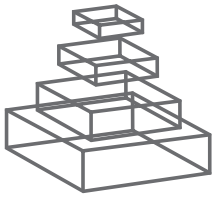
frontiers
RESEARCH TOPICS

FACTORS MEDIATING PERFORMANCE MONITORING IN HUMANS – FROM CONTEXT TO PERSONALITY

Topic Editors
Patrizia Thoma and Christian Bellebaum

frontiers in
HUMAN NEUROSCIENCE



frontiers

Cover image provided by Ibbl sarl, Lausanne CH

ISSN 1664-8714
ISBN 978-2-88919-112-3
DOI 10.3389/978-2-88919-112-3

ABOUT FRONTIERS

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

FRONTIERS JOURNAL SERIES

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing.
All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

DEDICATION TO QUALITY

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.
By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

WHAT ARE FRONTIERS RESEARCH TOPICS?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area!
Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

FACTORS MEDIATING PERFORMANCE MONITORING IN HUMANS – FROM CONTEXT TO PERSONALITY

Topic Editors:
Patrizia Thoma, Ruhr University Bochum, Germany
Christian Bellebaum, Ruhr University Bochum, Germany

In our everyday life, we constantly monitor our behaviour and adapt our responses following performance errors and feedback information from our environment. Receiving positive or negative feedback, which can be social, monetary or some other type of feedback classifiable as good or bad, can encourage us to continue with a specific action or may lead us to discontinue the same behaviour, respectively. Additionally, we daily observe errors being committed by other people or other people receiving feedback for their behaviour. We are able to infer how they feel in response to errors or feedback, and whether we feel sorry for their failures and happy about their successes may depend on our empathic concern and on the relationship to the observed person. At the same time, we can also learn from other people's errors by adaptively modifying our own behaviour.

Recently, a growing number of researchers in the neuroscientific community has begun to establish links between the ability to empathize with others and error/feedback processing. The ACC seems to be strongly involved in both error/feedback processing and in affective empathic responding, and positive relationships between error- and feedback-related ACC activity and self-rated dispositional empathy have been reported. Various contextual factors, like the relationship between the observer and the observed person, or person-related characteristics, like age, gender and psychopathological symptoms, may potentially modify this relationship.

In spite of these theoretical advances, there are still crucial gaps in our knowledge of the different contextual factors and personality characteristics that affect performance monitoring in humans. For instance, it is not well understood how different empathy components might relate to different stages and different forms of error/feedback processing. Also, the ability to engage in empathic perspective taking might be more related to observational than to active learning; and empathy should become more relevant when the behaviour observed in someone else is also relevant for one's own actions. One promising account in studying the relationship between person characteristics, performance context and action monitoring is

the investigation of these concepts across the lifespan. While performance monitoring might be increasingly compromised in older individuals due to structural and functional changes in the relevant brain areas, it might be partly compensated for by a heightened tendency and experience to engage in affective perspective taking. Furthermore, studying clinical populations may help us to disentangle the complex interdependence between performance monitoring and psychopathological symptoms.

Overall, for the current Research Topic issue, we would like to solicit original research articles, reviews as well as opinion and method papers, which investigate the neurocognitive mechanisms supporting performance monitoring providing a link to contextual factors or personality traits. Studies using a range of different methods (behavioural, imaging, electrophysiological, etc.), investigating healthy populations with or without a lifespan perspective or clinical populations are welcome, and authors with different academic backgrounds and working in different disciplines are encouraged to participate in order to promote a lively and integrative debate.

Table of Contents

frontiers in
HUMAN NEUROSCIENCE

EDITORIAL
published: 31 January 2013
doi: 10.3389/fnhum.2013.00023

Factors mediating performance monitoring in humans—from context to personality

Patrizia Thoma* and Christian Bellebaum

Department of Neuropsychology, Institute of Cognitive Neuroscience, Ruhr-University Bochum, Bochum, Germany
**Correspondence: patrizia.thoma@rub.de*

Edited by:
Hauke R. Heekeren, Freie Universität Berlin, Germany

Reviewed by:
Hauke R. Heekeren, Freie Universität Berlin, Germany

In everyday life, we constantly monitor our behavior and adapt our responses following performance errors and feedback information from the environment. Receiving social, monetary or some other type of feedback can encourage us to continue with a specific action or may lead us to discontinue the same behavior. Additionally, we daily observe other people performing various tasks and we can not only learn from their errors or the feedback they receive but also infer how they feel. Whether we feel sorry for their failures and happy about their successes may depend on our empathic concern and on the relationship to the observed person.

The present e-book, which is based on our Frontiers Research Topic entitled "Factors mediating performance monitoring in humans—from context to personality," encompasses both reviews and original research articles which explore the neurocognitive mechanisms supporting performance monitoring providing a link to contextual factors or personality traits.

The overarching theoretical framework for the current Research Topic is presented in three review articles: Thoma and Bellebaum (2012) aimed to link the electrophysiological correlates of performance monitoring, in particular, the mediofrontal negative components error-related negativity (ERN) and feedback-related negativity (FRN), to the concept of empathy. One of the main conclusions they reached is that empathy might be more strongly related to observational than to active learning. This makes sense intuitively given that learning from another person's errors or performance feedback might also involve inferring how the other person feels in response to these events. van Noordt and Segalowitz (2012) adopted a broader perspective on this issue. They reviewed the performance monitoring literature taking into account a variety of potential interindividual differences (such as temperament, different genetic endowments, and various personality factors) as well as different task contexts and linked them to MPFC functioning, as reflected by the mediofrontal negativities. The authors emphasize the highly complex effects of these factors and their interactions on performance monitoring. Brown and Brüne (2012) surveyed the related social neuroscience literature from yet a different angle by focusing on the role of predictive internal representations of one's own and other people's actions, emotions, and outcomes for successful performance monitoring. They postulate that non-social predictive mechanisms, such as prediction error and efference copy signals, also contribute to the processing of social information.

Two original studies in our e-book highlight the importance of the personality dimension discussed in all three previous articles in relation to performance monitoring: Hoffmann et al. (2012) investigated the relationship between the ERN and personality factors, finding a negative association between the ERN and the personality dimensions of "Openness," "Impulsiveness," and "Emotionality" as well as a positive relationship between the ERN and "Social Orientation." The authors conclude that the way people respond to their errors is modulated by their overall emotional and social rigidity. In a comment to this study, Tops and Koole (2012) extended the discussion of the findings arguing that traits related to higher task engagement predict ERN amplitude. Unger et al. (2012), on the other hand, reported a positive association between higher punishment sensitivity and higher FRN amplitudes, independent of feedback validity, which at the same time appeared to be related to poorer behavioral learning performance.

Three further original studies addressed the meaning of contextual factors for performance monitoring: Wu et al. (2011) investigated how recipients in the Ultimatum Game responded when they were not only informed about their own offers but also about the offers of other recipients. The results suggest that, on a neural level, evaluation of fairness in asset division involves an earlier automatic component (mediofrontal negativity) responding to fairness at an abstract level and a later appraisal process (late positive potential) affected by social comparison. Zhang et al. (2012) investigated neural responses to feedback stimuli with a social dimension (female faces). Participants were asked to judge the attractiveness of blurred faces and were shown unblurred faces as feedback. A late FRN-like component showed higher amplitudes in response to feedback faces that were inconsistent with the initial attractiveness judgment than to faces consistent with the judgment. For wave forms in the P300 time window, an opposite effect was found only with more sophisticated data analysis techniques involving a principle component analysis. The authors conclude that complex social feedback stimuli are processed in a similar way as non-social feedback stimuli. Schuermann et al. (2012) investigated how low and high risk for gains and losses affected event-related potentials. FRN amplitudes were enhanced following high-risk decisions but only for gains, while the early positivity (P200) was increased in response to losses following high-risk choices. Finally, P300 amplitudes were increased in high-risk decisions, and in an additive way, following losses compared to gains, suggesting that the P300 may process additional information related to the motivational significance of the processed rewards.

