of the FRN quantification method has considerable impact on the results we receive and the conclusions we draw. Using mean amplitudes, we come to the conclusion that there is a larger FRN amplitude after negative feedback although no negative deflection can be identified. Using other methods, we have to conclude that there is no FRN difference between positive and negative feedback at all. Using the Peak-to-Peak (P2-FRN) analysis there is even a tendency for a larger FRN in the positive feedback conditions. The heterogeneity of our results demonstrates that there should be a synopsis of different methods, individual and grand average ERPs before a conclusion is drawn.

We can only speculate about the reasons for the absence of a classical FRN in the present study. As the FRN was sensitive for involvement in the task (Yeung et al., 2005), there might have been a lack of motivation in the participants of this study. Further on, problems during EEG recording cannot be ruled out. The unusual absence of early sensorial responses in the grand average ERPs gives another hint at possible problems during EEG recording. Problems at the task level are unlikely because the behavioral data showed that the participants responded appropriately to negative and positive feedback with a greater change in time estimation and a better behavioral adjustment in the first case.

No difference between positive and negative feedback was found regarding the P300 amplitude. This finding is consistent with previous studies demonstrating that the P300 is insensitive for the feedback valence (Pfabigan et al., 2011a; Yeung, & Sanfey, 2004).

4. STUDY 2

4.1 METHODS

4.1.1 Participants

Twenty right-handed participants (9 male, 11 female) were recruited by students in the context of an EEG course at the Faculty of Psychology (University of Vienna). The participants had a mean age of 25.80 years (SD = 3.91) and no one reported any history of neurological, psychiatric or severe physical illness. No monetary remuneration was provided for their participation. Written informed consent was obtained from all participants and handedness was assessed by the German version of the Edinburgh Handedness Inventory (Oldfield, 1971). The study was conducted in accordance with the Declaration of Helsinki and local ethical guidelines of the University of Vienna. Two subjects had to be excluded from further EEG analysis because of abnormal noise in their EEG signal.

4.1.2 Task and Procedure

The same time-estimation task and procedure as in Study 1 were used. The only change to Study 1 concerns the given feedback stimuli. Two types of positive and negative feedback stimuli were presented in Study 2 (symbols and faces). The angry face and the happy face (Figure 5) were taken from the Ekman and Friesen's Facial Action Coding System (Ekman, & Friesen, 1978). The plus-sign and the happy face served as positive feedback, the minus-sign and the angry face as negative feedback. As described in detail for Study 1, the valence of given feedback (positive vs. negative) was a function of whether the time-estimation fell in a defined time-window or not. In the current study, the feedback type (symbols and faces) changed randomly from trial to trial. Consequently, all trials were approximately equally separated into four conditions:

25% positive facial feedback (happy face), 25% positive symbol feedback (plus-sign), 25% negative facial feedback (angry face) and 25% negative symbol feedback (minus-sign).

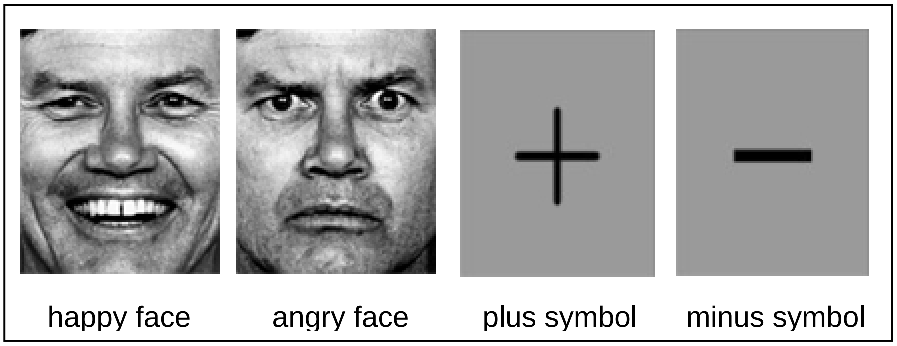

Figure 5: Used feedback stimuli in Experiment 2 and 3 (100 x 122 pixels in size). In Experiment 2, the feedback type (faces vs. symbols) changed randomly from trial to trial. In Experiment 3, the feedback type remained stable within blocks of 50 trials.

4.1.3 Electrophysiological recording

The same EEG recording procedure was used as in Study 1 with the exception of the number of the applied electrodes. In Study 2, only the central scalp electrode sites Fz, FCz, Cz, PCz, Pz and Oz according to the international 10/20 system (Bocker, Van Avermaete, & Van Den Berg-Lenssen, 1994) were used because of economical reasons and the fact that the FRN and P300 amplitudes commonly reach their maximum at the central line of the scull.

4.1.4 Data analysis and statistical analysis

The same procedure as in Study 1 was used for EEG preprocessing and analysis. EEG data were treated separately for the four conditions (happy face, angry face, plus-sign, minus-sign). For the statistical analysis a 2x2 ANOVA was applied to analyze the influence of the within subject factors *feedback valence* and *feedback type* on the behavioral data, FRN and P300 amplitudes and latencies. Concerning the FRN, this 2x2 ANOVA was applied separately for the Fz and FCz amplitude values. Effect sizes are provided in terms of partial eta² (η_p^2). According to Kirk (1996) values of $\eta_p^2 = .01$, $\eta_p^2 = .06$ and $\eta_p^2 = .14$ represent a small, medium and large effect. Pearson correlation coefficients were calculated to analyze the association between the individual ERPs and the individual mean trial-to-trial time estimation change.

4.2 RESULTS

Descriptive statistics of the behavioral data, the FRN- and P300 amplitudes and latencies are presented in the appendix (A.2-A.5).

4.2.1 Behavioral data

Due to technical problems, behavioral data were only available for 16 of 20 participants. The four feedback stimuli were approximately evenly distributed on the trials (26.0% happy face, 25.2% minus-sign, 24.0% angry face, 24.8% plus-sign). The participants received negative feedback on 51.2% of the trials. The average time window required for receiving positive feedback lay between 800 and 1200ms and the participants slightly underestimated the one-second-interval (967ms ± 279ms). There was a greater change in time estimation after negative feedback ($F_{[1,15]}$ = 20.020, p < .001, η_p^2 = 0.572) No significant effect of the feedback type ($F_{[1,15]}$ = 0.820, p = .379, η_p^2 = 0.052) and no interaction effect was observed ($F_{[1,15]}$ = 0.894, p = .359, η_p^2 = 0.056). The subjects correctly adjusted (towards 1000ms) their time estimation in 76.7% after negative feedback trials, compared to 51.3% in the positive feedback trials. A correct adjustment was significantly more frequent after negative than after positive feedback ($F_{[1,15]}$ = 70.387, p < .001, η_p^2 = 0.824).

4.2.2 FRN

The FRN was absent for four participants in the negative feedback conditions. No FRN amplitude was found for eight participants in the happy-face condition and for five participants in the plus-sign condition. The grand average ERPs are depicted in Figure 6. The main effect of ***feedback valence*** was significant across all FRN quantification methods and both electrodes (Fz and FCz). The FRN amplitude was larger in the negative compared to the positive feedback conditions. *Base-to-Peak:* Fz ($F_{[1,9]}$ = 17.902, p = .002, η_p^2 = 0.665), FCz ($F_{[1,8]}$ = 15.508, p = .004, η_p^2 = 0.660), *Peak-to-Peak (P2-FRN):* Fz ($F_{[1,17]}$ = 22.245, p < .001, η_p^2 = 0.567), FCz ($F_{[1,17]}$ = 20.282, p < .001, η_p^2 = 0.544), *Peak-to-Peak (P3-FRN):* Fz ($F_{[1,17]}$ = 16.230, p < .001, η_p^2 = 0.488), FCz ($F_{[1,17]}$ = 15.250, p < .001, η_p^2 = 0.473), *Mean Amplitudes (280-312ms):* Fz ($F_{[1,17]}$ = 30.440, p < .001, η_p^2 = 0.642), FCz ($F_{[1,17]}$ = 29.965, p < .001, η_p^2 = 0.613), *Individual Mean*

Amplitudes: Fz ($F_{[1,9]} = 13.824$, $p = .005$, $\eta_p^2 = 0.606$), FCz ($F_{[1,17]} = 15.885$, $p = .003$, $\eta_p^2 = 0.638$). The main effect of ***feedback type*** was only significant using the following three methods, the *Base-to-Peak* method (Fz: $F_{[1,9]} = 10.645$, $p = .010$, $\eta_p^2 = 0.542$; FCz: $F_{[1,8]} = 10.143$, $p = .013$, $\eta_p^2 = 0.559$), the *Mean-Amplitude(280-312 ms)* method (Fz. $F_{[1,17]} = 12.053$, $p = .003$, $\eta_p^2 = 0.415$; FCz: $F_{[1,17]} = 11.377$, $p = .004$, $\eta_p^2 = 0.401$) and the *Individual Mean Amplitude* method (Fz: $F_{[1,9]} = 7.895$, $p = .020$, $\eta_p^2 = 0.467$; FCz: $F_{[1,9]} = 9.754$, $p = .012$, $\eta_p^2 = 0.520$). The means (see appendix A.4) indicate that the FRN amplitude was larger following symbols than following facial feedback. No significant feedback type effects were observed using the both peak-to-peak measures: *Peak-to-Peak (P2-FRN):* Fz: $F_{[1,17]} = 0.001$, $p = .972$, $\eta_p^2 = 0.000$; FCz: $F_{[1,17]} = 0.390$, $p = .541$, $\eta_p^2 = 0.022$; *Peak-to-Peak (P3-FRN):* Fz: $F_{[1,17]} = 2.799$, $p = .113$, $\eta_p^2 = 0.141$; FCz: $F_{[1,17]} = 2.426$, $p = .138$, $\eta_p^2 = 0.125$. The ***interaction effect (valence x type)*** was only observed for the *Mean Amplitude(280-312 ms)* method (Fz: $F_{[1,17]} = 4.403$, $p = .051$, $\eta_p^2 = 0.206$; FCz: $F_{[1,17]} = 12.628$, $p = .002$, $\eta_p^2 = 0.426$) and was non-significant using all other measurements (all $p_s > .306$). No significant **FRN latency** effects were observed. Neither the main effect of feedback valence (Fz: $F_{[1,9]} = 2.492$, $p = .149$, $\eta_p^2 = 0.217$; FCz: $F_{[1,9]} = 0.689$, $p = .428$, $\eta_p^2 = 0.071$), the main effect of feedback type (Fz: $F_{[1,9]} = 2.840$, $p = .126$, $\eta_p^2 = 0.240$; FCz: $F_{[1,9]} = 3.912$, $p = .079$, $\eta_p^2 = 0.303$), nor the interaction valence x type (Fz: $F_{[1,9]} = 3.080$, $p = .113$, $\eta_p^2 = 0.255$; FCz: $F_{[1,9]} = 0.354$, $p = .567$, $\eta_p^2 = 0.038$) reached statistical significance.

4.2.3 P300

The Grand Average ERPs at electrode site Pz are depicted in Figure 7. Regarding the P300 amplitude, statistical analysis revealed a significant main effect of ***feedback valence*** for each used method: *Base-to-Peak*: $F_{[1,17]} = 23.285$, $p < .001$, $\eta_p^2 = 0.607$, *Peak-to-Peak (P3-N2):* $F_{[1,17]} = 28.107$, $p < .001$, $\eta_p^2 = 0.623$, *Mean Amplitudes (352-404ms):* $F_{[1,17]} = 28.074$, $p < .001$, $\eta_p^2 = 0.623$, *Individual Mean Amplitudes:* $F_{[1,17]} = 25.968$, $p < .001$, $\eta_p^2 = 0.604$. In each case, the P300 amplitude was larger in the positive feedback conditions. We also found a significant main effect of ***feedback type*** which was constantly observed across all used methods with a larger P300 amplitude in the facial feedback conditions compared to symbols: *Base-to-Peak*: $F_{[1,17]} = 19.804$, p <

.001, η_p^2 = 0.538, *Peak-to-Peak (P3-N2):* $F_{[1,17]}$ = 9.249, p = .007, η_p^2 = 0.352, *Mean Amplitudes (352-404ms):* $F_{[1,17]}$ = 11.684, p = .003, η_p^2 = 0.455, *Individual Mean Amplitudes:* $F_{[1,17]}$ = 18.140, p < .001, η_p^2 = 0.516. A significant **interaction *valence x type*** was found using the *Mean Amplitude(352-404ms)* method ($F_{[1,17]}$ = 9.506, p = .007, η_p^2 = 0.359), *Individual Mean Amplitudes* ($F_{[1,17]}$ = 5.359, p = .033, η_p^2 = 0.240) and the *Base-to-Peak* method ($F_{[1,17]}$ = 4.891, p = .041, η_p^2 = 0.223) but not for the *Peak-to-Peak(P3-N2)* method ($F_{[1,17]}$ = 1.424, p = .249, η_p^2 = 0.077). The interaction diagram showing the amplitude courses in Figure 6 reveals a special large P300 amplitude in the happy face condition. Pairwise post-hoc tests comparing the P300 amplitude value (based on mean amplitudes) between the happy face condition and the other conditions all reached statistical significance even when the alpha level is adjusted for multiple comparison by a Bonferroni correction (α_{adj} = .017): happy face vs. angry face: t(17) = 4.935, p < .001; happy face vs. minus-sign: t(17) = 5.128, p < .001; happy face vs. plus-sign: t(17) = 2.894, p = .010. Those post-hoc tests demonstrate that the P300 amplitude is significantly larger in the happy face condition compared to the other conditions. Regarding the **P300 latencies**, no main effects for neither *feedback valence* ($F_{[1,17]}$ = 0.000, p = .992, η_p^2 = 0.000) nor *feedback type* ($F_{[1,17]}$ = 0.170, p = .685, η_p^2 = 0.010) was found, but a significant interaction occurred revealing a longer latency in the angry face and plus-sign condition, a shorter latency in the happy face and minus-sign condition respectively ($F_{[1,17]}$ = 8.279, p = .010, η_p^2 = 0.328). Pairwise comparisons revealed that there were no significant latency differences between the four conditions (angry face vs. minus-sign: t(17) = 1.501, p = .152; angry face vs. happy face: t(17) = 1.406, p = .178; angry face vs. plus-sign: t(17) = 0.263, p = .795; minus-sign vs. happy face: t(17) = -0.359, p = .724; minus-sign vs. plus-sign: t(17) = -1.763, p = .096; happy face vs. plus-sign: t(17) = -1.663, p = .115).