The authors of all three review articles (Brown and Brüne, 2012; Thoma and Bellebaum, 2012; van Noordt and Segalowitz, 2012) advocate the investigation of clinical populations to inform theories about the interactions between context, personality, and performance monitoring. Accordingly, three articles in this e-book involved subclinical or clinical populations: Pfabigan et al. (2011) demonstrated that in comparison with individuals scoring low on anti-social personality traits, individuals with more pronounced antisocial personality traits show enhanced FRN amplitudes to monetary, but not to social feedback. This highlights that these individuals might attribute higher motivation valence to financial assets. Morris et al. (2011) reported that while schizophrenia patients showed diminished ERN amplitudes relative to controls following erroneous responses, groups did not differ on feedback-related activity. Using fMRI, Mainz et al. (2012) investigated the effects of alcohol-related cue exposure on inhibition performance in alcohol-dependent participants. While they did not find any behavioral effects, exposition to alcohol cues was associated with subjectively stronger urges to drink and differential neural activation in amygdala and hippocampus.

Although healthy aging does of course not constitute a pathological condition, it is accompanied by a number of neurobehavioral changes which may alter performance monitoring. Drueke et al. (2012) investigated the effects of performance feedback on executive control, as exerted during a flanker task, in younger and older adults. They found that, although performance feedback improved executive performance in younger individuals, this was not the case in older adults. Error rates, on the other hand, were increased by performance feedback in both groups.

Taken together, we hope that the diverse articles comprised in our e-book may help to illustrate some of the complexities and exiting new developments regarding the intricate relationships between different environmental and personality factors affecting performance monitoring.

REFERENCES

Brown, E. C., and Brüne, M. (2012). The role of prediction in social neuroscience. *Front. Hum. Neurosci.* 6:147. doi: 10.3389/fnhum.2012.00147

Drueke, B., Boecker, M., Mainz, V., Gauggel, S., and Mungard, L. (2012). Can executive control be influenced by performance feedback? Two experimental studies with younger and older adults. *Front. Hum. Neurosci.* 6:90. doi: 10.3389/fnhum.2012.00090

Hoffmann, S., Wascher, E., and Falkenstein, M. (2012). Personality and error monitoring: an update. *Front. Hum. Neurosci.* 6:171. doi: 10.3389/fnhum.2012.00171

Mainz, V., Drueke, B., Boecker, M., Kessel, R., Gauggel, S., and Forkmann, T. (2012). Influence of cue exposure on inhibitory control and brain activation in patients with alcohol dependence. *Front. Hum. Neurosci.* 6:92. doi: 10.3389/fnhum.2012.00092

Morris, S. E., Holroyd, C. B., Mann-Wrobel, M. C., and Gold, J. M. (2011). Dissociation of response and feedback negativity in schizophrenia: electrophysiological and computational evidence for a deficit in the representation of value. *Front. Hum. Neurosci.* 5:123. doi: 10.3389/fnhum.2011.00123

Pfabigan, D. M., Alexopoulos, J., Bauer, H., Lamm, C., and Sailer, U. (2011). All about the money – external performance monitoring is affected by monetary, but not by socially conveyed feedback cues in more antisocial individuals. *Front. Hum. Neurosci.* 5:100. doi: 10.3389/fnhum.2011.00100

Schuermann, B., Endrass, T., and Kathmann, N. (2012). Neural correlates of feedback processing in decision-making under risk. *Front. Hum. Neurosci.* 6:204. doi: 10.3389/fnhum.2012.00204

Thoma, P., and Bellebaum, C. (2012). Your error's got me feeling – how empathy relates to the electrophysiological correlates of performance monitoring. *Front. Hum. Neurosci.* 6:135. doi: 10.3389/fnhum.2012.00135

Tops, M., and Koole, S. L. (2012). An updated update to personality and error monitoring. *Front. Hum. Neurosci.* 6:283. doi: 10.3389/fnhum.2012.00283

Unger, K., Heintz, S., and Kray, J. (2012). Punishment sensitivity modulates the processing of negative feedback but not error-induced learning. *Front. Hum. Neurosci.* 6:186. doi: 10.3389/fnhum.2012.00186

van Noordt, S. J. R., and Segalowitz, S. J. (2012). Performance monitoring and the medial prefrontal cortex: a review of individual differences and context effects as a window on self-regulation. *Front. Hum. Neurosci.* 6:197. doi: 10.3389/fnhum.2012.00197

Wu, Y., Zhou, Y., van Dijk, E., Leliveld, M. C., and Zhou, X. (2011). Social Comparison affects brain responses to fairness in asset division: an ERP study with the ultimatum game. *Front. Hum. Neurosci.* 5:131. doi: 10.3389/fnhum.2011.00131

Zhang, Y., Li, X., Qian, X., and Zhou, X. (2012). Brain responses in evaluating feedback stimuli with a social dimension. *Front. Hum. Neurosci.* 6:29. doi: 10.3389/fnhum.2012.00029

Received: 18 October 2012; accepted: 18 January 2013; published online: 31 January 2013.

Citation: Thoma P and Bellebaum C (2013) Factors mediating performance monitoring in humans—from context to personality. Front. Hum. Neurosci. 7:23. doi: 10.3389/fnhum.2013.00023

frontiers in
HUMAN NEUROSCIENCE

MINI REVIEW ARTICLE
published: 17 May 2012
doi: 10.3389/fnhum.2012.00135

Your error's got me feeling – how empathy relates to the electrophysiological correlates of performance monitoring

Patrizia Thoma* and Christian Bellebaum

Department of Neuropsychology, Institute of Cognitive Neuroscience, Ruhr-University Bochum, Bochum, Germany

Edited by:
Hauke R. Heekeren, Freie Universität Berlin, Germany

Reviewed by:
Grit Hein, University of Zurich, Switzerland
Sid Segalowitz, Brock University, Canada

****Correspondence:***
Patrizia Thoma, Department of Neuropsychology, Institute of Cognitive Neuroscience, Ruhr-University Bochum, Universitätsstraße 150, 44780 Bochum, Germany.
e-mail: patrizia.thoma@rub.de

The error-related and feedback-related negativities (ERN and FRN) represent negative event-related potentials associated with the processing of errors and (negative) response outcomes. The neuronal source of these components is considered to be in the anterior cingulate cortex (ACC). Monitoring one's own behavior and the impact it may have on other people or observing other individuals perform and receive feedback for their actions may also engage empathy-related processes. Empathy is conceived of as a multifaceted construct involving both cognitive and affective components, partly also supported by the ACC. The present mini-review aims to summarize the sparse database linking the electrophysiological correlates of performance monitoring to empathy. While most studies so far provide largely indirect evidence for such an association – e.g., by pointing toward altered ERN/FRN signaling in populations characterized by deviations in empathic responding – fewer investigations establish more explicit links between the two concepts. The relationship between state and, less consistently, trait measures of empathy and action monitoring might be more pronounced for observational than for active participation.

Keywords: error negativity, feedback negativity, empathy, perspective taking, observation learning

INTRODUCTION

The capacity to modify our behavior based on the feedback that we receive for our actions forms an integral part of our everyday life. It enables us to flexibly adapt to distinct environments characterized by different response-outcome contingencies. However, it is not only active learning that allows us to adjust our behavior but also the observation of other individuals being rewarded or punished for their actions. The evaluation of the affective consequences of the outcomes for the observed person and for oneself might involve empathy-related processes. The mirror neuron system, activated during self-performed but also observed actions, is thought to support both observational learning and our ability to resonate with other people's emotions (Gallese, 2003; Gallese et al., 2004). In our mini-review, we will consider empirical evidence for a link between action monitoring and empathy. Given space limitations, regarding the former, we will focus on the error-related negativity (ERN) and the feedback-related negativity (FRN) as electrophysiological correlates of action monitoring in the brain. The ERN and FRN components as well as current empathy concepts will be briefly introduced before we move on to studies linking these concepts.

ACTION MONITORING: THE ERROR- AND FEEDBACK-RELATED NEGATIVITIES

In event-related potential (ERP) studies, characteristic patterns of activity at fronto-central scalp electrodes have been associated with the monitoring of performance. While the response-locked *ERN* represents a negative deflection peaking within 100 ms after error commission (Falkenstein et al., 1991; Gehring et al., 1993), the *FRN* reaches a maximum after 200–300 ms following stimulus onset and is more pronounced for unfavorable as opposed to favorable performance feedback (Gehring and Willoughby, 2002; Nieuwenhuis et al., 2004). The neuronal generator for both components is assumed to be in the anterior cingulate cortex (ACC; Dehaene et al., 1994; Gehring and Willoughby, 2002), a region which has been related to various aspects of cognitive and emotional control [see the reviews by Allman et al., 2001; Van Veen and Carter, 2002; Rushworth et al., 2004; for evidence from functional magnetic resonance imaging (fMRI) studies]. The most common model explaining the functional significance of the ERN and FRN is the reinforcement learning theory (Holroyd and Coles, 2002). Within this framework, errors are conceptualized as "worse than expected outcomes" or negative "prediction errors," leading to an attenuation of phasic dopamine activity in the mesolimbic reward system. Unexpected reward (i.e., a positive prediction error), on the other hand, has been associated with increased phasic dopaminergic signaling. This signal is thought to guide action selection by the ACC, which will either be disinhibited or inhibited, affecting the probability with which an action that has or has not been reinforced in the past will be shown in the future. Alternative theories hold that the ERN reflects the motivational salience attributed to errors and the FRN the motivational and affective evaluation of outcomes (e.g., Gehring and Willoughby, 2002; Yeung et al., 2005), rendering it plausible that empathy might also play a role. This notion is further supported by a hypothesized link between dopaminergic prediction error signaling and context-dependent updating of our representations of other people's emotional states (Abu-Akel and Shamay-Tsoory, 2011).

MONITORING THE EMOTIONAL STATES OF OTHERS: EMPATHY

Empathy broadly refers to the capacity to respond to the emotional experiences of someone else. It is thought of as a multidimensional

construct involving at least a *cognitive component* enabling an individual to understand another person's emotional perspective and an *affective component* based on the ability to affectively share and respond to the emotional experiences of others (Shamay-Tsoory, 2011). Empathic responding is modulated by the context of the interpersonal interaction and characteristics of the observer or the observed person (Hein and Singer, 2008), mediated by executive mechanisms which also keep track of the emotions' source (self vs. other), delineating empathy from pure emotional contagion (Decety and Lamm, 2006). The ventromedial and dorsomedial prefrontal cortices have been associated with cognitive empathy. Affective empathy might partly rely on more simple mechanisms such as emotion recognition, and on shared representations of affective experiences. The inferior frontal gyrus, inferior parietal lobe, anterior insula, and ACC have been linked to affective empathy, with the latter two structures playing a pivotal role in the "empathy for pain" network (see review by Shamay-Tsoory, 2011), although some argue that the anterior insula is more important than the ACC (Gu et al., 2010b). Electrophysiological evidence has repeatedly related state and trait empathic responding to enhanced mu/alpha suppression (e.g., Yang et al., 2009; Perry et al., 2010; Woodruff et al., 2011) and to a modulation of early fronto-central and late centro-parietal ERP amplitudes, partially affecting time windows, in which the ERN/FRN typically occur (Decety et al., 2010; Li and Han, 2010). Although most probably working in concert in most everyday situations, cognitive and affective empathy components can be impaired independently, e.g., in psychiatric disorders like autism (Dziobek et al., 2008), alcoholism (Maurage et al., 2011) and borderline personality disorder (Harari et al., 2010).

INDIRECT EVIDENCE FOR A RELATIONSHIP BETWEEN EMPATHY AND THE ERN/FRN

REPRESENTATION OF AVERSIVE EMOTIONAL STATES

Given the prominent role of the ACC in cognitive and emotional control, it is not surprising that this structure has been related to empathy, particularly in response to aversive emotional states such as physical (Singer et al., 2004) and social pain (Eisenberger and Lieberman, 2004; Krach et al., 2011). Evidence from fMRI studies consistently suggests an overlap between the ACC activation during the first-hand experience of pain or other aversive emotions and during the mere observation of someone else experiencing these events with the strength of this overlapping ACC activity correlating positively with self-reported trait empathy (e.g., Singer et al., 2004; Krach et al., 2011). As mentioned previously, the FRN might also reflect the affective evaluation of negative performance feedback (Gehring and Willoughby, 2002; Yeung et al., 2005), potentially evoking aversive emotions. Transient negative affect and negative affect-related personality traits modulate the FRN and ERN. An enhanced FRN to negative but not to positive or neutral feedback has been related to increased state negative affect and anxiety (Gu et al., 2010a; Santesso et al., 2011) and clinical depression (Mies et al., 2011). Even with depression and anxiety being controlled for, the FRN remained increased in patients with remitted depression (Santesso et al., 2008). However, there are also reports of FRN reductions in association with depressive symptoms (Foti and Hajcak, 2009). Similarly, the ERN amplitude seems to be enhanced in participants with obsessive–compulsive disorder (Xiao et al., 2011), generalized anxiety disorder (Weinberg et al., 2010), and remitted clinical depression (Georgiadi et al., 2011), but reduced during severe depressive episodes (Ruchsow et al., 2004, 2006) with impaired differentiation between errors and correct responses (Olvet et al., 2010). The relationship between the ERN and negative affect seems to be further modulated by factors like psychomotor retardation (Schrijvers et al., 2008), perfectionism (Schrijvers et al., 2010), and neuroticism (Olvet and Hajcak, 2011). Healthy individuals learning better from negative than positive feedback also show increased ERN and FRN signaling (Frank et al., 2005). Overall, enhancement of these components in association with negative affect might point toward a hypervigilant ACC action monitoring system. Interestingly, individuals with clinical depression appear to show increased self-reported trait affective empathy (O'Connor et al., 2002; Thoma et al., 2011), indirectly highlighting an association between a hypervigilant action monitoring system, as indexed by the ERN/FRN, on the one hand and enhanced affective empathic responding on the other.