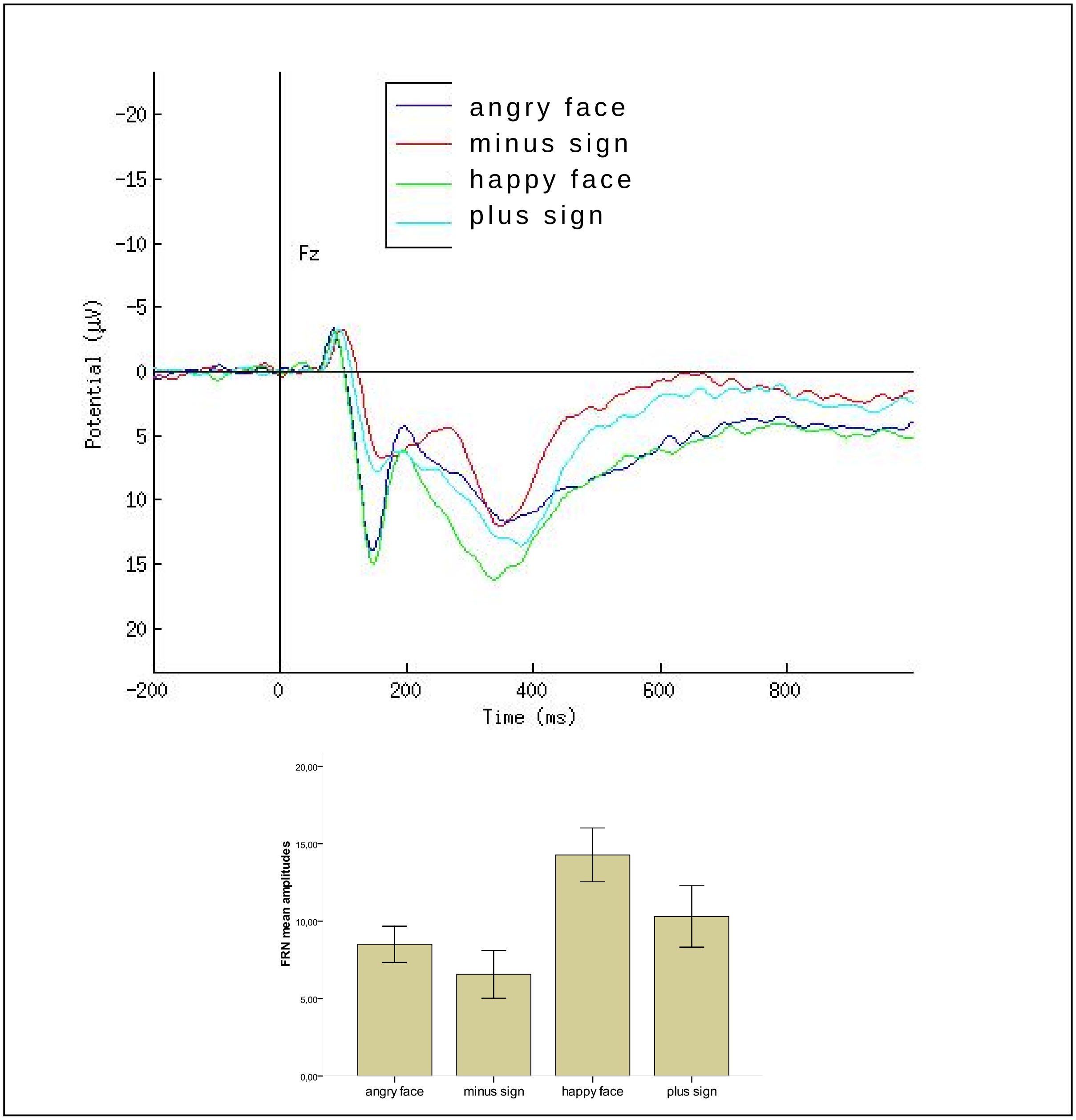

Figure 6: Grand average ERPs of Experiment 2 at electrode site Fz for all feedback conditions. Negative is drawn upward by convention. Feedback presentation started at 0ms. The bar chart depicts the respective FRN mean amplitude values. Note that a low value indicates a large FRN amplitude. Error bars indicate 1 SE.

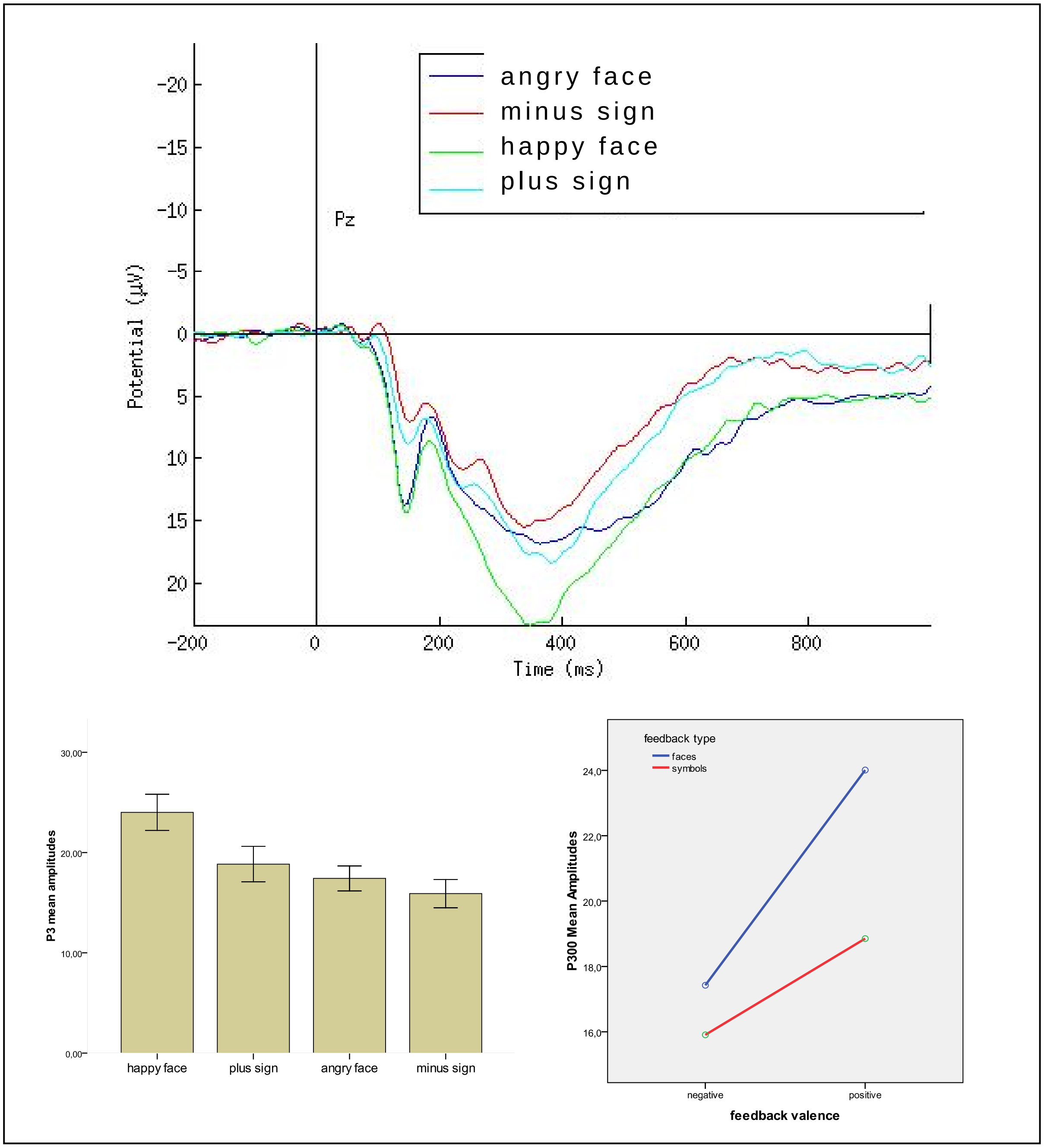

Figure 7: Grand average ERPs of Experiment 2 at electrode site Pz for all feedback conditions. Negative is drawn upward by convention. Feedback presentation started at 0ms. The bar chart depicts the respective P300 mean amplitude values and error bars indicate 1 SE. The interaction graph depicts the P300 mean amplitude values.

4.2.3 Correlations of ERPs and behavioral data

FRN and P300 mean amplitude values were chosen for the correlation analysis because the whole sample is included using this method, in contrast to the base-to-peak method where the sample size was reduced due to absent peaks. No significant correlations were found when the individual FRN (obtained by mean amplitudes at electrode Fz) across both negative conditions was related to the individual mean time estimation change after negative feedback trials ($r = -.365$, $n = 15$, $p = .181$) and when the FRN amplitude across both positive feedback conditions was related to the mean time estimation change after positive feedback trials ($r = -.136$, $n = 15$, $p = .628$). The overall FRN amplitude across all four conditions was not significantly associated with the overall trial-to-trial time estimation change as well ($r = -.263$, $n = 15$, $p = .344$). Note that a low value represents a large FRN amplitude, thus a negative correlation means that a large FRN amplitude is accompanied by a great time estimation change.

The individual P300 amplitude (obtained by mean amplitudes at electrode Pz) across both negative conditions was not significantly associated with the trial-to-trial time estimation change after negative feedback trials ($r = -.439$, $n = 15$, $p = .102$). In an analogous manner, the P300 amplitude across both positive conditions was not significantly correlated with the time estimation change after positive feedback trials ($r = -.164$, $n = 15$, $p = 559$). The overall P300 amplitude across all conditions was not correlated with the overall time estimation change as well ($r = -.321$, $n = 15$, $p = .243$).

Although there are no significant effects, there is a tendency of a larger FRN amplitude and a lower P300 amplitude related to a greater change in time estimation from trial to trial.

4.3 DISCUSSION

In Study 2 the FRN and P300 responses to two feedback types (symbols and faces) were compared. The aim of this study was to investigate whether both, the FRN and P300, are sensitive to the feedback valence and feedback type or if they reflect just either feedback valence or type. In this study, the feedback type changed randomly from trial to trial. Statistical analysis revealed that the FRN was sensitive for the feedback valence and type with a larger FRN amplitude after negative feedback and symbols. The interpretation of the FRN effects are based on the results of all used methods except for the peak-to-peak measures which leaded to a non-significant result concerning the feedback type dimension. The P300 was sensitive for both outcome features as well, but in the opposite way. The P300 amplitude was larger for positive feedback and faces.

Consistent with previous studies, the FRN was larger in the negative feedback conditions. It was assumed that an emotional feedback stimulus should enhance the error-related ERP response because of preliminary findings demonstrating that an induced negative affect by emotional pictures (Wiswede et al., 2009a), derogatory verbal feedback (Wiswede et al., 2009b), or a negative affect trait (Sato et al., 2005) was related to a larger FRN amplitude. As opposed by this, we found a larger FRN amplitude after feedback symbols than emotional facial expressions. Regarding the grand average ERPs, a clear FRN peak could only be recognized in the condition with the minus-sign. One possible explanation for this finding in the present study is that we tried to induce affect by the feedback stimulus itself and not by pictures or verbal statements before the presentation of the real feedback stimulus. Our results indicate that these two ways of experimental affect manipulation make a difference regarding the FRN amplitude. Furthermore, the larger FRN amplitude after symbols could have something to do with the stimulus complexity which is definitely lower for feedback symbols than pictures of facial expressions. This assumption has to be investigated in further studies by varying systematically the level of feedback stimuli complexity.

The finding of larger P300 amplitudes after positive and facial feedback is consistent with previous studies which reported that the P300 was sensitive for feedback salience and valence (Bellebaum et al., 2010; Wu, & Zhou, 2009). Assuming that facial feedback stimuli induce a short-term negative affect, we conclude from our findings that a positive affect is reflected by the P300 amplitudes. Thus we observed a dissociation between FRN and P300 in respect of the feedback salience dimension. Whereas affective stimuli did not come along with larger FRN amplitudes, they had a significant influence on the P300 amplitude. The interaction effect between feedback valence and type indicates that the P300 main effect of feedback type goes beyond a simple effect of stimulus complexity. The interaction reached statistical significance only when using mean amplitudes, but can be easily identified by looking at the grand average ERPs at electrode Pz.

Considering the FRN and P300 effects found, the double dissociation theory of Yeung and Sanfey (2004) which suggests that the FRN should reflect the feedback valence but not the feedback salience and the P300 should reflect the feedback salience but not the feedback valence, cannot be supported by the present study. Our results indicate that the FRN and P300 effects are a function of multiple outcome features which was already argued by Goyer et al. (2008).

No obvious difference between the P300 quantification methods was observed. Both main effects reached statistical significance no matter which method was used. Concerning the FRN analysis the Peak-to-Peak measurements didn't reached statistical significance when the main effect of feedback type was analyzed. It can be argued that this was found because the P2 and P3 peak to which the FRN is compared, vary across the conditions. We performed additional statistical analysis (repeated measurement ANOVAs) to test for P2 and P3 amplitude differences at electrode site Fz and FCz. Indeed, the P2 amplitude (*Main effect of feedback valence:* Fz: $F_{[1, 17]} = 13.554$, $p = .002$, FCz: $F_{[1, 17]} = 13.628$, $p = .002$; *Main effect of feedback type:* Fz: $F_{[1, 17]} = 28.320$, $p < .001$, FCz: $F_{[1, 17]} = 29.820$, $p < .001$) as well as the P3 amplitude (*Main effect of feedback valence:* Fz: $F_{[1, 17]} = 11.202$, $p = .004$, FCz: $F_{[1, 17]} = 11.977$, $p = .003$; *Main effect of feedback type:* Fz: $F_{[1, 17]} = 2.428$, $p = .138$, FCz: $F_{[1, 17]} = 3.236$, $p = .090$) varied across the conditions.

5. STUDY 3

5.1 METHODS

5.1.1 Participants

The sample consisted of 17 healthy right-handed participants (six men, eleven women) with a mean age of 26.00 (SD = 6.43) years. The participants were free of any cardiovascular and neurological disease. Written informed consent was obtained from each participant prior to the experiment. They were recruited by undergraduate students in the circle of their acquaintances. The study was conducted in accordance with the Declaration of Helsinki and local ethical guidelines of the University of Vienna. Due to unusual noise in the EEG signal, one participant had to be excluded from further EEG analysis. The behavioral data are analyzed on the bases of the whole sample.

5.1.2 Task and Procedure

The same time-estimation task (Miltner et al., 1997) as in Study 2 and 3 was administered. As in Study 2 four different feedback stimuli (happy and angry face, plus and minus sign) were presented. In contrast to Study 2, the feedback types (faces and symbols) did not change from trial to trial but were presented in blocks. Each block consisted of 50 trials containing either faces or symbols. There were eight blocks in the whole experiment and the feedback type alternated from block to block. It was randomized whether the first block contained facial feedback or symbols. Consequently, facial feedback was presented in four blocks and feedback in terms of symbols was presented in four blocks as well. In exception of the feedback presentation mode, the experimental procedure was exactly the same as in Study 2.

5.1.3 Electrophysiological recording

EEG recording was identical to Study 2. EEG data were recorded from six central scalp electrodes (Fz, FCz, Cz, CPz, Pz, Oz).

5.1.4 Data analysis and statistical analysis

The same procedure for EEG preprocessing, artifact correction, ERP quantification and statistical analysis as in Study 2 was used.

5.2 RESULTS

Descriptive statistics concerning the behavioral data, the FRN- and P300 amplitudes and latencies are depicted in the appendix (A.2-A.5).

5.2.1 Behavioral data

The participants received negative feedback in 50.7% of the trials. Negative and positive feedback was approximately equally-distributed on the facial feedback and symbol feedback conditions (25.3% angry face, 25.4% minus-sign, 24.7% happy face, 24.6% plus-sign). The participants slightly overestimated the one-second interval (M = 1021.09, SD = 266.15). The average time window required for receiving positive feedback was between 905ms and 1095ms. The mean trial-to-trial change in time estimation was significantly greater in trials after negative feedback than after positive feedback ($F_{[1,16]}$ = 62.939, p < .001, η_p^2 = 0.797), but there were no impact of the feedback type ($F_{[1,16]}$ = 0.446, p = .514, η_p^2 = 0.027) and no valence x type interaction effects ($F_{[1,16]}$ = 0.006, p = .939, η_p^2 < 0.001. The participants correctly adjusted their time-estimation more often towards 1000ms after negative feedback than after positive feedback trials (77.40 % vs. 53.50%; $F_{[1,16]}$ = 100.181, p < .001, η_p^2 = 0.862).

5.2.2 FRN

Figure 8 presents the grand average ERPs at electrode site Fz. No FRN peak could be assessed for one participant in the angry-face condition, for two participants in the minus-sign condition and for approximately half of the participants in the positive feedback conditions. Due to missing FRN peaks, the sample size is greatly reduced when the *Base-to-Peak* method is used and *Individual Mean Amplitudes* are calculated. Thus, those methods have limited validity regarding the present data. With the exception of the Base-to-Peak method at electrode FCz, statistical analysis revealed a significant main effect of ***feedback valence*** for each method used towards a larger FRN amplitude in the negative feedback conditions. *Base-to-Peak:* Fz ($F_{[1,6]}$ = 35.561, p < .001, η_p^2 = 0.856), FCz ($F_{[1,3]}$ = 9.942, p =.051, η_p^2 = 0.768), *Peak-to-Peak (P2-FRN):* Fz ($F_{[1,15]}$ = 29.330, p < .001, η_p^2 = 0.662), FCz ($F_{[1,15]}$ = 10.327, p = .006, η_p^2 = 0.408), *Peak-to-Peak (P3-FRN):* Fz ($F_{[1,15]}$ = 25.874, p < .001, η_p^2 = 0.633), FCz ($F_{[1,15]}$ = 24.557, p < .001, η_p^2 = 0.621), *Mean Amplitudes (272-312ms):* Fz ($F_{[1,15]}$ = 44.934, p < .001, η_p^2 = 0.750), FCz ($F_{[1,15]}$ = 51.848, p < .001, η_p^2 = 0.776); *Individual Mean Amplitudes:* Fz ($F_{[1,6]}$ = 32.999, p < .001, η_p^2 = 0.846), FCz ($F_{[1,3]}$ = 12.006, p =.040, η_p^2 = 0.800). Regarding the within factor ***feedback type*** no significant difference between faces and symbols was observed except for the Base-to-Peak and the Individual Mean Amplitude methods at electrode FCz ($F_{[1,3]}$ = 14.530, p = .032, η_p^2 = 0.829, respectively $F_{[1,3]}$ = 14.087, p = .031, η_p^2 = 0.832) with a tendency of a larger FRN after symbols compared to faces, but is has to be noted that the sample size is reduced to four subjects in these cases. Using all other methods, no significant differences between faces and symbols were observed: *Base-to-Peak:* Fz ($F_{[1,6]}$ = 0.020, p = .893, η_p^2 = 0.036), *Peak-to-*

Peak(P2-FRN): Fz ($F_{[1,15]}$ = 0.006, p = .939, η_p^2 < 0.000), FCz ($F_{[1,15]}$ = 0.304, p = .590, η_p^2 = 0.020), *Peak-to-Peak(P3-FRN):* Fz ($F_{[1,15]}$ = 0.062, p = .807, η_p^2 = 0.004), FCz ($F_{[1,15]}$ = 1.026, p = .327, η_p^2 = 0.064), *Mean Amplitudes (272-312ms):* Fz ($F_{[1,15]}$ = 0.038, p = .848, η_p^2 = 0.003), FCz ($F_{[1,15]}$ = 0.028, p = .601, η_p^2 = 0.019), *Individual Mean Amplitudes:* Fz ($F_{[1,6]}$ = 0.022, p = .888, η_p^2 = 0.004). No ***valence x type*** **interaction** effect was observed expect for the Peak-to-Peak (P3-FRN) method at electrode site Fz: *Base-to-Peak:* Fz ($F_{[1,6]}$ = 0.227, p = .650, η_p^2 = 0.036), FCz ($F_{[1,3]}$ = 0.184, p = .697, η_p^2 = 0.058), *Peak-to-Peak(P2-FRN):* Fz ($F_{[1,15]}$ = 2.341, p = .147, η_p^2 = 0.135), FCz ($F_{[1,15]}$ = 1.105, p = .310, η_p^2 = 0.069), *Peak-to-Peak(P3-FRN):* Fz ($F_{[1,15]}$ = 5.033, p = .040, η_p^2 = 0.251), FCz ($F_{[1,15]}$ = 1.695, p = .213, η_p^2 = 0.102), *Mean Amplitudes(272-312ms):* Fz ($F_{[1,15]}$ = 0.720, p = .409, η_p^2 = 0.046), FCz ($F_{[1,15]}$ = 0.121, p = .733, η_p^2 = 0.008), *Individual Mean Amplitudes*: Fz ($F_{[1,6]}$ = 0.100, p = .762, η_p^2 = 0.016), FCz ($F_{[1,3]}$ = 0.161, p = .715, η_p^2 = 0.051. No significant **FRN latency** differences between the conditions were observed, neither for the main effect of feedback valence (Fz: $F_{[1,6]}$ = 2.227, p = .186, η_p^2 = 0.271, FCz: $F_{[1,3]}$ = 0.024, p = .888, η_p^2 = 0.008), nor for the main effect of feedback type (Fz: $F_{[1,6]}$ = 0.114, p = .747, η_p^2 = 0.019, FCz: $F_{[1,3]}$ = 0.050, p = .838, η_p^2 = 0.016). There were further no valence x type interaction effects: Fz: $F_{[1,6]}$ = 3.687, p = .107, η_p^2 = 0.374, FCz: $F_{[1,3]}$ = 0.680, p = .470, η_p^2 = 0.185.

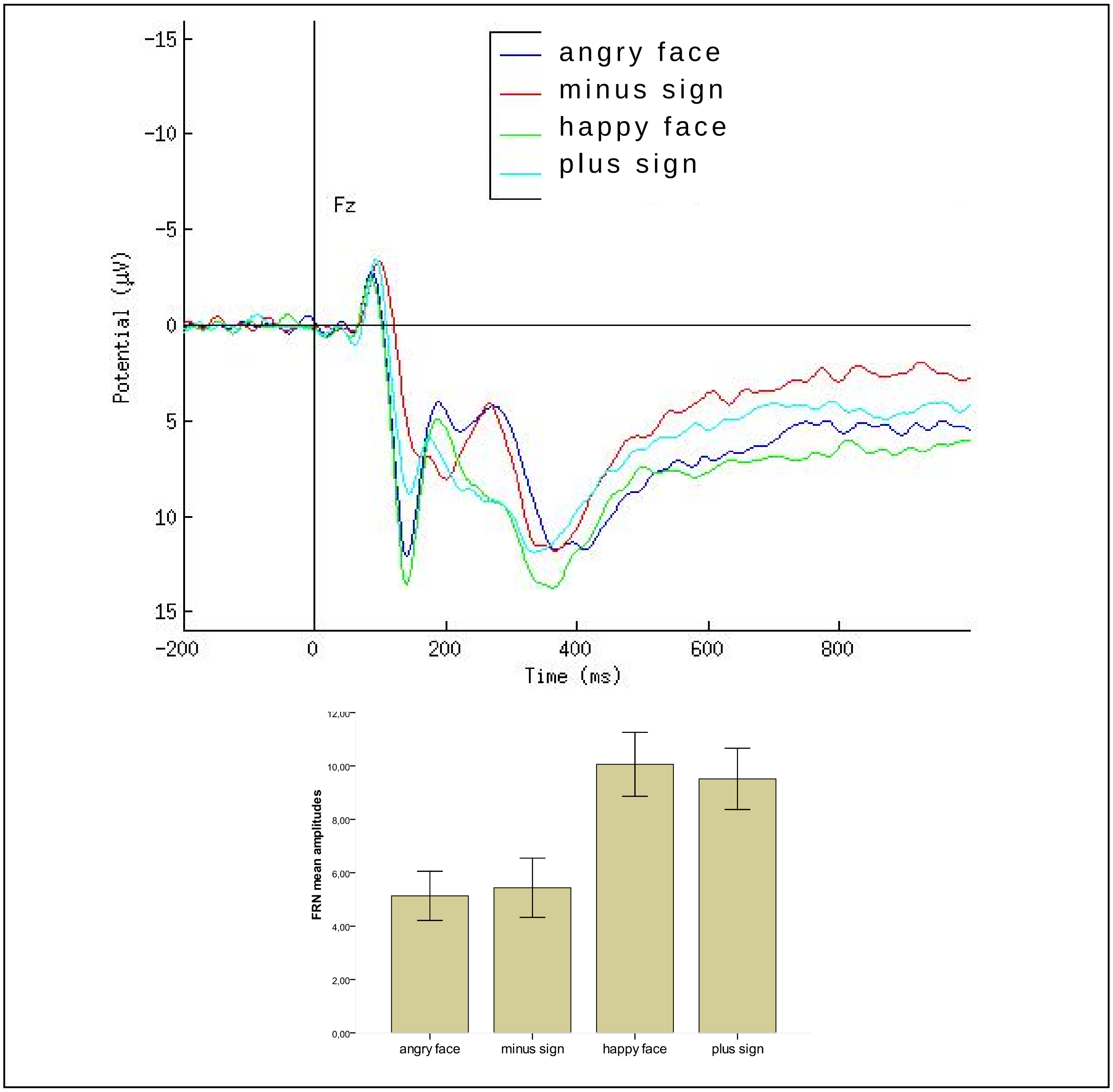

Figure 8: Grand average ERPs of Experiment 3 at electrode sites Fz for all feedback conditions. Negative is drawn upward by convention. Feedback presentation started at 0ms. The bar chart depicts the respective FRN mean amplitude values. Note that a low value indicates a large FRN amplitude. Error bars indicate 1 SE.

5.2.3 P300

The grand average ERPs at electrode Pz are depicted in figure 9. The main effect of ***feedback valence*** was significant with the exception of the peak-to-peak method towards a larger P300 amplitude in the positive feedback conditions: *Base-to-Peak:* $F_{[1,15]} = 12.493$, $p = .003$, $\eta_p^2 = 0.454$; *Peak-to-Peak(P3-N2):* $F_{[1,15]} = 2.601$, $p = .128$, $\eta_p^2 = 0.148$, *Mean Amplitudes (360-412ms):* $F_{[1,15]} = 5.617$, $p = .032$, $\eta_p^2 = 0.272$; *Individual Mean Amplitudes:* $F_{[1,15]} = 13.468$, $p = .002$, $\eta_p^2 = 0.473$. Regarding the main effect of ***feedback type*** a larger P300 amplitude after faces were observed across all used methods. *Base-to-Peak:* $F_{[1,15]} = 5.753$, $p = .030$, $\eta_p^2 = 0.277$; *Peak-to-Peak (P3-N2):* $F_{[1,15]} = 19.424$, $p < .001$, $\eta_p^2 = 0.564$; *Mean Amplitudes (360-412ms):* $F_{[1,15]} = 12.556$, $p = .003$, $\eta_p^2 = 0.456$; *Individual Mean Amplitudes:* $F_{[1,15]} = 18.801$, $p < .001$, $\eta_p^2 = 0.556$. The ***valence x type* interaction** was not significant regardless of the used method: *Base-to-Peak:* $F_{[1,15]} = 0.912$, $p = .355$, $\eta_p^2 = 0.057$, *Peak-to-Peak(P3-N2):* $F_{[1,15]} = 1.557$, $p = .231$, $\eta_p^2 = 0.094$, *Mean Amplitudes(360-412ms):* $F_{[1,15]} = 1.460$, $p = .246$, $\eta_p^2 = 0.089$, *Individual Mean Amplitudes:* $F_{[1,15]} = 0.305$, $p = .589$, $\eta_p^2 = 0.020$. Considering the **P300 latencies**, statistical analysis revealed a significant main effect of *feedback valence* ($F_{[1,15]} = 16.516$, $p < .001$, $\eta_p^2 = 0.524$) and *feedback type* ($F_{[1,15]} = 16.027$, $p < .001$, $\eta_p^2 = 0.517$) with longer latencies for negative feedback and faces. The interaction was not significant ($F_{[1,15]} = 0.716$, $p = .411$, $\eta_p^2 = 0.046$).

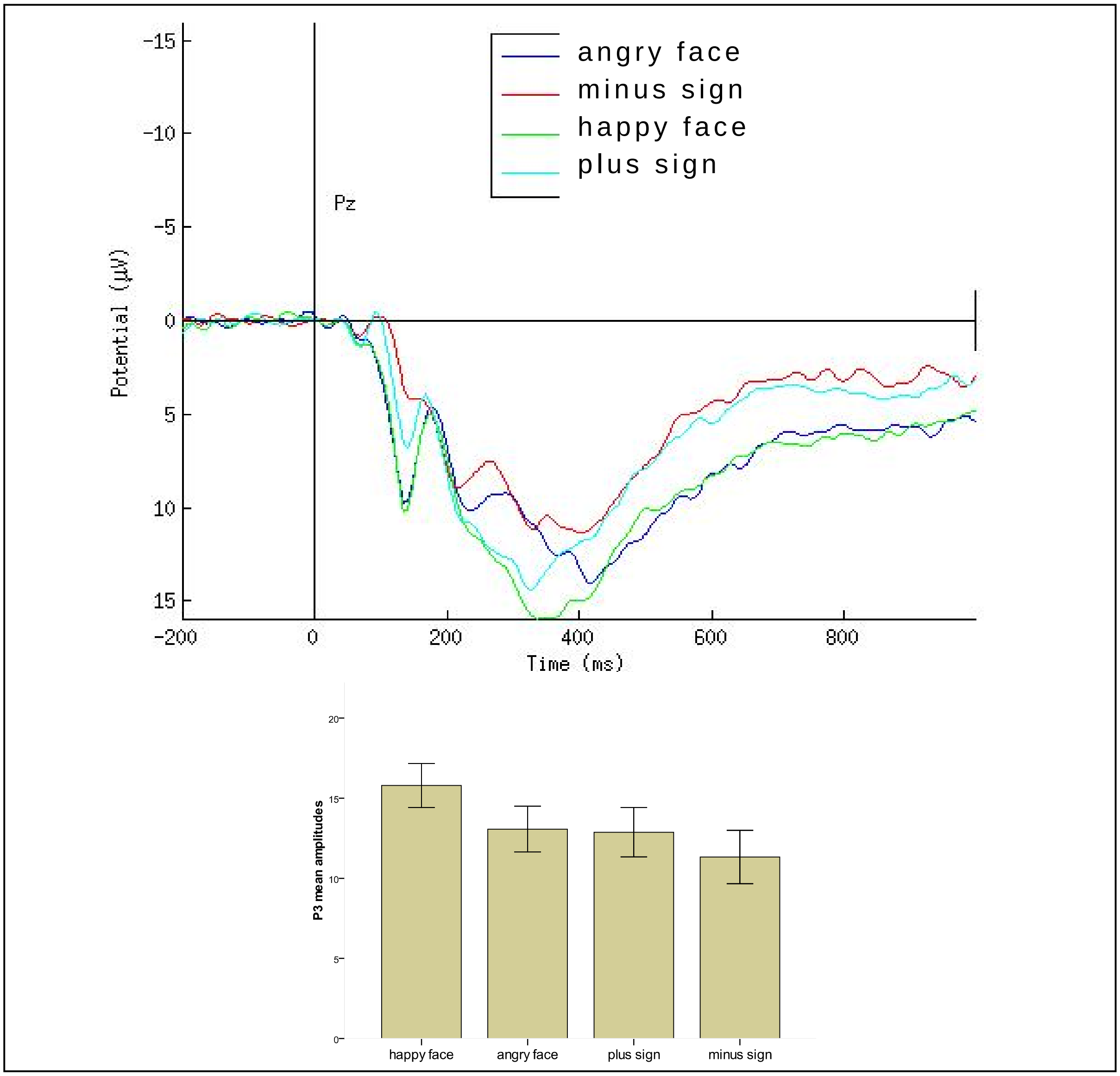

Figure 9: Grand average ERPs of Experiment 3 at electrode sites Pz for all feedback conditions. Negative is drawn upward by convention. Feedback presentation started at 0ms. The bar chart depicts the respective P300 mean amplitude values and error bars indicate 1 SE.

5.2.4 Correlation ERPs and behavioral data

The FRN amplitude (obtained by mean amplitudes at electrode Fz) across both negative feedback conditions was significantly associated with the mean trial-to-trial change in time-estimation after negative feedback trials ($r = -.634$, $n = 16$, $p = .008$). Since a low value stands for a large FRN amplitude, a large FRN was associated with a greater behavioral adjustment after negative feedback trials. In contrast, the FRN amplitude across both positive feedback conditions was not significantly related to the time estimation change after positive feedback trials ($r = -.304$, $n = 16$, $p = .252$). The correlation between the overall FRN amplitude across all conditions and the general mean of trial-to-trial changes in time estimation was significant ($r = -.522$, $n = 16$, $p = .038$).

The P300 amplitude (obtained by mean amplitudes at electrode Pz) across both negative conditions was not significantly correlated with the mean change in time estimation after negative feedback trials ($r = -.395$, $n = 16$, $p = .130$) and the correlation between the P300 in the positive feedback conditions and the time estimation change after positive feedback trials did not reach statistical significance as well ($r = -.437$, $n = 16$, $p = .090$). The correlation between the overall P300 amplitude across all conditions and the mean overall change in time estimation just failed statistical significance ($r = -.461$, $n = 16$, $p = .072$). Although no significant results were found concerning the P300, there was a tendency of larger P300 amplitudes associated with a lower change in time estimation from trial to trial.