ALTERED ACTION MONITORING IN POPULATIONS EXHIBITING ABNORMAL EMPATHIC RESPONDING

The electrophysiological correlates of action monitoring are also altered in other populations typically exhibiting abnormal empathic responding. Diminished ERN amplitudes have been reported in individuals with autism spectrum disorder (Vlamings et al., 2008; Sokhadze et al., 2010; South et al., 2010), a population characterized by below average empathy (Baron-Cohen, 2010), possibly particularly regarding cognitive empathy and less so in terms of impaired affective empathy (Dziobek et al., 2008). Reduced ACC activity has been associated with attenuated ERN amplitudes, more severe social impairment and more pronounced psychopathology in adults and children with autism (Henderson et al., 2006; Santesso et al., 2010). On the other hand, the FRN was comparable in individuals with autism and controls, suggesting that the patients might primarily have difficulty with internal, more abstract regulation of performance and less so with feedback processing (Larson et al., 2011). Compared with autism, psychopathy has been associated with the reverse pattern of relatively intact or even superior cognitive and diminished affective empathy (Blair, 2008). While some authors have found reduced ERN, but intact FRN amplitudes (von Borries et al., 2010), others did not find any ERN changes (Brazil et al., 2009) in incarcerated, violent offenders with psychopathy. As these individuals are frequently involved in physical fights, potential previous head injury may confound interpretation of results. Munro et al. (2007a,b) controlled for this and reported reduced ERN and N2 amplitudes following errors in a flanker task with emotional faces but not with neutral letter stimuli, which illustrates an interaction of personality and context on error-related brain activation. Interestingly, compared with healthy controls, offenders with psychopathy performing a social flanker task showed similar ERN amplitudes during active performance, and diminished amplitudes when observing the performance of another individual (Brazil et al., 2011). This suggests a relatively specific impairment of other-related performance monitoring and possibly lower concern about other people's actions in this population.

ACTION MONITORING AND EMPATHY IN OBSERVATIONAL LEARNING

The findings by Brazil et al. (2011) support the relationship between action monitoring and empathy playing a pivotal role in observation situations. The "observational ERN" or oERN reflects similar underlying neural mechanisms as the ERN elicited by active learning, although the peak of the former component seems to occur later and with an attenuated amplitude (van Schie et al., 2004). Similarly, the observational FRN (oFRN) is somewhat reduced in magnitude relative to the active FRN (Bellebaum et al., 2010). fMRI studies have confirmed that overlapping networks encompassing the dorsal ACC, the orbitofrontal cortex, the posterior medial frontal cortex, and supplementary motor regions mirror responses to one's own and to other people's errors (Shane et al., 2008; Brazil et al., 2011). This resembles evidence of overlapping ACC activations for one's own emotional experiences and during the observation of similar emotions in others, as cited above. Witnessing another individual's actions, the observer may rely on cognitive and affective empathy to infer how the other person might feel about her outcomes and what these might entail for one's own performance and outcomes. To date, few studies investigated these associations in observational learning, either indirectly or directly.

Based on the reasoning that empathic responding and the associated neural representations of other people's emotional states might be more pronounced toward individuals we feel emotionally closest to (e.g., Singer et al., 2004), a modulation of the ERN/FRN by the relationship between performer and observer might partly reflect empathy-related processes. While larger perceived similarity between observer and performer has been associated with a decreased oERN when observing confederates perform a flanker task (Carp et al., 2009), a more pronounced oFRN has been reported for participants observing friends vs. strangers complete a Stroop task, with the effect being mediated by the degree to which participants included the observed person in their self-concept (Kang et al., 2010). The fact that the participants' real-life friends were involved might have increased the probability of empathic reactions modulating the oFRN, while in the former study, larger perceived similarity with strangers might not have sufficed to do so. Decreased oERN amplitudes might even mirror the tendency to underestimate error commission by similar others.

According to Marco-Pallares et al. (2010), two different processes may affect the neural signal corresponding to the processing of observed response feedback: one might evaluate the consequences for oneself, while an empathy-related process might evaluate the outcome for the observed person. Depending on the social context, one or the other process might prevail and both may be modulated by different factors. In the betting task these authors used, a "neutral" observer group merely observed a performer's action; for a "parallel" group, losses or wins of the performer entailed similar outcomes for the observer, and in a "reverse" group, losses and wins of the performer signaled reverse outcomes for the observer. Participants showed a pronounced FRN to losses vs. gains, both as active players and as "neutral" or "parallel" observers. In the "reverse" group, however, an oFRN was elicited only in response to *wins* of the performer corresponding to losses for the observer. Similarly, active participation in a task may elicit competitive feelings, highlighting the need to evaluate outcome-related consequences for oneself and attenuating empathic responding toward the observed competitor. Accordingly, Ma et al. (2011) reported that an increased oFRN to a friend's relative to a stranger's performance could only be observed if the observer was not actively involved in the game.

STUDIES ASSESSING BOTH EMPATHY AND THE ERN/FRN

In contrast to the studies reported in the previous paragraphs, some authors used self-report measures of state or trait empathy allowing for a more direct investigation of the relationship between empathy and the ERN/FRN, although overall the result pattern does not appear consistent as yet.

Complementing the Ma et al. (2011) findings, Koban et al. (2010) showed that when participants' attentional resources were taken up by focusing on their own actions, the ERN during active learning was unaffected by a cooperative vs. competitive social context. In an observation condition, participants showed an "early" oERN after 125–145 ms during cooperation and a "late" oERN (280–320 ms) during competition. Trait empathy was unrelated to any of these components, but state measures of rivalry and competition toward the observed player were associated with a diminished early oERN, while the late oERN was smaller for participants who felt more sympathy and friendship toward the co-player. Having their participants play a competitive card game, Yamada et al. (2011) found larger FRN amplitudes on trials signaling "gain" for the participant and simultaneous "loss" for the confederate player (incongruent condition) relative to trials where both opponents lost (congruent condition), interpreting this as an effect of "counterempathy" or "schadenfreude." Larger FRN differences (incongruent–congruent loss) were related to higher subjective ratings of pleasantness about one's own winnings, but not to trait empathy. Only male participants were investigated, and there is evidence that gender may modulate empathy-related ACC activation (Singer et al., 2006) and the neural correlates of action monitoring in competitive situations. In a gambling task, where one player's monetary gain resulted in the opponent's loss, perception of the opponent's negative outcome elicited a small but discernible oFRN (loss–gain) in female, but not in male participants, even if the other individual's loss incurred wins for them (Fukushima and Hiraki, 2006). The authors attributed this to a more pronounced tendency of women to feel empathy for their opponents. Overall, the more the participants felt empathic concern about the opponent's outcomes, the less the oFRN diminished. Habitual tendencies to empathize and systemize (i.e., to focus on the analysis of physical objects and systems; Baron-Cohen et al., 2003) were also assessed. A higher "empathizing minus systemizing" score was negatively related to the amplitude of the oFRN, but not of the active FRN. The authors concluded that individual differences in empathy-related neural activity are best illustrated as a ratio between empathetic and non-empathetic (systemizing) functions.

Further support for the notion that the oFRN might be modulated by empathic responding specifically characterizing human interactions comes from a later study by Fukushima and Hiraki (2009). Participants performed actively, but also observed the performance of real-life friends or computer players, with the

outcomes of the players being unrelated to each other. A significant oFRN was elicited only when humans were observed. Larger oFRN amplitudes were associated with higher dispositional cognitive and affective empathy, while there were no such relationships in the "computer player" condition. During active performance, higher cognitive empathy and smaller FRN amplitudes were marginally significantly correlated. This illustrates that while empathy might positively affect the monitoring of other people's actions, the tendency to habitually focus on other people's emotions might actually disrupt the monitoring of one's own performance. Depending on stimulus-feedback contingencies, this might also apply to observational learning. Kobza et al. (2011) had participants observe virtual others receiving positive or negative feedback for choosing between two symbols. The probability of positive feedback varied for different stimuli. Higher trait affective empathy was associated with poorer performance and higher trait cognitive empathy with smaller oFRN differences (positive–negative feedback) only when contingencies were most difficult to learn and feedback difficult to predict. The authors suggest that highly empathic individuals might tend to focus on the observed person's choice behavior rather than on response feedback contingencies to make sense of unpredictable feedback. Together with cognitive resources being taken up by attending to the emotional consequences for the observed person, this might particularly disrupt the learning of difficult associations. An alternative interpretation was suggested in an fMRI investigation by Newman-Norlund et al. (2009) who reported an association between higher trait empathic concern and weaker ventral ACC activation following error observation. According to the authors, empathic concern may also represent a disposition to regulate negative affect elicited by the observation of other people committing errors that might be relevant for oneself (in this case missed penalty shots of a soccer club one does or does not support). In cases where the observed errors might lack significance for the observer, empathic concern and ventral ACC activity seem to be associated positively (Shane et al., 2009).

While in the ERP studies presented earlier, ERN/FRN amplitudes correlated more consistently with state rather than with trait measures of empathy-related affective responding, FRN differences were associated with trait empathy in the two latter studies. Based on the evidence presented so far, it is difficult to decide which factors contributed to the inconsistencies. Note that at least in the Kobza et al. (2011) study, participants actually had to transfer the knowledge they acquired by observation to their own performance assessed in subsequent active test trials. This might have induced highly empathic individuals to try to benefit from the other person's coping with an ambiguous feedback situation to guide their own actions. Other studies focusing on active learning yielded significant correlations of the ERN with trait empathy. Santesso and Segalowitz (2009) reported significant associations between increased risk taking propensity and diminished ERN amplitudes and between higher trait empathy and larger ERN amplitudes in adolescents performing a flanker task. While high risk takers might not care about their errors and/or show diminished ability to learn from negative feedback, highly empathic individuals might be implicitly more concerned about the impact their actions might have on others. As empathy and risk taking were not correlated, they account for separate variance in the ERN. Larson et al. (2010) confirmed the association between larger ERN amplitudes and higher dispositional empathy, controlling for state negative affect. According to these authors, one construct that may explain the relationship between ERN and empathy might be vigilance to one's own performance and to the environment. Alternatively, both empathy and the ERN might be related to caring about positive or negative outcomes of one's own or other people's behavior.

CONCLUSION

Taken together, the evidence available so far points toward an association between empathy-related affective responding and the ERN/FRN components as electrophysiological correlates of action monitoring. The exact nature of the relationship is subject to modulation depending on state (negative) affect, personality and psychopathology, the type of learning (active or observational), gender, the specific stimulus-feedback contingencies and the interactive context (competitive vs. cooperative, relationship between performer and observer) among other factors. Currently, few studies explicitly assessed self-reported state or trait empathy in association with the ERN/FRN. Although part of the evidence nicely supports the association between ACC, empathy and ERN/FRN, some inconsistencies have to be borne in mind. For instance, while in fMRI studies, it is rather the rostral ACC that has been associated with error monitoring (Van Veen and Carter, 2002), mid-ACC (Lamm et al., 2011), or anterior insular activity (Gu et al., 2010b) seems to be more strongly related to empathy. On the other hand, trait empathy was found to relate to ventral ACC activity elicited following error observation, which possibly reflects affective aspects of error processing (Newman-Norlund et al., 2009). Future studies should try to more precisely disentangle the nature of the relationship between distinct empathy components and the ERN/FRN.

ACKNOWLEDGMENTS

Christian Bellebaum was supported by the Ministry of Innovation, Science and Research of the federal state of Nordrhein-Westfalen, Germany (Ministeriumfür Innovation, Wissenschaft und Forschung des Landes Nordrhein-Westfalen; MIWF – grant number 334-4). This work was also supported by the German Research Foundation (Deutsche Forschungsgemeinschaft – DFG), grant number TH 1535/2-1, awarded to Patrizia Thoma.

REFERENCES

Abu-Akel, A., and Shamay-Tsoory, S. (2011). Neuroanatomical and neurochemical bases of theory of mind. *Neuropsychologia* 49, 2971–2984.

Allman, J. M., Hakeem, A., Erwin, J. M., Nimchinsky, E., and Hof, P. (2001). The anterior cingulate cortex. The evolution of an interface between emotion and cognition. *Ann. N. Y. Acad. Sci.* 935, 107–117.

Baron-Cohen, S. (2010). Empathizing, systemizing, and the extreme male brain theory of autism. *Prog. Brain Res.* 186, 167–175.

Baron-Cohen, S., Richler, J., Bisarya, D., Gurunathan, N., and Wheelwright, S. (2003). The systemizing quotient: an investigation of adults with Asperger syndrome or high-functioning autism, and normal sex differences. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358, 361–374.

Bellebaum, C., Kobza, S., Thiele, S., and Daum, I. (2010). It was not MY fault: event-related brain potentials in active and observational learning from feedback. *Cereb. Cortex* 20, 2874–2883.

Blair, R. J. (2008). Fine cuts of empathy and the amygdala: dissociable deficits in psychopathy and autism. *Q. J. Exp. Psychol. (Colchester)* 61, 157–170.

Brazil, I. A., De Bruijn, E. R., Bulten, B. H., von Borries, A. K., van Lankveld, J. J., Buitelaar, J. K., and Verkes, R. J. (2009). Early and late components of error monitoring in violent offenders with psychopathy. *Biol. Psychiatry* 65, 137–143.

Brazil, I. A., Mars, R. B., Bulten, B. H., Buitelaar, J. K., Verkes, R. J., and De Bruijn, E. R. (2011). A neurophysiological dissociation between monitoring one's own and others' actions in psychopathy. *Biol. Psychiatry* 69, 693–699.

Carp, J., Halenar, M. J., Quandt, L. C., Sklar, A., and Compton, R. J. (2009). Perceived similarity and neural mirroring: evidence from vicarious error processing. *Soc. Neurosci.* 4, 85–96.

Decety, J., and Lamm, C. (2006). Human empathy through the lens of social neuroscience. *ScientificWorldJournal* 6, 1146–1163.

Decety, J., Yang, C. Y., and Cheng, Y. (2010). Physicians down-regulate their pain empathy response: an event-related brain potential study. *Neuroimage* 50, 1676–1682.

Dehaene, S., Posner, M. I., and Tucker, D. M. (1994). Localization of an neural system for error detection and compensation. *Psychol. Sci.* 5, 303–305.

Dziobek, I., Rogers, K., Fleck, S., Bahnemann, M., Heekeren, H. R., Wolf, O. T., and Convit, A. (2008). Dissociation of cognitive and emotional empathy in adults with Asperger syndrome using the Multifaceted Empathy Test (MET). *J. Autism Dev. Disord.* 38, 464–473.

Eisenberger, N. I., and Lieberman, M. D. (2004). Why rejection hurts: a common neural alarm system for physical and social pain. *Trends Cogn. Sci. (Regul. Ed.)* 8, 294–300.

Falkenstein, M., Hohnsbein, J., Hoormann, J., and Blanke, L. (1991). Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. *Electroencephalogr. Clin. Neurophysiol.* 78, 447–455.

Foti, D., and Hajcak, G. (2009). Depression and reduced sensitivity to non-rewards versus rewards: evidence from event-related potentials. *Biol. Psychol.* 81, 1–8.

Frank, M. J., Woroch, B. S., and Curran, T. (2005). Error-related negativity predicts reinforcement learning and conflict biases. *Neuron* 47, 495–501.