5.3 DISCUSSION

The aim of Study 3 was to analyze FRN and P300 amplitudes as a function of positive and negative feedback which was presented in terms of symbols (minus-sign, plus-sign) and faces (angry face, happy face). The feedback type was presented in eight blocks of 50 trials each containing either symbols or faces. As in Study 2, the aim was to investigate whether the factor feedback valence and / or the factor feedback type were reflected by the FRN and / or P300. Statistical analysis revealed that the FRN was only influenced by the valence dimension but not by the feedback type and that the P300 reflected both, valence and type, with a larger P300 amplitude in the positive and facial feedback conditions.

A clear FRN peak is detectable in both negative feedback conditions in the grand average ERPs. This observation is also reflected in large effect sizes in statistical analysis. The fact that the feedback type had no influence on the FRN amplitude supports the Reinforcement learning theory (Holroyd, & Coles, 2002; Nieuwenhuis et al., 2004) which predicts that the FRN reflects a simple classification of good vs. bad outcomes. Our findings are consistent with other studies that failed to detect a larger

FRN amplitude for more salient feedback stimuli in terms of larger monetary losses (e.g. Hajcak et al., 2006). The results of the present study point against the influence of induced affect on FRN amplitudes as it was found in previous studies (Sato et al., 2005; Wiswede et al., 2009a, 2009b). We interpret the FRN effect as a general error signal that indicates whether an outcome is satisfying or not (Holroyd et al., 2006).

The behavioral results demonstrated that the participants were able to adjust their time estimation very well dependent on positive and negative feedback. As before, the feedback type had no influence on the trial-to-trial change in time-estimation. Assuming that the FRN is generated within or near the ACC (Miltner et al., 1997; Ruchsow et al., 2002), a brain area known to be relevant for error detection and compensation (Holroyd, & Coles, 2002), we suggest that the large and clear FRN amplitudes are linked to the performance in the time-estimation task. To test this additional hypothesis a correlation between individual FRN amplitudes and the individual change in time estimation after positive and negative feedback was calculated. Indeed, the FRN was significantly correlated with the change in time estimation from trial to trial. This finding provided further evidence that the FRN signals future behavioral modifications.

The P300 amplitude was larger in the positive feedback conditions and when faces were presented. This result is consistent with studies which reported that the P300 reflected both feedback valence and salience (Bellebaum et al., 2010). The longer P300 latencies in the negative and facial feedback conditions can be explained by effects of attention. According to Yamamoto et al. (2000) it was hypothesized that negative emotional faces attract more attention than positive or non-emotional stimuli. Thus, P300 latencies were prolonged for these stimuli in comparison to happy faces and non-emotional symbols.

The FRN and P300 quantification methods used generally yielded very similar results. Most of the little discrepancies can be explained by a reduced sample size because of missing FRN peaks in the positive feedback conditions.

6. EFFECT SIZE CALCULATIONS

6.1 METHODOLOGICAL PROCEDURE

In order to compare the effect sizes of FRN and P300 analysis between the three conducted studies and the applied ERP-quantification methods, the meta-analysis software CMA v2.2.030 (Comprehensive Meta-Analysis; Biostat™, Englewood, USA) was used to calculate and compare effect size estimates. Effect sizes in terms of Hedges' g[3] were calculated for the main effect of feedback valence (negative vs. positive feedback) and the main effect of feedback type (symbols vs. facial feedback). Effect sizes were analyzed for each ERP-quantification method and each study. In case of the FRN analysis, the ERPs at electrode sites Fz and FCz were averaged for effect size calculations, whereas the electrode Pz was used for the P300 effect size calculations. As a result we received an effect size value in the form of Hedges' g (with a corresponding variance and confidence interval) for both main effects and separately for each method and study. Forest plots will be presented illustrating effect sizes along with their 95% confidence intervals.

In order to compare the conducted studies with regard to the main effect of feedback valence and type, the effect sizes obtained by each method were averaged. Consequently, we received just one effect size per study for the feedback valence and the feedback type dimension. This procedure correspond to subgroup analysis used to calculate mean effect sizes for different groups of outcomes (Borenstein, Hedges, Higgins, & Rothstein, 2009). Due to dependent measurements (the same FRN and P300 amplitudes were measured repeatedly by various methods), new variances of the averaged effect sizes were calculated including the correlations between the applied methods. Treating each single method as independent and uncorrelated to the other methods would lead to an underestimation of the error and to an overestimation of the precision of the summed effect (Borenstein et al., 2009). Thus, we had to adjust the

[3] Hedges' g is an effect size that is very similar to Cohen's d. Both are based on standard difference in means. Since Cohen's d tends to overestimate the true effect in small samples, Hedges' g includes a simple correction factor (Borenstein et al., 2009). Therefore, it is recommended to use Hedges' s instead of Cohen's d in small samples. The interpretation of Hedges' effect size values are equal to the Cohen's d effect size values.

variance of the averaged effect sizes. The formula to calculate the variance of the averaged effect size was derived from Borenstein et al. (2009): *V* represents variance, *y* the effect size, *m* the number of measurements.

$$V_y = \left(\frac{1}{m}\right)^2 \operatorname{var}\left(\sum_{j=1}^{m} Y_i\right) = \left(\frac{1}{m}\right)^2 \left(\sum_{j=1}^{m} V_i + \sum_{j \neq k}\left(r_{jk}\sqrt{V_j}\sqrt{V_k}\right)\right)$$

As recommended by the authors, the averaged effect sizes with their new variances were put into a separate meta-analysis to compare the studies' outcomes by subgroup analysis. Each study formed its own subgroup. A random effects model was chosen because it was assumed that the effect sizes did not share an underlying common true effect. In order to analyze whether the effect sizes derived from the three studies were significantly different from each other or not, a homogeneity test of the effect size distribution based on a Q statistic was applied. As described by Lipsey and Wilson (2001) a statistical significant Q indicates a heterogeneous effect size distribution. This is equivalent to a difference in the effect sizes. The formula for Q (depicted in Lipsey, & Wilson, 2001) is presented below. ES_i represents the effect size of the i^{th} study, $\overline{ES}$ is the weighted mean effect size across all effect sizes, and w_i is the individual weight for the respective effect size.

$$Q = \sum w_i \left(ES_i - \overline{ES}\right)^2$$

Subgroup analyses were also performed concerning the research question whether or not the ERP-quantification methods produce significantly different effect sizes. In that case, each subgroup contained all outcomes of one applied method. For example, the base-to-peak effect sizes concerning the P300 main effect of feedback valence of Study 1, Study 2 and Study 3 formed one subgroup and all effect sizes obtained by the mean amplitude method formed another subgroup. In contrast to above, independent measurements in each subgroup were assumed because each subgroup just contained outcomes of different studies. Therefore, no adjustment of the variance of the summed effect sizes was necessary. Once again, a heterogeneity analysis was done in order to analyze if the effect sizes provided by the used methods were significantly different from each other.

Additionally, Pearson intercorrelation coefficients between the FRN amplitude values obtained by different FRN quantification methods and between P300 amplitude values obtained by different P300 quantification methods were calculated. At the first step, such an intercorrelation matrix was created separately for each study and condition. Due to the observation that those matrices were very similar, a mean intercorrelation matrix (across all studies and conditions) was calculated, each correlation weighted by its underlying sample size.

6.2 RESULTS

6.2.1 FRN

Figure 10 shows the effect sizes concerning the main effect of feedback valence. It can be seen that, excepting for one measurement in Study 1, all effect sizes were positive, meaning a larger FRN amplitude following negative compared to positive feedback. Regarding Study 2 and 3, no confidence interval is including zero, which indicates that a significant result was reached independently from the method used.

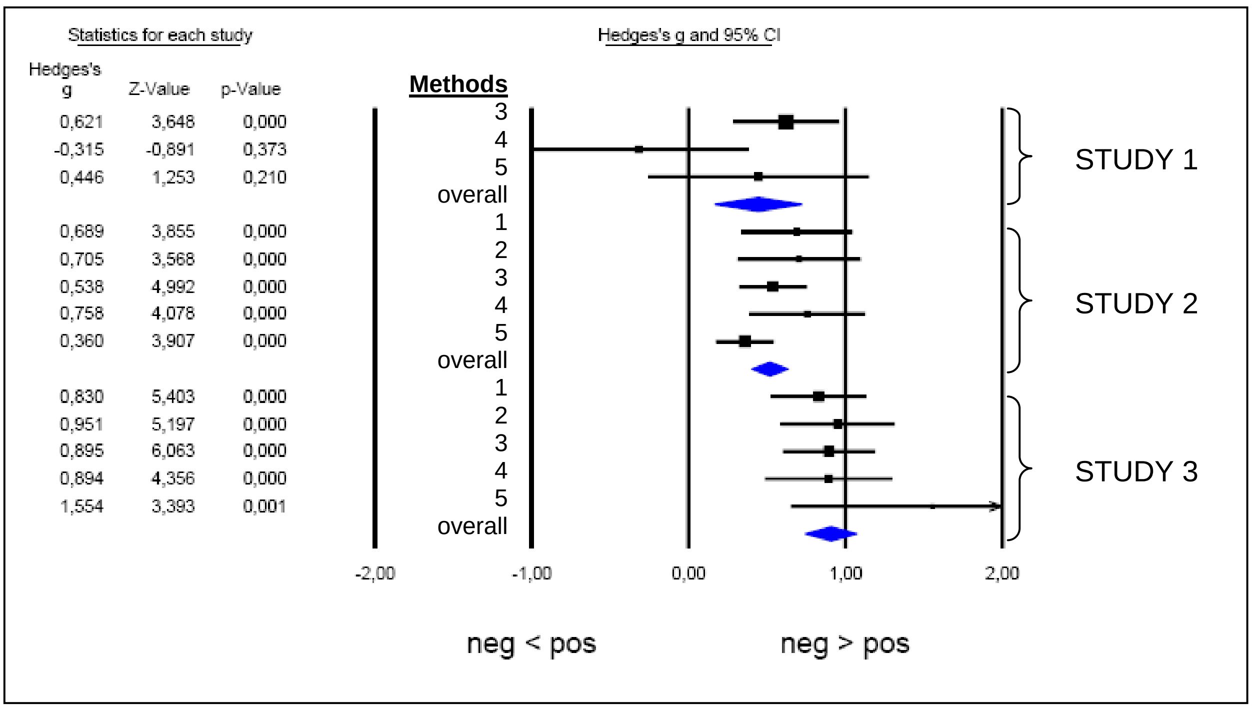

Figure 10: Effect sizes of the FRN main effect of feedback valence: A positive effect size (right side of the plot) means a larger FRN amplitude following negative than positive feedback. The black square represents the Hedges' g point estimate, the line the 95% confidence interval. The blue rhombus represents the summed mean effect of each study. Note that this is uncorrected for dependent measurements.
Abbreviations (Methods): 1… Base-to-Peak, 2…Individual Mean Amplitudes, 3…Mean Amplitudes (MW, SD), 4…Peak-to-Peak (P2-FRN), 5…Peak-to-Peak (P3-FRN)

The overall effects for each study corrected for dependent measurements are depicted in figure 11. The overall effect of Study 3 (Hedges' g = 1.025) was larger than in Study 2 (Hedges' g = 0.610) and in Study 1 (Hedges' g = 0.251). The heterogeneity analysis revealed a significant difference between those effect sizes (Q = 10.677, df = 2, p = .005). Only regarding Study 2 and 3, a significantly larger main effect of feedback valence was found for Study 3 compared to Study 2 (Q = 5.720, df = 1, p = .017). The overall valence effect across all three studies was significant (Hedges' g = 0.650, z = 3.381, p = .001).

Study name	Statistics for each study						
	Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Study1	0,251	0,206	0,042	-0,153	0,655	1,219	0,223
Study2	0,610	0,098	0,010	0,417	0,803	6,197	0,000
Study3	1,025	0,143	0,020	0,745	1,305	7,177	0,000
	0,650	0,192	0,037	0,273	1,026	3,381	0,001

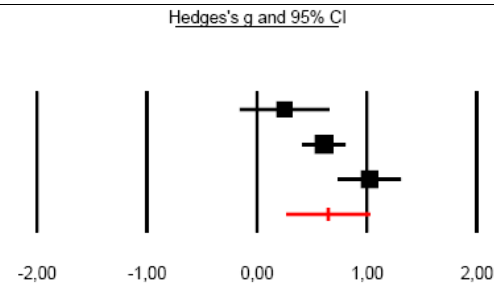

Figure 11: Corrected overall effect sizes for the FRN main effect of feedback valence: The black squares and lines represent the overall valence effects for each experiment (across all methods) in terms of Hedges' g with their 95% confidence intervals. The red line represents the overall valence effect across all three studies.

Considering the main effect of feedback type, all effect sizes except one (the Individual Mean Amplitude method in Study 3) were positive indicating a greater FRN effect for symbol feedback than for facial feedback (Figure 12). As easily recognizable from the plot below, no measurement reached statistical significance in Study 3, whereas the majority of the measurements in Study 2 did.

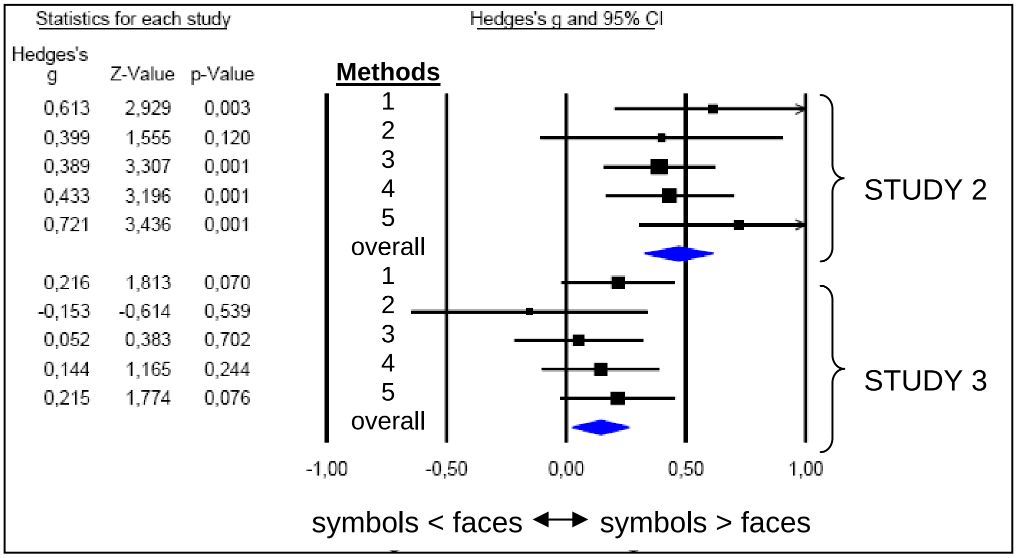

Figure 12: Effect sizes of the FRN main effect of feedback type: A positive effect size (right side of the plot) means a larger FRN amplitude following symbols than facial feedback. The black square represents the Hedges' g point estimate, the line the 95% confidence interval. The blue rhombus represents the summed mean effect of each study. Note that this is uncorrected for dependent measurements.
Abbreviations (Methods): 1… Base-to-Peak, 2…Individual Mean Amplitudes, 3…Mean Amplitudes (MW, SD), 4…Peak-to-Peak (P2-FRN), 5…Peak-to-Peak (P3-FRN)

The corresponding corrected overall effects are presented in Figure 13. The main effect of feedback type was significantly larger in Study 2 (Hedges' g = 0.511) than in Study 3 (Hedges' g = 0.095): $Q(df = 1) = 7.459$ $p = .006$. The overall effect across both studies (Hedges' g = 0.226) was significant (z = 3.544, $p < .001$) indicating a larger FRN amplitude following symbols than faces.