Fukushima, H., and Hiraki, K. (2006). Perceiving an opponent's loss: gender-related differences in the medial-frontal negativity. *Soc. Cogn. Affect. Neurosci.* 1, 149–157.

Fukushima, H., and Hiraki, K. (2009). Whose loss is it? Human electrophysiological correlates of non-self reward processing. *Soc. Neurosci.* 4, 261–275.

Gallese, V. (2003). The manifold nature of interpersonal relations: the quest for a common mechanism. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358, 517–528.

Gallese, V., Keysers, C., and Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends Cogn. Sci. (Regul. Ed.)* 8, 396–403.

Gehring, W. J., Goss, G. M. G. H. G., Meyer, D. E., and Donchin, E. (1993). A neural system for error detection and compensation. *Psychol. Sci.* 4, 385–390.

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

Georgiadi, E., Liotti, M., Nixon, N. L., and Liddle, P. F. (2011). Electrophysiological evidence for abnormal error monitoring in recurrent major depressive disorder. *Psychophysiology* 48, 1192–1202.

Gu, R., Ge, Y., Jiang, Y., and Luo, Y. J. (2010a). Anxiety and outcome evaluation: the good, the bad and the ambiguous. *Biol. Psychol.* 85, 200–206.

Gu, X., Liu, X., Guise, K. G., Naidich, T. P., Hof, P. R., and Fan, J. (2010b). Functional dissociation of the frontoinsular and anterior cingulate cortices in empathy for pain. *J. Neurosci.* 30, 3739–3744.

Harari, H., Shamay-Tsoory, S. G., Ravid, M., and Levkovitz, Y. (2010). Double dissociation between cognitive and affective empathy in borderline personality disorder. *Psychiatry Res.* 175, 277–279.

Hein, G., and Singer, T. (2008). I feel how you feel but not always: the empathic brain and its modulation. *Curr. Opin. Neurobiol.* 18, 153–158.

Henderson, H., Schwartz, C., Mundy, P., Burnette, C., Sutton, S., Zahka, N., and Pradella, A. (2006). Response monitoring, the error-related negativity, and differences in social behavior in autism. *Brain Cogn.* 61, 96–109.

Holroyd, C. B., and Coles, M. G. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychol. Rev.* 109, 679–709.

Kang, S. K., Hirsh, J. B., and Chasteen, A. L. (2010). Your mistakes are mine: self-other overlap predicts neural response to observed errors. *J. Exp. Soc. Psychol.* 46, 229–232.

Koban, L., Pourtois, G., Vocat, R., and Vuilleumier, P. (2010). When your errors make me lose or win: event-related potentials to observed errors of cooperators and competitors. *Soc. Neurosci.* 5, 360–374.

Kobza, S., Thoma, P., Daum, I., and Bellebaum, C. (2011). The feedback-related negativity is modulated by feedback probability in observational learning. *Behav. Brain Res.* 225, 396–404.

Krach, S., Cohrs, J. C., Echeverria Loebell, N. C., Kircher, T., Sommer, J., Jansen, A., and Paulus, F. M. (2011). Your flaws are my pain: linking empathy to vicarious embarrassment. *PLoS ONE* 6, e18675. doi:10.1371/journal.pone.0018675

Lamm, C., Decety, J., and Singer, T. (2011). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *Neuroimage* 54, 2492–2502.

Larson, M. J., Fair, J. E., Good, D. A., and Baldwin, S. A. (2010). Empathy and error processing. *Psychophysiology* 47, 415–424.

Larson, M. J., South, M., Krauskopf, E., Clawson, A., and Crowley, M. J. (2011). Feedback and reward processing in high-functioning autism. *Psychiatry Res.* 187, 198–203.

Li, W., and Han, S. (2010). Perspective taking modulates event-related potentials to perceived pain. *Neurosci. Lett.* 469, 328–332.

Ma, Q., Shen, Q., Xu, Q., Li, D., Shu, L., and Weber, B. (2011). Empathic responses to others' gains and losses: an electrophysiological investigation. *Neuroimage* 54, 2472–2480.

Marco-Pallares, J., Kramer, U. M., Strehl, S., Schroder, A., and Munte, T. F. (2010). When decisions of others matter to me: an electrophysiological analysis. *BMC Neurosci.* 11, 86. doi:10.1186/1471-2202-11-86

Maurage, P., Grynberg, D., Noel, X., Joassin, F., Philippot, P., Hanak, C., Verbanck, P., Luminet, O., de Timary, P., and Campanella, S. (2011). Dissociation between affective and cognitive empathy in alcoholism: a specific deficit for the emotional dimension. *Alcohol. Clin. Exp. Res.* 35, 1662–1668.

Mies, G. W., van der Veen, F. M., Tulen, J. H., Birkenhager, T. K., Hengeveld, M. W., and van der Molen, M. W. (2011). Drug-free patients with major depression show an increased electrophysiological response to valid and invalid feedback. *Psychol. Med.* 41, 2515–2525.

Munro, G. E., Dywan, J., Harris, G. T., McKee, S., Unsal, A., and Segalowitz, S. J. (2007a). ERN varies with degree of psychopathy in an emotion discrimination task. *Biol. Psychol.* 76, 31–42.

Munro, G. E., Dywan, J., Harris, G. T., McKee, S., Unsal, A., and Segalowitz, S. J. (2007b). Response inhibition in psychopathy: the frontal N2 and P3. *Neurosci. Lett.* 418, 149–153.

Newman-Norlund, R. D., Ganesh, S., van Schie, H. T., De Bruijn, E. R., and Bekkering, H. (2009). Self-identification and empathy modulate error-related brain activity during the observation of penalty shots between friend and foe. *Soc. Cogn. Affect. Neurosci.* 4, 10–22.

Nieuwenhuis, S., Holroyd, C. B., Mol, N., and Coles, M. G. (2004). Reinforcement-related brain potentials from medial frontal cortex: origins and functional significance. *Neurosci. Biobehav. Rev.* 28, 441–448.

O'Connor, L. E., Berry, J. W., Weiss, J., and Gilbert, P. (2002). Guilt, fear, submission, and empathy in depression. *J. Affect. Disord.* 71, 19–27.

Olvet, D. M., and Hajcak, G. (2011). The error-related negativity relates to sadness following mood induction among individuals with high neuroticism. *Soc. Cogn. Affect. Neurosci.* 7, 289–295.

Olvet, D. M., Klein, D. N., and Hajcak, G. (2010). Depression symptom severity and error-related brain activity. *Psychiatry Res.* 179, 30–37.

Perry, A., Bentin, S., Bartal, I. B., Lamm, C., and Decety, J. (2010). "Feeling" the pain of those who are different from us: modulation of EEG in the mu/alpha range. *Cogn. Affect. Behav. Neurosci.* 10, 493–504.

Ruchsow, M., Herrnberger, B., Beschoner, P., Gron, G., Spitzer, M., and Kiefer, M. (2006). Error processing in major depressive disorder: evidence from event-related potentials. *J. Psychiatr. Res.* 40, 37–46.

Ruchsow, M., Herrnberger, B., Wiesend, C., Gron, G., Spitzer, M., and Kiefer,

M. (2004). The effect of erroneous responses on response monitoring in patients with major depressive disorder: a study with event-related potentials. *Psychophysiology* 41, 833–840.

Rushworth, M. F., Walton, M. E., Kennerley, S. W., and Bannerman, D. M. (2004). Action sets and decisions in the medial frontal cortex. *Trends Cogn. Sci. (Regul. Ed.)* 8, 410–417.