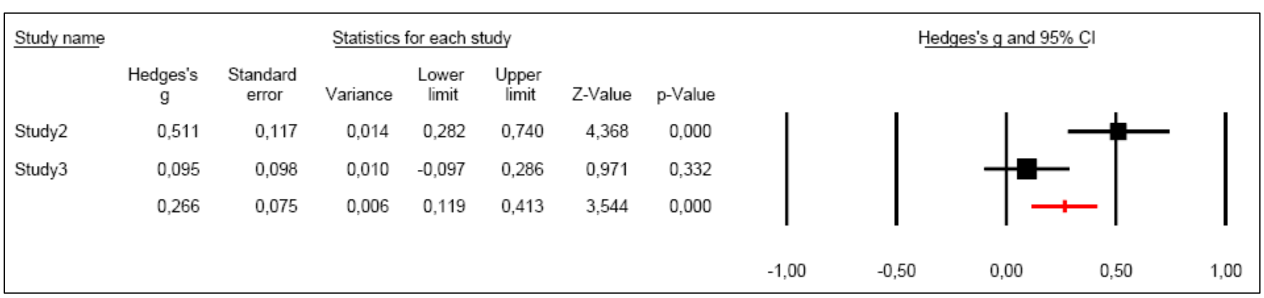

Figure 13*: Corrected overall effect sizes for the FRN main effect of feedback type:* The black square and lines represent the overall feedback type effects for each experiment (across all methods) in terms of Hedges' g with their 95% confidence intervals. The red line represents the overall feedback type effect across all three studies.

Subgroup analyses comparing the five FRN quantification methods used yielded a non-significant result considering the main effect of feedback valance (Q = 1.734, df = 4, p = .785), as well as for the main effect of feedback type (Q = 1.510, df = 4, p =.825). The intercorrelation matrix including all FRN amplitude values obtained by different methods correlated with each other is presented in Table 3. A very high intercorrelation between the base-to-peak method and both mean-amplitude measures is remarkable, as well as very low to null correlations between the Peak-to-Peak (P3-FRN) method with the Base-to-Peak method and both mean amplitude methods. The correlation between the two Peak-to-Peak methods was middle to high.

Table 3: Intercorrelations of FRN amplitude values obtained by different quantification methods (across all studies and conditions) electrode sites Fz

Method	1	2	3	4	5
1 Base-to-Peak	-				
2 Peak-to-Peak (P2-FRN)	-.255*	-			
3 Peak-to-Peak (P3-FRN)	- .027	.586***	-		
4 Mean Amplitudes (Mean, SD)	.972***	-.314**	-.128	-	
5 Individual Mean Amplitudes	.994***	-.373**	-.065	.959***	-

Note: * p < .05, ** p < .01, ***p < .001
Negative correlations can be explained because of the fact that using method 1, 4 and 5 a low and using method 2 and 3 a high value represents a large FRN amplitude.

6.2.2 P300

Considering the main effect of feedback valence, the P300 amplitude was consistently larger in the positive compared to the negative feedback condition (Figure 14) which reached statistical significance in Study 2 regardless of the method used, for the majority of measurements in Study 3, but not in Study 1.

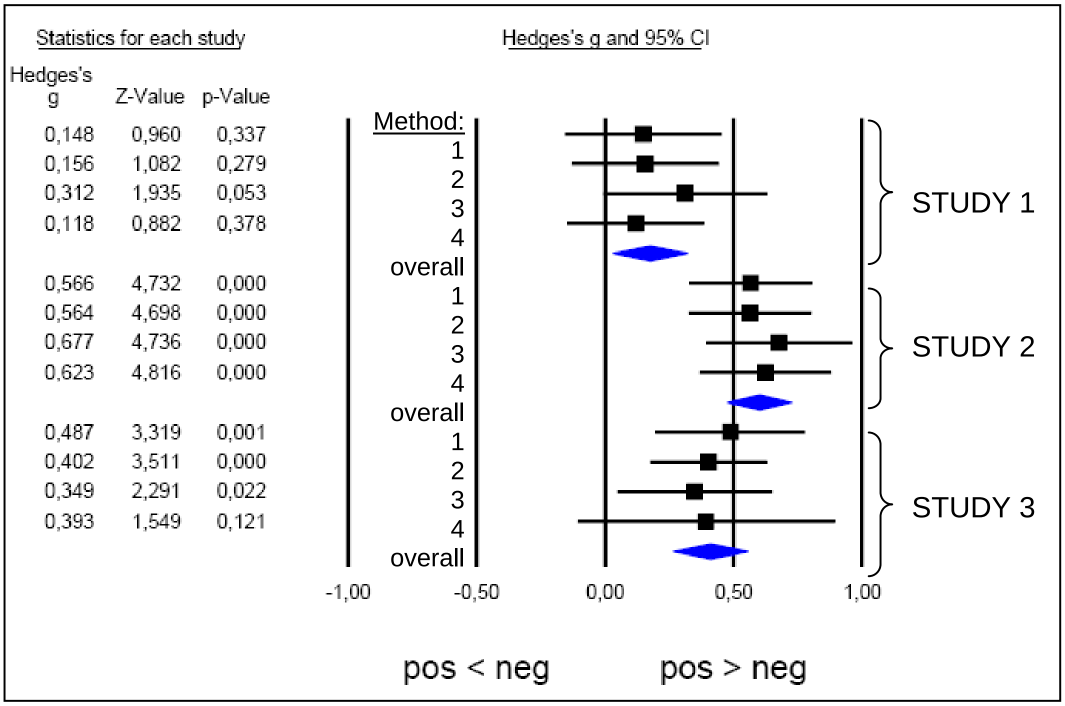

Figure 14: Effect sizes of the P300 main effect of feedback valence: A positive effect size (right side of the plot) represents a larger P300 amplitude following positive feedback than negative feedback. The black square represents the Hedges' g point estimate, the line the 95% confidence interval. The blue rhombus represents the summed mean effect of each study. Note that this is uncorrected for dependent measurements.
Abbreviations (Methods): 1… Base-to-Peak, 2…Individual Mean Amplitudes, 3…Mean Amplitudes (MW, SD), 4…Peak-to-Peak (P3-N2)

Corrected for dependent measurements, there was an overall effect (Hedges' g) of 0.184 in Study 1, 0.608 in Study 2 and 0.408 in Study 3 (Figure 15). The difference between those effect sizes was significant (Q = 8.606, df = 2, p = .014). Just comparing Study 2 and 3, the effect size difference failed to reach statistical significance (Q = 1.679, df = 1, p = .195). The overall P300 valence effect across all three studies was significant (Hedges' g = 0.424, z = 6.877, p < .001).

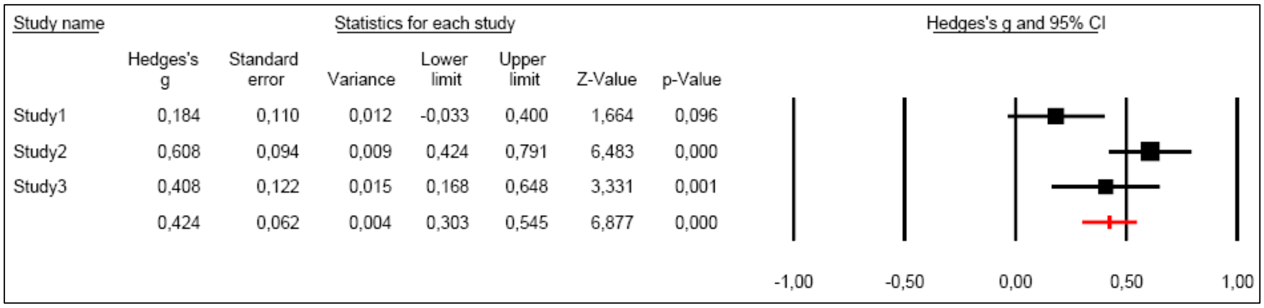

Figure 15: Corrected overall effect sizes for the P300 main effect of feedback valence: The black squares and lines represents the overall valence effects for each study (across the methods) in terms of Hedges' g with their 95% confidence intervals The red line represents the overall effect across all studies.

Regarding the P300 main effect of feedback type, a consistent tendency of a larger P300 amplitude after facial feedback compared to symbol feedback was observed across all studies and methods (Figure 16).

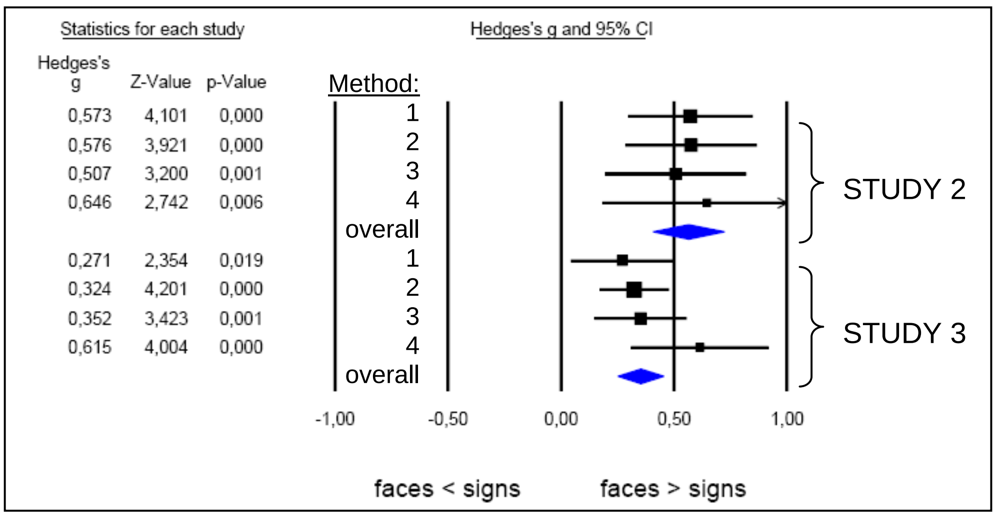

Figure 16: Effect sizes of the P300 main effect of feedback type: A positive effect size (right side of the plot) means a larger P300 amplitude following facial feedback than symbols. The black square represents the Hedges' g point estimate, the line the 95% confidence interval. The blue rhombus represents the summed mean effect of each study. Note that this is uncorrected for dependent measurements.
Abbreviations (Methods): 1… Base-to-Peak, 2…Individual Mean Amplitudes, 3…Mean Amplitudes (MW, SD), 4…Peak-to-Peak (P3-N2)

There was no significant difference between the corrected overall effects of Study 2 (Hedges' g = 0.576) and Study 3 (Hedges' g = 0.391): $Q(df = 1) = 0.057$, $p = .812$. The overall effect across both studies (Hedges' g = 0.424) was significant (z = 5.662, $p < .001$) indicating a larger P300 amplitude following facial feedback (figure 17).

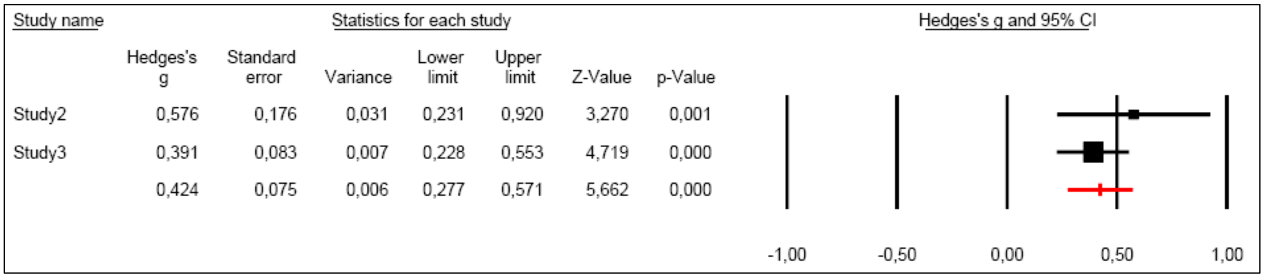

Figure 17: Corrected overall effect sizes for the P300 main effect of feedback type: The black squares and lines represents the overall feedback type effects for each study (across the methods) with their 95% confidence interval. The red line represents the overall effect across all studies.

Subgroup analysis comparing the P300 quantification methods yielded a non-significant result, neither for the main effect of feedback valence (Q = 0.198, df = 3, p = .978), nor for the main effect of feedback type (Q = 1.990, df = 3, p = .575). Table 4 depicts the correlation matrix, including the intercorrelations between the P300 amplitude values obtained by different quantification methods. The base-to-peak method and both mean-amplitude measurements intercorrelated highly. The correlations of the peak-to-peak method with the other methods were on a medium to high level. In contrast to the FRN quantification methods, all methods for the P300 quantification intercorrelated significantly with each other.

Table 4*:* Intercorrelations of P300 amplitude values obtained by different quantification methods (across all studies and conditions) at electrode sites Pz

Method	1	2	3	4
1 Base-to-Peak	-			
2 Peak-to-Peak (P3-N2)	.622**	-		
3 Mean Amplitudes (Mean, SD)	.945***	.570***	-	
4 Individual Mean Amplitudes	.991***	.620***	.964***	-

Note: * p < .05, ** p < .01, ***p < .001

7. GENERAL DISCUSSION

7.1 Summary of research intentions and main results

The conducted three studies investigated the influence of two feedback types (symbols and faces) on event-related brain potentials, the Feedback-Related-Negativity (FRN) and P300. Negative and positive feedback was presented as a function of participants' performance in a time estimation task. In the first study, feedback was provided in terms of symbols (plus-sign, minus-sign), whereas symbols and faces (happy face, angry face) were presented in Study 2 and 3. We used two different feedback presentation modes. The feedback type either changed randomly from trial to trial (Study 2) or remained constant within blocks of 50 trials (Study 3). We were interested in the effects of feedback valence, feedback type and possible interaction effects and how they differ from study to study. We used different methods to quantify FRN and P300 amplitudes in order to investigate whether or not they are equally suitable for detecting differences between experimental conditions.

Due to unusual noisy ERPs in the first study with absent FRN peaks for the majority of participants, we concentrate our interpretations on Study 2 and 3. The valence of feedback had an impact on the FRN and P300 amplitudes. The FRN amplitude was larger in the negative feedback conditions. This effect was not significant in the first study and significantly greater in the third compared to the second study. As opposed to this, the P300 amplitude was larger in the positive feedback conditions. This effect reached statistical significance in Study 2 and 3. The feedback type had an influence on the P300 amplitude, but only partially on the FRN amplitude. There was a tendency of larger FRN amplitudes after symbols compared to faces, which was only significant in Study 2. Furthermore, the effect size of the feedback type dimension was significantly larger in Study 2 than in Study 3. In contrast, the P300 amplitude was larger after facial feedback independent of the feedback valence. No latency effects were observed regarding the FRN. Results concerning the P300 latencies were quite inconsistent across the three studies, but there seems to be a tendency of a longer latency in the negative and facial feedback conditions.

The FRN and P300 quantification methods did not differ significantly or systematically from each other in respect of the produced effect sizes. Focusing on the studies separately, the used methods partially differed in respect of whether a result reached statistical significance or not. Whereas the P300 amplitude values obtained by the different methods all intercorrelated with $r = .570$ at least, some FRN amplitude values revealed by different FRN quantification methods failed to correlate with each other.

7.2 Behavioral data

The task-performance data of all three studies demonstrated that the feedback valence affected the subsequent behavior. As it was hypothesized, participants changed their time estimation more after negative than after positive feedback trials. This finding is consistent with previous studies that used the same time estimation task (Holroyd, & Krigolson, 2007; Miltner et al., 1997; Van der Veen et al., 2011; Van Veen et al., 2004). Furthermore, the participants' time estimation change after negative feedback approached 1000ms in the majority of cases. These results demonstrated that the participants adequately used the feedback to guide and adjust their behavior. Interestingly, the behavioral performance was worse in Study 2 than in Study 1 and 3 which was reflected in a larger time window required for positive feedback and a lower mean time estimation change in trials after negative feedback in Study 2. It can be hypothesized that the perpetual change of the feedback type in Study 2 was responsible for the difference in the behavioral performance by distracting attentional resources. The behavioral adjustments were uninfluenced by the feedback type which is consistent with other studies where the feedback modality (Miltner at al., 1997) and different feedback types (Van der Veen et al., 2011) had no influence on the task performance. We conclude that the emotional value of a feedback stimulus has no influence with regard to the behavioral data. We might assume that the behavioral results were a function of a rational “correct vs. incorrect” classification of the feedback stimulus. However, this assumption can only be made for the time estimation task used. In fact, it cannot be ruled out that the feedback type would not have had an influence when, for example, a gambling task had been used instead where subjects could win or loose money.

The association of behavioral results with the ERPs is discussed below.