Santesso, D. L., Bogdan, R., Birk, J. L., Goetz, E. L., Holmes, A. J., and Pizzagalli, D. A. (2011). Neural responses to negative feedback are related to negative emotionality in healthy adults. *Soc. Cogn. Affect. Neurosci.* doi:10.1093/scan/nsr054. [E-pub ahead of print].

Santesso, D. L., Drmic, I. E., Jetha, M. K., Bryson, S. E., Goldberg, J. O., Hall, G. B., Mathewson, K. J., Segalowitz, S. J., and Schmidt, L. A. (2010). An event-related source localization study of response monitoring and social impairments in autism spectrum disorder. *Psychophysiology* 48, 241–251.

Santesso, D. L., and Segalowitz, S. J. (2009). The error-related negativity is related to risk taking and empathy in young men. *Psychophysiology* 46, 143–152.

Santesso, D. L., Steele, K. T., Bogdan, R., Holmes, A. J., Deveney, C. M., Meites, T. M., and Pizzagalli, D. A. (2008). Enhanced negative feedback responses in remitted depression. *Neuroreport* 19, 1045–1048.

Schrijvers, D., De Bruijn, E. R., Maas, Y., De Grave, C., Sabbe, B. G., and Hulstijn, W. (2008). Action monitoring in major depressive disorder with psychomotor retardation. *Cortex* 44, 569–579.

Schrijvers, D. L., De Bruijn, E. R., Destoop, M., Hulstijn, W., and Sabbe, B. G. (2010). The impact of perfectionism and anxiety traits on action monitoring in major depressive disorder. *J. Neural Transm.* 117, 869–880.

Shamay-Tsoory, S. G. (2011). The neural bases for empathy. *Neuroscientist* 17, 18–24.

Shane, M. S., Stevens, M., Harenski, C. L., and Kiehl, K. A. (2008). Neural correlates of the processing of another's mistakes: a possible underpinning for social and observational learning. *Neuroimage* 42, 450–459.

Shane, M. S., Stevens, M. C., Harenski, C. L., and Kiehl, K. A. (2009). Double dissociation between perspective-taking and empathic-concern as predictors of hemodynamic response to another's mistakes. *Soc. Cogn. Affect. Neurosci.* 4, 111–118.

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., and Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science* 303, 1157–1162.

Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., and Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature* 439, 466–469.

Sokhadze, E., Baruth, J., El Baz, A., Horrell, T., Sokhadze, G., Carroll, T., Tasman, A., Sears, L., and Casanova, M. F. (2010). Impaired error monitoring and correction function in autism. *J. Neurother.* 14, 79–95.

South, M., Larson, M. J., Krauskopf, E., and Clawson, A. (2010). Error processing in high-functioning autism spectrum disorders. *Biol. Psychol.* 85, 242–251.

Thoma, P., Zalewski, I., von Reventlow, H. G., Norra, C., Juckel, G., and Daum, I. (2011). Cognitive and affective empathy in depression linked to executive control. *Psychiatry Res.* 189, 373–378.

Van Veen, V., and Carter, C. S. (2002). The anterior cingulate as a conflict monitor: fMRI and ERP studies. *Physiol. Behav.* 77, 477–482.

van Schie, H. T., Mars, R. B., Coles, M. G., and Bekkering, H. (2004). Modulation of activity in medial frontal and motor cortices during error observation. *Nat. Neurosci.* 7, 549–554.

Vlamings, P. H., Jonkman, L. M., Hoeksma, M. R., van Engeland, H., and Kemner, C. (2008). Reduced error monitoring in children with autism spectrum disorder: an ERP study. *Eur. J. Neurosci.* 28, 399–406.

von Borries, A. K., Brazil, I. A., Bulten, B. H., Buitelaar, J. K., Verkes, R. J., and De Bruijn, E. R. (2010). Neural correlates of error-related learning deficits in individuals with psychopathy. *Psychol. Med.* 40, 1559–1568.

Weinberg, A., Olvet, D. M., and Hajcak, G. (2010). Increased error-related brain activity in generalized anxiety disorder. *Biol. Psychol.* 85, 472–480.

Woodruff, C. C., Martin, T., and Bilyk, N. (2011). Differences in self- and other-induced Mu suppression are correlated with empathic abilities. *Brain Res.* 1405, 69–76.

Xiao, Z., Wang, J., Zhang, M., Li, H., Tang, Y., Wang, Y., Fan, Q., and Fromson, J. A. (2011). Error-related negativity abnormalities in generalized anxiety disorder and obsessive-compulsive disorder. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 35, 265–272.

Yamada, M., Lamm, C., and Decety, J. (2011). Pleasing frowns, disappointing smiles: an ERP investigation of counterempathy. *Emotion* 11, 1336–1345.

Yang, C. Y., Decety, J., Lee, S., Chen, C., and Cheng, Y. (2009). Gender differences in the mu rhythm during empathy for pain: an electroencephalographic study. *Brain Res.* 1251, 176–184.

Yeung, N., Holroyd, C. B., and Cohen, J. D. (2005). ERP correlates of feedback and reward processing in the presence and absence of response choice. *Cereb. Cortex* 15 535–544.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 13 February 2012; accepted: 26 April 2012; published online: 17 May 2012.

Citation: Thoma P and Bellebaum C (2012) Your error's got me feeling – how empathy relates to the electrophysiological correlates of performance monitoring. Front. Hum. Neurosci. ***6****:135. doi: 10.3389/fnhum.2012.00135*

frontiers in
HUMAN NEUROSCIENCE

REVIEW ARTICLE
published: 11 July 2012
doi: 10.3389/fnhum.2012.00197

Performance monitoring and the medial prefrontal cortex: a review of individual differences and context effects as a window on self-regulation

Stefon J. R. van Noordt and Sidney J. Segalowitz*

Cognitive and Affective Neuroscience Laboratory, Department of Psychology, Brock University, ON, Canada

Edited by:
Patrizia Thoma, Ruhr-University Bochum, Germany

Reviewed by:
Maarten A. S. Boksem, Erasmus University Rotterdam, Netherlands
Daniela M. Pfabigan, University of Vienna, Austria

***Correspondence:**
Sidney J. Segalowitz, Cognitive and Affective Neuroscience Laboratory, Department of Psychology, Brock University, 500 Glenridge Avenue, St. Catharines, ON L2S 3A1, Canada.
e-mail: sid.segalowitz@brocku.ca

The medial prefrontal cortex (MPFC) is central to self-regulation and has been implicated in generating a cluster of event-related potential components, collectively referred to as medial frontal negativities (MFNs). These MFNs are elicited while individuals monitor behavioral and environmental consequences, and include the error-related negativity, Nogo N2, and the feedback-related negativity. A growing cognitive and affective neuroscience literature indicates that the activation of the anterior cingulate cortex (ACC) and surrounding medial prefrontal regions during performance monitoring is not only influenced by task context, but that these patterns of activity also vary as a function of individual differences (e.g., personality, temperament, clinical and non-clinical symptomatology, socio-political orientation, and genetic polymorphisms), as well as interactions between individual differences and task context. In this review we survey the neuroscience literature on the relations between performance monitoring, personality, task context, and brain functioning with a focus on the MPFC. We relate these issues to the role of affect in the paradigms used to elicit performance-monitoring neural responses and highlight some of the theoretical and clinical implications of this research. We conclude with a discussion of the complexity of these issues and how some of the basic assumptions required for their interpretation may be clarified with future research.