7.3 FRN and P300

As it was expected, the FRN amplitude was generally larger in the negative than in the positive feedback conditions. This provides further evidence that the FRN serves as an error signal indicating that an outcome is not satisfying (Hajcak et al., 2006; Miltner et al., 1997; Yeung, & Sanfey, 2004). This effect was not obviously observed in Study 1 which could possibly be explained by problems during data recording or a noisy EEG signal that complicated the identification of FRN peaks.

Surprisingly, the feedback type had an influence on the FRN amplitude in Study 2 but not in Study 3 and, contrary to our hypothesis there was the tendency of larger FRN amplitudes in the symbol feedback conditions. This result did not confirm previous studies that observed an enhanced FRN amplitude when a negative emotion was induced experimentally (Wiswede et al., 2009a, 2009b) and for individuals with a general negative affective trait (Sato et al., 2005). It has to be noted that none of those studies induced positive or negative affect by a feedback stimulus itself. This point discriminates those studies from our study design because we tried to induce positive and negative emotions by the presentation of emotional facial expressions as feedback stimuli. This may be the reason for this unexpected result. It can be presumed that emotions have just long-term effects on the FRN in the sense of a general higher sensitivity to negative feedback for subjects with a negative affective trait. This view is supported by further studies demonstrating that a negative trait affect is associated with enhanced error-related brain activity (Hajcak, McDonald, & Simons, 2004; Luu et al., 2000) and by clinical studies showing enhanced FRN amplitudes for patients with anxiety disorders and depression (Olvet, & Hajcak, 2008). We suggest that the FRN amplitude variation reflects rather affective traits than affective states that are induced by experimental manipulation. Further studies have to address this question by the combination of affective trait and state aspects.

Furthermore, it was assumed that the higher feedback salience of greater monetary rewards in gambling tasks is comparable to the higher feedback salience of emotional faces compared to non-emotional symbols. Using gambling tasks, the FRN was associated with higher monetary rewards or losses in a few studies (e.g. Kreussel et al., 2012; Santesso et al., 2011), but no study was identified where the effect was the other way round, with larger FRN amplitudes in conditions with lower monetary rewards or losses. Thus, our findings (larger FRN amplitudes after symbols compared to faces) are surprising in respect of the present literature. We conclude that the reward magnitude dimension in gambling tasks and the feedback type dimension (with affective and non-affective stimuli) in the current study are not situated on a common feedback salience dimension and cannot be compared directly.

According to our results, the feedback presentation mode was identified as a moderator variable of the feedback valence and type effects. The main effect of feedback valence was significantly larger in Study 3 compared to Study 2. By contrast, the effect of the feedback type was larger in Study 2. These findings demonstrate that the FRN also reflects the frequency of the change of the feedback type within an experiment. If the FRN codes a binary evaluation of good vs. bad outcomes, which is strongly supported by the results of Study 3, the findings of Study 2 suggest that this good vs. bad classification is complicated by a frequent change of feedback types within an experiment. The perpetual change of the feedback type prevents from developing expectations about which type of feedback stimulus would occur in the following trial. Previous studies demonstrated that the FRN amplitude also varies as a function of the feedback expectation level (Hajcak et al., 2007; Liao et al., 2011). It is conceivable that the difference in the stimulus complexity between the two feedback types had an impact on the FRN effects in Study 2 as well. The stimulus complexity was identified as a moderating variable in relation to visual evoked ERPs with larger anterior negativity (N2 component) and larger posterior positivity (late positive potential) by Shigeto, Ishiguro, and Nittono (2011). It may be that a good vs. bad outcome classification can be performed easier with symbols than with more complex facial stimuli which could have influenced the results of Study 2. It can be argued that the factor of stimulus complexity did not play a role in Study 3 where the participants could habituate one feedback type for a longer time interval. However, further studies have to clarify the role of stimulus complexity in context of feedback-related brain potentials.

The P300 amplitudes were generally larger after positive feedback. This result supports those findings in previous research which demonstrated that the P300 is influenced by the valence of feedback (Bellebaum et al., 2010; Hajcak et al., 2007; Kreussel et al., 2012) and argues against theories saying that the feedback valence is only reflected in the FRN amplitudes (Yeung, & Sanfey, 2004). Additionally, the P300 amplitude was larger after facial feedback than after symbols regardless of the feedback valence. This finding is in line with a range of previous studies that found an effect of reward magnitude on the P300 (Bellebaum et al., 2010; Moser et al., 2004; Wu, & Zhou, 2009). It is presumable that facial feedback stimuli have an enhanced motivational significance compared to non-emotional symbols which is reflected in the P300. This argumentation is supported by the theory that the P300 reflects the activity of the locus coeruleus–norepinephrine system which was associated with the motivational significance of events (Nieuwenhuis, Aston-Jones, & Cohen, 2005). Motivational significant events are associated with larger P300 amplitudes compared to less significant events. Nevertheless, further studies have to include considerations about a possible influence of stimulus complexity on the P300. It cannot be ruled out that the P300 main effect of feedback type were found just because of the varying stimulus complexity of symbols and faces. Reviewing studies about emotional and non-emotional picture processing (Keil et al., 2002), the most presumable explanation for feedback type differences in the P300 seem to be differences in the affective and motivational value of the feedback stimuli.

Although there were no significant differences between Study 2 and 3, the P300 effect sizes tend to be larger in Study 2. As P300 effects of expectation were found previously (Hajcak et al., 2005, 2007; Pfabigan et al., 2011b), it is conceivable that expectation effects also played a role in Study 2 in respect of to the uncertainty which feedback type would occur in the next trial.

P300 latency effects were completely inconsistent across the three conducted experiments. In Study 1 there were no latency effects, in Study 2 only the interaction valence x type was significant and in Study 3 the latencies were significantly longer in the negative and facial feedback conditions. No general interpretation can be made according to our findings. The results of Study 3 come the closest to our hypothesis which said that the P300 latency is longer after negative feedback. Such latency effects were previously found by Yeung et al. (2005) and Pfabigan et al. (2011b). As the P300 latency reflects the evaluation process of a stimulus (Polich, 2007) it can be argued that negative feedback and facial feedback tend to be evaluated more thoroughly. However, the impact of feedback valence and type on the P300 latencies remains still unresolved.

An important issue we tried to address with our research project was to make a contribution to the research question whether feedback valence and salience are processed by distinct or similar neural mechanism. According to the results of Experiment 2 and 3, the independent coding theory (Yeung, & Sanfey, 2004) cannot be supported. The P300 has been shown to be sensitive to the feedback valence and type dimensions in both experiments and the feedback type had an influence on the FRN in Study 2 which is not consistent with this theory. Although the latter possibly can be explained by the experimental procedure, our results demonstrated that feedback valence and type are reflected by both the FRN and P300 in some way. Our data indicate that the FRN and P300 are influenced by multiple factors as it was already proposed by Goyer et al. (2008).

Due to the observation that a worse behavioral performance in the time estimation task in Study 2 was accompanied by ambiguous FRN peaks and that a better behavioral performance in Study 3 came along with obvious FRN peaks in both negative conditions, we created the additional hypothesis that the FRN amplitude might be associated with the behavioral performance. Therefore, we correlated the individual FRN amplitude with the individual trial-to-trial change in time-estimation. We found that the individual FRN amplitudes were highly correlated with the performance data in Study 3. Larger FRN amplitudes were associated with a greater mean change in time estimation from trial-to-trial. A significant association between the individual FRN amplitudes and the task performance was also observed in previous studies (Holroyd, & Krigolson, 2007; Van der Helden, Boksem, & Blom, 2010). This is consistent with the assumption that the ACC, which was identified as the most likely source of the FRN, is responsible for the transformation of motor intentions into motor responses and behavioral modification (Holroyd, & Coles, 2002). By contrast, there were no significant correlations of the FRN amplitudes and performance data in Study 2. It is possible that the lower FRN valence effect in Study 2 was linked to the worse behavioral performance which could have been the result of the permanently change of the feedback type in Study 2. This exploratory hypothesis has to be investigated in further studies. The individual P300 amplitude was not significantly correlated with the performance data, neither in Study 2 nor in Study 3.

The differences between Study 2 and 3 can be summarized as follows. The FRN main effect of feedback valence was lower in Study 2 accompanied by smaller FRN amplitudes in the angry-face condition compared to the minus-sign condition. The P300 effects tended to be larger in Study 2. The participants' time-estimation performance was worse in Study 2 and, in contrast to Study 3, there was no association between the individual FRN amplitudes and the behavioral performance. These facts demonstrate that the feedback presentation mode has an important impact on different levels of the data analysis. This should be considered whenever different types of feedback stimuli are compared in further experiments.

7.4 Methodological issues of ERP-quantification

We have to discuss methodical differences of ERP quantification at three levels. The first perspective focuses on the level of the individual ERPs. Handy (2005) and Picton et al. (2000) made clear that the choice of the appropriate quantification method depends on the characteristics of the analyzed components. This became apparent in our studies. A P300 peak could have been clearly identified in the majority of cases, whereas there were sometimes difficulties to clearly determine an FRN peak. The identification of the smaller FRN component was often complicated by noise in the EEG signal. Regarding the FRN, the base-to-peak method and the individual mean amplitude method were at a disadvantage compared to the other methods used because of absent FRN peaks primarily in the positive feedback conditions. An absent peak results in a missing value (i.e., a missing participant in the analysis) using those methods. The peak-to-peak analysis provided the opportunity to set the amplitude value at zero if there is no peak and mean amplitudes provided an amplitude value in each case. Along these lines, the peak-to-peak method and mean amplitudes have an advantage over methods that requires the presence of a peak.

Secondly, methodological difference should be regarded on the level of one study. Statistical analysis demonstrated that a main effect or interaction effect can be statistically significant or not depending on which method was used. The disadvantages of the base-to-peak and the individual mean amplitude methods became apparent at this level. Missing FRN peaks in some conditions had an impact on the statistical results due to a partially reduced sample size. Thus, it is not surprising that those methods could not have been applied in Study 1 and that we found differing results in Study 3 when the FRN main effect of feedback type was analyzed. The significant results using the base-to-peak and the individual mean amplitude method can be labeled as “chance findings” that occur because of a reduced sample size. Study 1 demonstrated that mean amplitudes were able to find a statistically significant FRN main effect of feedback valence although there were no obvious FRN peaks in the majority of cases. Thus, this result is the consequence of a general larger negativity in the negative feedback condition. It may be an advantage of mean amplitudes to detect differences between conditions for smaller components. In Study 2, the peak-to-peak methods yielded a non-significant result regarding the FRN main effect of feedback type in contrast to the other methods used. This can be explained by not stable P300 and P200 peaks across the conditions to which the FRN is compared. Regarding the P300 analysis, no meaningful differences between the methods used were found. Merely the interaction effect in Study 2 could not have been consistently found by all methods and the peak-to-peak method failed to reveal a significant P300 main effect of feedback valence in Study 3 contrary to the other methods used.

Thirdly, differences of ERP quantification methods have to be regarded on the level of several studies. It was analyzed whether any method yielded higher effect sizes compared to the others across our three studies. No significant differences between the effect sizes were observed, neither for the FRN- nor P300 quantification methods. This indicates that there is no general tendency that one method is “better” or “worse” than the other ones. A view on the effect size plots reveals that the effect sizes did not differ systematically across the three studies. In summary, it can be stated that differences between the different methods are small, non-significant, non-systematic, and more important for the FRN than for the P300. Although there were no significant differences between the methods, additional correlation analysis surprisingly revealed that there were partly very low correlations between the amplitude values obtained by the single methods. This was especially true for the FRN quantification methods where the peak-to-peak methods failed to correlate with the other methods. This result was found across all conditions and studies and raises the question whether the methods used measure something basically different. Further investigations are needed to address this question but it can be assumed that the low intercorrelations between the FRN quantification methods resulted from setting FRN amplitudes at zero using peak-to-peak analysis. Another explanation of the low inter-correlations could be the non-stable P200 and P300 amplitudes across the conditions to which the FRN peak is compared. In contrast to the FRN, the P300 quantification methods highly intercorrelated which confirms the perception that the P300 is a large component which can be easily identified by each method.

To sum up, different ERP quantification methods clearly differ in respect of how absent and non-clearly recognizable peaks can be handled. Especially for FRN analysis, the choice of the method is crucial in respect of whether a result reaches statistical significance or not, which can lead to different conclusion when the results are interpreted. We therefore recommend a synopsis of several methods and descriptive data. Finally, there seem to be no general difference with regard to the effect sizes obtained by different methods.

7.5 LIMITATIONS

A few limitations of the present study have to be mentioned. First, peak finding was done by a single person. Just when comparing various ERP-quantification methods, at least a second person performing peak finding independently from the other one, would increase the reliability of ERP analysis. However, the very similar results by the use of different methods are indicative of a sufficient precision during peak finding. Secondly, three studies seem to be insufficient to detect possible significant differences between the effect sizes obtained by different methods. A meta-analytic approach including existing ERP studies with various ERP quantification methods used could help to address this question. Thirdly, due to bad EEG data quality in Study 1, which was not satisfactorily compensated by artifact correction, the validity of Study 1 is limited. Thus, we had to focus our attention on the interpretation of Study 2 and 3.

7.6 CONCLUSION

To conclude, feedback valence and salience do not seem to be processed by strictly separate neural mechanism. The FRN seems to reflect a binary good vs. bad distinction of ongoing events which is complicated by a permanently change of the feedback type within an experiment, whereas the P300 reflects both, feedback valence and salience.

Multiple factors like feedback valence, type and the emotional value of a feedback stimulus have an influence on feedback-related brain potentials and should be considered in further experiments. The role of feedback stimulus complexity and the association between the individual ERPs and task performance have to be investigated in future studies. Due to well-recognizable ERPs and a better behavioral performance in Study 3 compared to Study 2, we recommend to use a blocked presentation mode when different feedback types are compared within one study.

Different methods of ERP quantification are differentially suitable for quantifying FRN and P300 components even though there are no systematic differences in the effect sizes elicited by different methods. However, a synopsis of different ERP quantification methods can avoid possible misinterpretations and enhance the validity of conclusions drawn by statistical results.

8. REFERENCES

Bauer, H., & Lauber W. (1979). Operant conditioning of brain steady potential shifts in man. *Biofeedback and Self-Regulation, 4*, 145-154. doi: 10.1007/BF01007109

Bellebaum, C., Polezzi, D., & Daum, I. (2010). It is less than you expected: The feedback-related negativity reflects violations of reward magnitude expectations. *Neuropsychologia, 48*(11), 3343-3350. doi: 10.1016/j.neuropsychologia.2010.07.023

Bocker, K. B. E., Van Avermaete, J. A. G., & Van Den Berg-Lenssen, M. M. C. (1994). The international 10-20 system revisited: Cartesian and spherical co-ordinates. *Brain Topography, 6*(3), 231-235. doi: 10.1007/bf01187714

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. Hoboken, NJ: John Wiley & Sons.

Briggs, K. E., & Martin, F. H. (2009). Affective picture processing and motivational relevance: Arousal and valence effects on ERPs in an oddball task. *International Journal of Psychophysiology, 72*(3), 299-306. doi: 10.1016/j.ijpsycho.2009.01.009

Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Sciences, 4*(6), 215-222. doi: 10.1016/s1364-6613(00)01483-2

Cohen, B. H. (2007). *Explaining Psychological Statistics* (3rd ed.). New York: John Wiley & Sons.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Cohen, M. X., Elger, C. E., & Ranganath, C. (2007). Reward expectation modulates feedback-related negativity and EEG spectra. *Neuroimage, 35*(2), 968-978. doi: 10.1016/j.neuroimage.2006.11.056

Compton, R. J., Dainer-Best, J., Fineman, S. L., Freedman, G., Mutso, A., & Rohwer, J. (2010). Anxiety and expectancy violations: Neural response to false feedback is exaggerated in worriers. *Cognition & Emotion, 24*(3), 465-479. doi: Pii 908297586

Dehaene, S., Posner, M. I., & Tucker, D. M. (1994). Localization of a neural system for error-detection and compensation. *Psychological Science, 5*(5), 303-305. doi: 10.1111/j.1467-9280.1994.tb00630.x

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9-21. doi: 10.1016/j.jneumeth.2003.10.009

Donkers, F. C. L., Nieuwenhuis, S., & van Boxtel, G. J. M. (2005). Mediofrontal negativities in the absence of responding. *Cognitive Brain Research, 25*(3), 777-787. doi: 10.1016/j.cogbrainres.2005.09.007

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17*(2), 124-129. doi: 10.1037/h0030377

Ekman, P., & Friesen, W. V. (1978). *Facial action coding system*. Palo Alto, CA: Consulting Psychologists Press.

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

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. doi: 10.1111/j.1467-9280.1993.tb00586.x

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. doi: 10.1126/science.1066893

Gibson, J., Krigolson, O. E., & Holroyd, C. B. (2006). Sensitivity of the feedback error-related negativity to reward probability. *Psychophysiology, 43*, S41-S42.

Goyer, J. P., Woldorff, M. G., & Huettel, S. A. (2008). Rapid Electrophysiological Brain Responses are Influenced by Both Valence and Magnitude of Monetary Rewards. *Journal of Cognitive Neuroscience, 20*(11), 2058-2069. doi: 10.1162/jocn.2008.20134

Gu, R., Ge, Y., Jiang, Y., & Luo, Y.-j. (2010). Anxiety and outcome evaluation: The good, the bad and the ambiguous. *Biological Psychology, 85*(2), 200-206. doi: 10.1016/j.biopsycho.2010.07.001

Gu, R., Lei, Z., Broster, L., Wu, T., Jiang, Y., & Luo, Y.-j. (2011). Beyond valence and magnitude: A flexible evaluative coding system in the brain. *Neuropsychologia, 49*(14), 3891-3897. doi: 10.1016/j.neuropsychologia.2011.10.006

Handy, T. C. (2005). Basic principles of ERP quantification. In T. C. Handy (Ed.), *Event-related potentials: A methods handbook* (pp. 33-55). Cambridge, MA: MIT Press.

Hajcak, G., Holroyd, C. B., Moser, J. S., & Simons, R. F. (2005). Brain potentials associated with expected and unexpected good and bad outcomes. *Psychophysiology, 42*(2), 161-170. doi: 10.1111/j.1469-8986.2005.00278.x

Hajcak, G., McDonald, N., & Simons, R. F. (2004). Error-related psychophysiology and negative affect. *Brain and Cognition, 56*(2), 189-197. doi: 10.1016/j.bandc.2003.11.001

Hajcak, G., Moser, J. S., Holroyd, C. B., & Simons, R. F. (2006). The feedback-related negativity reflects the binary evaluation of good versus bad outcomes. *Biological Psychology, 71*(2), 148-154. doi: 10.1016/j.biopsycho.2005.04.001

Hajcak, G., Moser, J. S., Holroyd, C. B., & Simons, R. F. (2007). It's worse than you thought: The feedback negativity and violations of reward prediction in gambling tasks. *Psychophysiology, 44*(6), 905-912. doi: 10.1111/j.1469-8986.2007.00567.x

Heldmann, M., Rüsseler, J., & Münte, T. F. (2008). Internal and external information in error processing. *Bmc Neuroscience, 9*. doi: 3310.1186/1471-2202-9-33

Hirsh, J. B., & Inzlicht, M. (2008). The Devil You Know: Neuroticism Predicts Neural Response to Uncertainty. *Psychological Science, 19*(10), 962-967. doi: 10.1111/j.1467-9280.2008.02183.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., Hajcak, G., & Larsen, J. T. (2006). The good, the bad and the neutral: Electrophysiological responses to feedback stimuli. *Brain Research, 1105*, 93-101. doi: 10.1016/j.brainres.2005.12.015

Holroyd, C. B., & Krigolson, O. E. (2007). Reward prediction error signals associated with a modified time estimation task. *Psychophysiology, 44*(6), 913-917. doi: 10.1111/j.1469-8986.2007.00561.x

Holroyd, C. B., Nieuwenhuis, S., Yeung, N., & Cohen, J. D. (2003). Errors in reward prediction are reflected in the event-related brain potential. *Neuroreport, 14*(18), 2481-2484. doi: 10.1097/01.wnr.0000099601/41403.a5

Holroyd, C. B., Nieuwenhuis, S., Yeung, N., Nystrom, L., Mars, R. B., Coles, M. G. H., & Cohen, J. D. (2004). Dorsal anterior cingulate cortex shows fMRI response to internal and external error signals. *Nature Neuroscience, 7*(5), 497-498. doi: 10.1038/nn1238

Hoormann, J., Falkenstein, M., Schwarzenau, P., & Hohnsbein, J. (1998). Methods for the quantification and statistical testing of ERP differences across conditions. *Behavior Research Methods Instruments & Computers, 30*(1), 103-109. doi: 10.3758/bf03209420

Hyvarinen, A., & Oja, E. (2000). Independent component analysis: algorithms and applications. [Article]. *Neural Networks, 13*(4-5), 411-430. doi: 10.1016/s0893-6080(00)00026-5

Jia, S. W., Li, H., Luo, Y. J., Chen, A. T., Wang, B. X., & Zhou, X. L. (2007). Detecting perceptual conflict by the feedback-related negativity in brain potentials. *Neuroreport, 18*(13), 1385-1388. doi: 10.1097/WNR.0b013e3282c48a90

Kahneman, D., & Tversky, A. (1979). Prospect Theory – Analysis of decision under risk. *Econometrica, 47*(2), 263-291. doi: 10.2307/1914185

Keil, A., Bradley, M. M., Hauk, O., Rockstroh, B., Elbert, T., & Lang, P. J. (2002). Large-scale neural correlates of affective picture processing. *Psychophysiology, 39*(5), 641-649. doi: 10.1017/s0048577202394162

Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement, 56*(5), 746-759. doi: 10.1177/0013164496056005002

Kobza, S., Thoma, P., Daum, I., & Bellebaum, C. (2011). The feedback-related negativity is modulated by feedback probability in observational learning. *Behavioral Brain Research, 225*(2), 396-404. doi: 10.1016/j.bbr.2011.07.059

Kreussel, L., Hewig, J., Kretschmer, N., Hecht, H., Coles, M. G. H., & Miltner, W. H. R. (2012). The influence of the magnitude, probability, and valence of potential wins and losses on the amplitude of the feedback negativity. *Psychophysiology, 49*(2), 207-219. doi: 10.1111/j.1469-8986.2011.01291.x

Lamm, C., Fischmeister, F. P. S., & Bauer, H. (2005). Individual differences in brain activity during visuo-spatial processing assessed by slow cortical potentials and LORETA. *Cognitive Brain Research, 25*(3), 900-912. doi: 10.1016/j.cogbrainres.2005.09.025

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1999). *International affective picture system (IAPS): Instruction manual and affective ratings. Technical Report A-4.* Gainsville, FL: The Center for Research in Psychophysiology, University of Florida.

Li, P., Han, C., Lei, Y., Holroyd, C. B., & Li, H. (2011). Responsibility modulates neural mechanisms of outcome processing: An ERP study. *Psychophysiology, 48*(8), 1129-1133. doi: 10.1111/j.1469-8986.2011.01182.x

Liao, Y., Gramann, K., Feng, W., Deak, G. O., & Li, H. (2011). This ought to be good: Brain activity accompanying positive and negative expectations and outcomes. *Psychophysiology, 48*(10), 1412-1419. doi: 10.1111/j.1469-8986.2011.01205.x

Lipsey, M. W., & Wilson, D. B. (2001). *Practical Meta-Analysis*. Thousand Oaks, CA: Sage Publications, Inc.

Luck, S. (2005). *An introduction to event-related potential technique.* London, England: MIT Press.

Luu, P., Collins, P., & Tucker, D. M. (2000). Mood, personality, and self-monitoring: Negative affect and emotionality in relation to frontal lobe mechanisms of error monitoring. *Journal of Experimental Psychology-General, 129*(1), 43-60. doi: 10.1037//0096-3445.129.1.43

Miltner, W., Braun, C., & Coles, M. (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*(6), 788-798. doi: 10.1162/jocn.1997.9.6.788

Moser, J. S., Hajcak, G., Holroyd, C., & Simons, R. F. (2004). The effects of feedback magnitude and valence on the P300. *Psychophysiology, 41*, S15-S15.

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

Nieuwenhuis, S., Holroyd, C. B., Mol, N., & Coles, M. G. H. (2004). Reinforcement-related brain potentials from medial frontal cortex: origins and functional significance. *Neuroscience and Biobehavioral Reviews, 28*(4), 441-448. doi: 10.1016/j.neubiorev.2004.05.003

Nieuwenhuis, S., Slagter, H. A., von Geusau, N. J. A., Heslenfeld, D. J., & Holroyd, C. B. (2005). Knowing good from bad: differential activation of human cortical areas by positive and negative outcomes. *European Journal of Neuroscience, 21*(11), 3161-3168. doi: 10.1111/j.1460-9568.2005.04152.x

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*(1), 97-113. doi: 10.1016/0028-3932(71)90067-4

Olvet, D. M., & Hajcak, G. (2008). The error-related negativity (ERN) and psychopathology: Toward an endophenotype. *Clinical Psychology Review, 28*(8), 1343-1354. doi: 10.1016/j.cpr.2008.07.003

Pfabigan, D. M., Alexopoulos, J., Bauer, H., Lamm, C., & Sailer, U. (2011a). All about the money - external performance monitoring is affected by monetary, but not by socially conveyed feedback cues in more antisocial individuals. *Frontiers in Human Neuroscience, 5*. doi: 10010.3389/fnhum.2011.00100

Pfabigan, D. M., Alexopoulos, J., Bauer, H., & Sailer, U. (2011b). Manipulation of feedback expectancy and valence induces negative and positive reward prediction error signals manifest in event-related brain potentials. *Psychophysiology, 48*(5), 656-664. doi: 10.1111/j.1469-8986.2010.01136.x

Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, R., Taylor, M. J. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology, 37*(2), 127-152. doi: 10.1017/s0048577200000305

Polezzi, D., Sartori, G., Rumiati, R., Vidotto, G., & Daum, I. (2010). Brain correlates of risky decision-making. *Neuroimage, 49*(2), 1886-1894. doi: 10.1016/j.neuroimage.2009.08.068

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

Potts, G. F. (2011). Impact of reward and punishment motivation on behavior monitoring as indexed by the error-related negativity. *International Journal of Psychophysiology, 81*(3), 324-331. doi: 10.1016/j.ijpsycho.2011.07.020

Ruchsow, M., Grothe, J., Spitzer, M., & Kiefer, M. (2002). Human anterior cingulate cortex is activated by negative feedback: evidence from event-related potentials in a guessing task. *Neuroscience Letters, 325*(3), 203-206. doi: 10.1016/s0304-3940(02)00288-4

Sailer, U., Fischmeister, F. P. S., & Bauer, H. (2010). Effects of learning on feedback-related brain potentials in a decision-making task. *Brain Research, 1342*, 85-93. doi: 10.1016/j.brainres.2010.04.051

Santesso, D. L., Dzyundzyak, A., & Segalowitz, S. J. (2011). Age, sex and individual differences in punishment sensitivity: Factors influencing the feedback-related negativity. *Psychophysiology, 48*(11), 1481-1488. doi: 10.1111/j.1469-8986.2011.01229.x

Sato, A., Yasuda, A., Ohira, H., Miyawaki, K., Nishikawa, M., Kumano, H., & Kuboki, T. L. (2005). Effects of value and reward magnitude on feedback negativity and P300. *Neuroreport, 16*(4), 407-411. doi: 10.1097/00001756-200503150-00020

Seifert, J. (2005). *Ereigniskorrelierte EEG-Aktivität.* Lengerich, Deutschland: Pabst Science Publishers.

Shackman, A. J., Salomons, T. V., Slagter, H. A., Fox, A. S., Winter, J. J., & Davidson, R. J. (2011). The integration of negative affect, pain and cognitive control in the cingulate cortex. *Nature Reviews Neuroscience, 12*(3), 154-167. doi: 10.1038/nrn2994

Shigeto, H., Ishiguro, J., & Nittono, H. (2011). Effects of visual stimulus complexity on event-related brain potentials and viewing duration in a free-viewing task. *Neuroscience Letters, 497*(2), 85-89. doi: 10.1016/j.neulet.2011.04.035

Smillie, L. D., Cooper, A. J., & Pickering, A. D. (2011). Individual differences in reward-prediction-error: extraversion and feedback-related negativity. *Social Cognitive and Affective Neuroscience, 6*(5), 646-652. doi: 10.1093/scan/nsq078

Streit, M., Wölwer, W., Brinkmeyer, J., Ihl, R., & Gaebel, W. (2000). Electrophysiological correlates of emotional and structural face processing in humans. *Neuroscience Letters, 278*(1-2), 13-16. doi: 10.1016/s0304-3940(99)00884-8

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potentials correlates of stimulus uncertainity. *Science, 150*(3700), 1187-1188. doi: 10.1126/science.150.3700.1187

Sutton, S., Tueting, P., Hammer, M., & Hakerem, G. (1978). Evoked potentials and feedback. In D. A. Otto (Ed.), *Multidisciplinary perspectives in event-related brain potential research* (pp. 184-188). Washington, DC: Environmental Protection Agency.

Ullsperger, M., & von Cramon, D. Y. (2003). Error monitoring using external feedback: Specific roles of the habenular complex, the reward system, and the cingulate motor area revealed by functional magnetic resonance imaging. *Journal of Neuroscience, 23*(10), 4308-4314.

Van der Helden, J., Boksem, M. A. S., & Blom, J. H. G. (2010). The Importance of Failure: Feedback-Related Negativity Predicts Motor Learning Efficiency. *Cerebral Cortex, 20*(7), 1596-1603. doi: 10.1093/cercor/bhp224

Van der Veen, F. M., Röder, C. H., Mies, G. W., van der Lugt, A., & Smits, M. (2011). Remedial action and feedback processing in a time-estimation task: Evidence for a role of the rostral cingulate zone in behavioral adjustments without learning. *Neuroimage, 54*(1), 447-454. doi: 10.1016/j.neuroimage.2010.08.018

Van Veen, V., Holroyd, C. B., Cohen, J. D., Stenger, V. A., & Carter, C. S. (2004). Errors without conflict: Implications for performance monitoring theories of anterior cingulate cortex. *Brain and Cognition, 56*(2), 267-276. doi: 10.1016/j.bandc.2004.06.007

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), 1063-1070. doi: 10.1037/0022-3514.54.6.1063

Wiswede, D., Münte, T. F., Goschke, T., & Rüsseler, J. (2009a). Modulation of the error-related negativity by induction of short-term negative affect. *Neuropsychologia, 47*(1), 83-90. doi: 10.1016/j.neuropsychologia.2008.08.016

Wiswede, D., Münte, T. F., & Rüsseler, J. (2009b). Negative affect induced by derogatory verbal feedback modulates the neural signature of error detection. *Social Cognitive and Affective Neuroscience, 4*(3), 227-237. doi: 10.1093/scan/nsp015

Wu, Y., & Zhou, X. L. (2009). The P300 and reward valence, magnitude, and expectancy in outcome evaluation. *Brain Research, 1286*, 114-122. doi: 10.1016/j.brainres.2009.06.032

Yamamoto, M., Morita, K., Tomita, Y., Tsuji, K., Kawamura, K., & Maeda, H. (2000). Effect of facial affect stimuli on auditory and visual P300 in healthy subjects. *Kurume Medical Journal, 47*(4), 285-290.

Yeung, N., Holroyd, C. B., & Cohen, J. D. (2005). ERP correlates of feedback and reward processing in the presence and absence of response choice. *Cerebral Cortex, 15*(5), 535-544. doi: 10.1093/cercor/bhh153

Yeung, N., & Sanfey, A. G. (2004). Independent coding of reward magnitude and valence in the human brain. *Journal of Neuroscience, 24*(28), 6258-6264. doi: 10.1523/jneurosci.4537-03.2004

Yu, R., & Zhou, X. (2006). Brain potentials associated with outcome expectation and outcome evaluation. *Neuroreport, 17*(15), 1649-1653. doi: 10.1097/01.wnr.0000236866.39328.1d

APPENDIX

A.1 German instruction of the time estimation task

At the beginning of the experiment:

„In diesem Experiment geht es um Zeitschätzung.
Auf dem Bildschirm sehen Sie gleich einen Fixationspunkt, gefolgt von einem "x".
Ihre Aufgabe ist es, EINE SEKUNDE AB dem
Erscheinen des "x" die Taste 1 zu drücken.
Sobald Sie also das "x" sehen, beginnt die zu schätzende Sekunde.

Wenn Sie die Sekunde richtig geschätzt haben,
erhalten Sie als Rückmeldung ein "+" (oder das Bild eines freundlichen Gesichts).

Wenn Sie die Sekunde falsch geschätzt haben,
erhalten Sie als Rückmeldung ein "-" (oder das Bild eines ärgerlichen Gesichts).

Sind Sie bereit? Falls Sie keine weiteren Fragen haben, drücken Sie bitte eine beliebige Taste.

Es folgen zunächst einige Übungsdurchgänge,
in denen Sie sich mit der Aufgabe vertraut machen können.

Bitte drücken Sie zum Starten der Übungsdurchgänge eine beliebige Taste.“

After the practice trials:

„Die Übungsphase ist nun beendet. Wenn Sie keine weiteren Fragen haben, drücken Sie bitte eine beliebige Taste zum Starten des Experiments. Viel Erfolg!”

At the end of the experiment:

„Vielen Dank für Ihre Teilnahme!“

A.2 Behavioral data of Study 1, 2 and 3

Table A.2: Behavioural data of study 1, 2 and 3

	Study 1	Study 2	Study 3
Ratio received posFB : negFB	49.2% : 50.8%	48.8% : 51.2%	49.3% : 50.7%
Mean Response times in ms (across all conditions)	1060 (± 228)	967 (± 279)	1021 (± 266)
Average time window required for positive feedback (in ms)	[918 ; 1082]	[800 ; 1200]	[905 ; 1095]
Mean trial-to-trial time estimation change (in ms)			
- following Neg.FB (minus-sign)	183 (± 48)	199 (± 95)	203 (± 72)
- following Pos.FB (plus-sign)	135 (± 37)	129 (± 31)	138 (± 43)
- following Neg.FB (angry face)	x	190 (± 80)	207 (± 72)
- following Pos.FB (happy face)	x	130 (± 31)	142 (± 54)
- Main effect of feedback valence (negative vs. positive)	$F_{(1,14)} = 62.250$*** ($\eta^2 = .816$)	$F_{(1,15)} = 20.020$*** ($\eta^2 = .572$)	$F_{(1,16)} = 62.939$*** ($\eta^2 = .797$)
- Main effect of feedback type (signs vs. faces)	x	$F_{(1,15)} = 0.820$ ($\eta^2 = .052$)	$F_{(1,14)} = 0.446$ ($\eta^2 = .027$)
- Interaction effect (*valence x type*)	x	none	none
Percentage of time estimation adjustments towards 1000ms			
- following negative vs. positive FB	77.3% vs. 52.0%	76.7% vs. 51.3%	77.4% vs. 53.5%
Mean frequencies of time estimation adjustments towards 1000ms			
- following Neg.FB (minus-sign)	156.27 (± 6.87)	69.13 (± 21.69)	76.71 (± 4.59)
- following Pos.FB (plus-sign)	101.73 (± 18.32)	45.06 (± 17.35)	52.06 (± 8.20)
- following Neg.FB (angry face)	x	69.75 (± 19.18)	76.47 (± 4.76)
- following Pos.FB (happy face)	x	44.00 (± 15.88)	52.88 (± 8.20)
- Main effect of feedback valence (negative vs. positive)	$F_{(1,14)} = 135.478$*** ($\eta^2 = .906$)	$F_{(1,15)} = 70.387$*** ($\eta^2 = .824$)	$F_{(1,16)} = 100.181$*** ($\eta^2 = .862$)
- Main effect of feedback type (signs vs. faces)	x	$F_{(1,15)} = 0.018$ ($\eta^2 = .001$)	$F_{(1,16)} = 0.062$ ($\eta^2 = .004$)
- Interaction effect (*valence x type*)	x	none	none

Note: * p < .05, ** p < .01, ***p < .001
Abbreviations: Neg.FB =negative feedback; Pos.FB = positive Feedback

A.3 Means and standard deviations of FRN amplitudes

Table A.3: Descriptive statistics (means and standard deviations) of FRN amplitudes (in µV) at electrode Fz and FCz

	Negative FB (minus-sign)	**Positive FB (plus-sign)**	**Negative FB (angry face)**	**Positive FB (happy face)**
STUDY 1				
- *Base-to-Peak*	not computed[1]	not computed[1]		
- *Peak-to-Peak (P2-FRN)*	Fz: 0.47 (± 0.71) FCz: 0.29 (± 0.45)	Fz: 0.75 (± 1.39) FCz: 0.75 (± 1.57)		
- *Peak-to-Peak (P3-FRN)*	Fz: 3.23 (± 4.80) FCz: 3.21 (± 5.00)	Fz: 1.36 (± 2.14) FCz: 1.44 (± 2.24)		
- *Mean Amplitude (Mean, SD)*	Fz: 7.78 (± 5.63) FCz: 10.96 (± 6.50)	Fz: 12.14 (± 6.49) FCz: 15.00 (± 6.74)		
- *Individual Mean Amplitude*	not computed[1]	not computed[1]		
STUDY 2				
- *Base-to-Peak*	Fz: 3.69 (± 3.49) FCz: 6.48 (± 4.24)	Fz: 8.86 (± 4.90) FCz: 10.74 (± 5.53)	Fz: 6.54 (± 1.94) FCz: 8.81 (± 3.17)	Fz: 11.49 (± 5.83) FCz: 14.62 (± 7.41)
- *Peak-to-Peak (P2-FRN)*	Fz: 4.49 (± 3.29) FCz: 3.88 (± 3.08)	Fz: 2.10 (± 1.91) FCz: 1.95 (± 1.95)	Fz: 4.21 (± 3.38) FCz: 4.24 (± 3.53)	Fz: 2.34 (± 2.84) FCz: 2.30 (± 2.82)
- *Peak-to-Peak (P3-FRN)*	Fz: 7.72 (± 5.69) FCz: 7.63 (± 6.01)	Fz: 5.42 (± 4.57) FCz: 5.13 (± 4.81)	Fz: 6.25 (5.05) FCz: 6.30 (± 5.11)	Fz: 3.20 (± 3.45) FCz: 2.82 (± 3.63)
- *Mean Amplitude (Mean, SD)*	Fz: 6.56 (± 6.55) FCz: 9.73 (± 6.69)	Fz: 10.31 (± 8.42) FCz: 12.85 (± 8.58)	Fz: 8.51 (± 4.95) FCz: 10.93 (± 5.23)	Fz: 14.27 (± 7.38) FCz: 17.19 (± 8.00)
- *Individual Mean Amplitude*	Fz: 4.72 (± 3.45) FCz: 6.82 (± 4.42)	Fz: 8.60 (± 5.48) FCz: 10.24 (± 5.60)	Fz: 7.47 (± 1.98) FCz: 8.95 (± 3.95)	Fz: 12.42 (± 6.05) FCz: 14.48 (± 7.88)
STUDY 3				
- *Base-to-Peak*	Fz: 2.28 (± 4.14) FCz: 4.83 (± 3.30)	Fz: 8.97 (± 5.56) FCz: 8.29 (± 6.51)	Fz: 2.88 (± 4.14) FCz: 5.75 (± 3.84)	Fz: 8.61 (± 5.90) FCz: 10.47 (± 4.80)
- *Peak-to-Peak (P2-FRN)*	Fz: 5.88 (± 3.25) FCz: 3.58 (± 2.41)	Fz: 1.93 (± 2.69) FCz: 1.40 (± 2.64)	Fz: 5.16 (± 4.60) FCz: 3.34 (± 2.86)	Fz: 2.79 (± 3.61) FCz: 2.12 (± 3.14)
- *Peak-to-Peak (P3-FRN)*	Fz: 10.62 (± 5.67) FCz: 7.67 (± 4.14)	Fz: 2.50 (± 3.25) FCz: 1.53 (± 2.71)	Fz: 9.30 (± 5.84) FCz: 7.71 (± 4.60)	Fz: 4.34 (± 4.26) FCz: 3.10 (± 3.89)
- *Mean Amplitude (Mean, SD)*	Fz: 5.44 (± 4.44) FCz: 7.98 (± 5.38)	Fz: 9.52 (± 4.60) FCz: 12.39 (± 6.19)	Fz: 5.14 (± 3.69) FCz: 8.22 (± 4.12)	Fz: 10.06 (± 4.79) FCz: 12.91 (± 5.51)
- *Individual Mean Amplitude*	Fz: 3.27 (± 3.78) FCz: 5.87 (± 3.70)	Fz: 9.69 (± 5.27) FCz: 9.16 (± 5.98)	Fz: 3.70 (± 4.23) FCz: 6.80 (± 3.90)	Fz: 9.51 (± 5.62) FCz: 11.08 (± 4.74)

Notes: [1] Insufficient sample size because of missing FRN peaks
For the Base-to-Peak and both Mean Amplitude methods a higher value, whereas for the Peak-to-Peak methods a lower value represents a greater FRN.

A.4 Means and standard deviations of P300 amplitudes

Table A.4: Descriptive statistics (means and standard deviations) of P300 amplitudes (in µV) at electrode Pz

	Negative FB (minus-sign)	Positive FB (plus-sign)	Negative FB (angry face)	Positive FB (happy face)
STUDY 1				
- *Base-to-Peak*	16.50 (± 5.45)	17.32 (± 4.09)		
- *Peak-to-Peak (P3-N2)*	17.65 (± 8.84)	18.76 (± 6.32)		
- *Mean Amplitude (Mean, SD)*	14.41 (± 5.44)	16.14 (± 4.78)		
- *Individual Mean Amplitude*	15.72 (± 5.42)	16.59 (± 4.04)		
STUDY 2				
- *Base-to-Peak*	18.41 (± 6.03)	21.02 (± 7.51)	21.09 (± 7.17)	26.63 (± 7.84)
- *Peak-to-Peak (P3-N2)*	11.61 (± 5.04)	14.62 (± 5.21)	15.30 (± 6.82)	19.36 (± 7.53)
- *Mean Amplitude (Mean, SD)*	15.91 (± 6.01)	18.85 (± 7.49)	17.42 (± 5.30)	24.01 (± 7.66)
- *Individual Mean Amplitude*	17.72 (± 6.00)	20.19 (± 7.36)	20.35 (± 7.18)	25.84 (± 7.75)
STUDY 3				
- *Base-to-Peak*	13.38 (± 6.22)	17.24 (± 7.34)	15.79 (± 5.65)	18.36 (± 6.55)
- *Peak-to-Peak (P3-N2)*	8.82 (± 4.17)	10.57 (± 6.30)	11.50 (± 5.53)	14.39 (± 6.76)
- *Mean Amplitude (Mean, SD)*	11.33 (± 6.66)	12.88 (± 6.16)	13.07 (± 5.71)	15.79 (± 5.49)
- *Individual Mean Amplitude*	12.74 (± 6.18)	15.21 (± 7.10)	14.72 (± 6.03)	17.66 (± 6.23)

Note: A higher value stands for a larger P300 Amplitude.

A.5 Mean and standard deviations of FRN and P300 latencies

Table A.5: Descriptive statistics (means and standard deviations) of FRN latencies (at electrodes Fz and FCz) and P300 latencies (at electrode Pz) in ms

	Negative FB (minus-sign)	**Positive FB (plus-sign)**	**Negative FB (angry face)**	**Positive FB (happy face)**
STUDY 1				
- *FRN*[1]	Fz: 274.40 (± 28.37) FCz: 282.40 (± 22.73)	Fz: 284.80 (± 24.72) FCz: 292.80 (± 27.19)		
- *P300*	Pz: 367.43 (± 38.14)	Pz: 334.86 (± 13.72)		
STUDY 2				
- *FRN*	Fz: 276.80 (± 18.84) FCz: 279.20 (± 23.84)	Fz: 302.80 (± 17.49) FCz: 292.00 (± 22.55)	Fz: 304.00 (± 23.63) FCz: 302.40 (± 22.80)	Fz: 306.00 (± 39.65) FCz: 307.20 (± 41.30)
- *P300*	Pz: 366.22 (± 43.17)	Pz: 385.56 (± 28.85)	Pz: 390.22 (± 75.09)	Pz: 370.67 (± 47.94)
STUDY 3				
- *FRN*	Fz: 278.89 (± 11.04) FCz: 308.00 (± 53.57)	Fz: 316.00 (± 24.87) FCz: 324.00 (± 33.78)	Fz: 300.00 (± 43.08) FCz: 332.00 (± 41.83)	Fz: 301.14 (± 23.41) FCz: 309.00 (± 15.79)
- *P300*	Pz: 396.50 (± 67.11)	Pz: 346.00 (± 37.95)	Pz: 431.75 (± 45.64)	Pz: 367.25 (± 48.58)

[1] greatly reduced sample size due to missing FRN peaks

LEBENSLAUF

Persönliche Daten:

Name: Michael Zeiler

Geburtsdatum: 25.07.1987

Geburtsort: Wien

Kontakt: michael.zeiler@univie.ac.at

Ausbildung:

1993 – 1997	Volksschule Kl. Sperlgasse, 1020 Wien
1997 – 2005	Gymnasium GRG II. Zirkusgasse, 1020 Wien
2005 – 2006	Zivildienst in der Caritas-Einrichtung „Am Himmel“, 1190 Wien
Okt. 2006	Beginn des Diplomstudiums Psychologie an der Universität Wien
Nov. 2008	Abschluss des 1. Abschnitts des Diplomstudiums Psychologie
seit Nov. 2008	Studienassistent am Institut für Psychologische Grundlagenforschung (Arbeitsbereich Methodenlehre)
SS 2010	6-Wochen-Praktikum am Institut für Psychologische Grundlagenforschung (Arbeitsbereich Methodenlehre)
Okt. 2011	Beginn des Diplomstudiums Katholische Fachtheologie an der Universität Wien
Nov. 2011	Abschluss des 2. Abschnitts des Diplomstudiums Psychologie

Kongressbesuche:

Dez. 2010	11th Annual Conference of the International Society of Intelligence Research (ISIR), Virginia, USA
Sept. 2011	11. Arbeitstagung der Fachgruppe Differentielle Psychologie, Persönlichkeitspsychologie und Psychologische Diagnostik, Saarbrücken, Deutschland
Apr. 2012	10. Tagung der Österreichischen Gesellschaft für Psychologie (ÖGP), Graz, Österreich

Kongressbeiträge:

Zeiler, M., Pfabigan, D., & Sailer, U. (2012). Faces vs. Symbols – Impact of feedback type on brain potentials during feedback processing. *10. Tagung der Österreichischen Gesellschaft für Psychologie (ÖGP), 12. – 14.04.2012*, Graz, Österreich

Zeiler, M., & Stieger, S. (2012). Nägelbeißen und Fingernagellänge im Kontext expliziter und impliziter Anteile von Selbstwert, Angst und Aggression. *10. Tagung der Österreichischen Gesellschaft für Psychologie (ÖGP), 12. – 14.04.2012*, Graz, Österreich

Pietschnig, J., Zeiler, M., Nader, I., & Voracek, M. (2011). Reporting-Bias bei Zusammenhängen von IQ und Gehirnvolumen: Eine Meta-Analyse. *DPPD 2011,* 26. - 28.09.2011, Saarbrücken.

Pietschnig, J., Zeiler, M., & Voracek, M. (2011). The bigger they are, the higher they score: Meta-analyzing associations of in vivo brain volume and IQ. *ISSID 2011,* 25. - 28.07.2011, London.

Pietschnig, J., Zeiler, M., & Voracek, M. (2010). Of valid concerns and invalid effects: Meta-analyzing associations of in-vivo brain volume and IQ. *The Eleventh Annual Conference of the International Society of Intelligence Research (ISIR), 09. - 11.12.2010*, Alexandria, VA.