Keywords: ERN, FRN, Nogo N2, anterior cingulate, medial prefrontal cortex, individual differences, performance monitoring, self-regulation

A hallmark of self-regulation is flexibility—the ability to maintain or disengage and establish different patterns of behavior in pursuit of adaptive outcomes (Baumeister et al., 2007). Monitoring, detecting, and evaluating behavioral and environmental consequences require the coordination of activity across multiple neural systems. In the human brain, areas of the prefrontal cortex are involved in mediating cognitive control processes of motor behavior (Ridderinkhof et al., 2004a,b; Polli et al., 2005), as well as appraisal and motivational responses to behavioral and environmental feedback (Ridderinkhof et al., 2004b; Schnider et al., 2005; Diekhof et al., 2011; Etkin et al., 2011). It is well established that activation of the anterior cingulate cortex (ACC) and surrounding medial prefrontal areas is associated with performance monitoring processes such as error detection and response correction, stimulus-response conflict resolution, inhibitory control, and feedback evaluation, all of which involve demands on the selection and maintenance of goal-directed behavior (Holroyd and Yeung, 2012). In addition, accumulating evidence is showing that the activation in the ACC and medial prefrontal cortex (MPFC) during performance monitoring is not only influenced by task context, but that these patterns of activity also vary as a function of individual differences in personality, as well as interactions between personality and task context.[1]

There is considerable interest in trying to understand the associations among personality, context, and brain activation during performance monitoring, as reflected by a growing body of literature in the cognitive, affective, and social neurosciences. Not only does this research have important theoretical implications, but these data can also inform our clinical understanding about how neurophysiological differences may reflect pathological patterns of performance monitoring and self-regulation. In order to synthesize the current understanding in the field, we have surveyed for this paper the neuroscience literature on the relations between performance monitoring, personality, task context, and brain functioning with a focus on the MPFC. The growth in

[1]Researchers vary on whether the MPFC includes all of, portions of, or none of the ACC (Sallet et al., 2011). For the sake of simplicity, we will consider the term MPFC to include the ACC in general, and will be more specific when needed. Several researchers offer a more detailed labelling of the cingulate cortex based on the results from structural (Vogt, 2009), functional (Shackman et al., 2011), connectivity (Beckmann et al., 2009) and receptor density distribution (Palomero-Gallagher et al., 2009) studies.

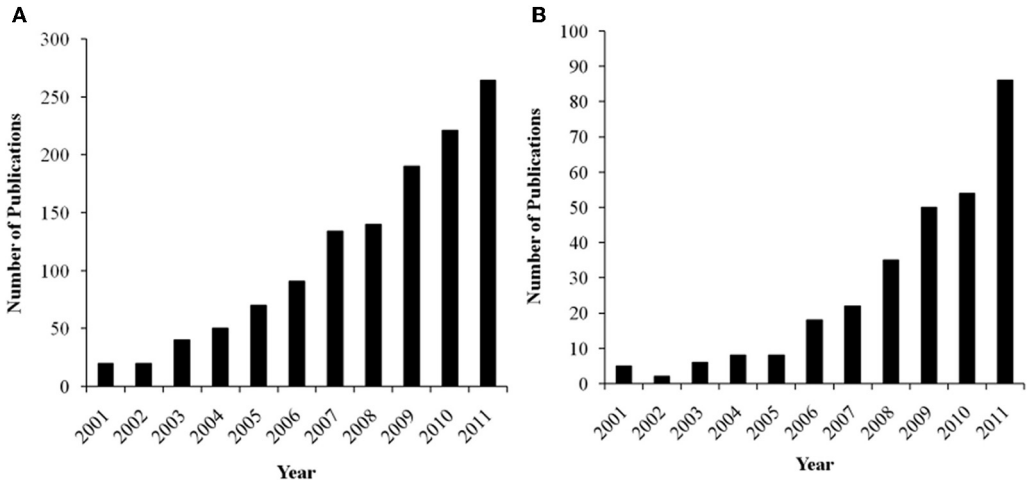


FIGURE 1 | (A) Number of annual publications listed in Google Scholar for key-terms "error-related negativity + personality." **(B)** Number of annual publications listed in Google Scholar for key-terms "feedback-related negativity + personality."

research on the functional relations between MPFC and these issues has been exponential in the last decade, as indicated by a literature search in Google Scholar (see **Figures 1A** and **B**). In this review, we relate these issues especially to the role of affect in the paradigms used to elicit the performance monitoring neural responses. To simplify the terminology, we will consider the DMPFC and VMPFC as a broad division, with each of these areas including several anatomically distinct regions (see **Figure 2**).

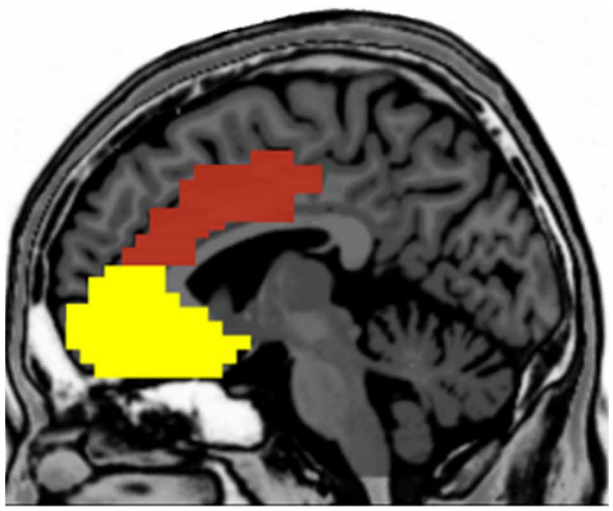


FIGURE 2 | Broad division between the dorsal (red) and ventral (yellow) medial prefrontal cortices.

THE MEDIAL PREFRONTAL CORTEX AND MEDIAL FRONTAL NEGATIVITIES

The MPFC generates several event-related potential (ERP) components associated with performance monitoring and self-regulation. For our focus, these ERP components include the error-related negativity (ERN), the Nogo N200 (N2), and the feedback-related negativity (FRN). Although some researchers have used the term medial frontal negativity (MFN) to describe specifically the FRN (e.g., Gehring and Willoughby, 2002), for this review we use the MFN label when referring to all three components. Each of these MFNs is elicited in a specific context (Luck, 2005) and, as described below, regions of the MPFC are consistently implicated as neuronal generators of all three (Gehring and Willoughby, 2002; van Veen and Carter, 2002a,b; Mathalon et al., 2003; Amodio et al., 2007; Gentsch et al., 2009; Segalowitz et al., 2010). Furthermore, there are theoretical constructs linking the three and therefore, while we will not at all claim that they are identical, there are good reasons to consider the three components together (see **Figure 3**).

ERN

The ERN, or error-negativity (Ne; Falkenstein et al., 1991), was first identified in the early 1990s and was thought to reflect the activation of a neural system sensitive to discrepancies between intended and actual responses. This ERP component can be observed as a negative-going deflection over central and frontal midline sites, peaking between 50 and 100 ms after an erroneous response has been delivered (Gehring et al., 1993). The ERN is traditionally examined using speeded response tasks in which conflicting stimulus-response mappings are equally likely to occur, such as in a stimulus discrimination task with incongruent flanking stimuli, or when prepotent responses to target stimuli must be inhibited, as is the case in a Go/Nogo task. The elicitation of the ERN is not specific to errors committed with the hand, as it has been observed after foot (Holroyd et al., 1998), vocal (Masaki et al., 2001) and saccadic motor errors (Van 't Ent and Apkarian, 1999; Murphy et al., 2006), and indeed even when making partial mistakes (Vidal et al., 2000; Masaki

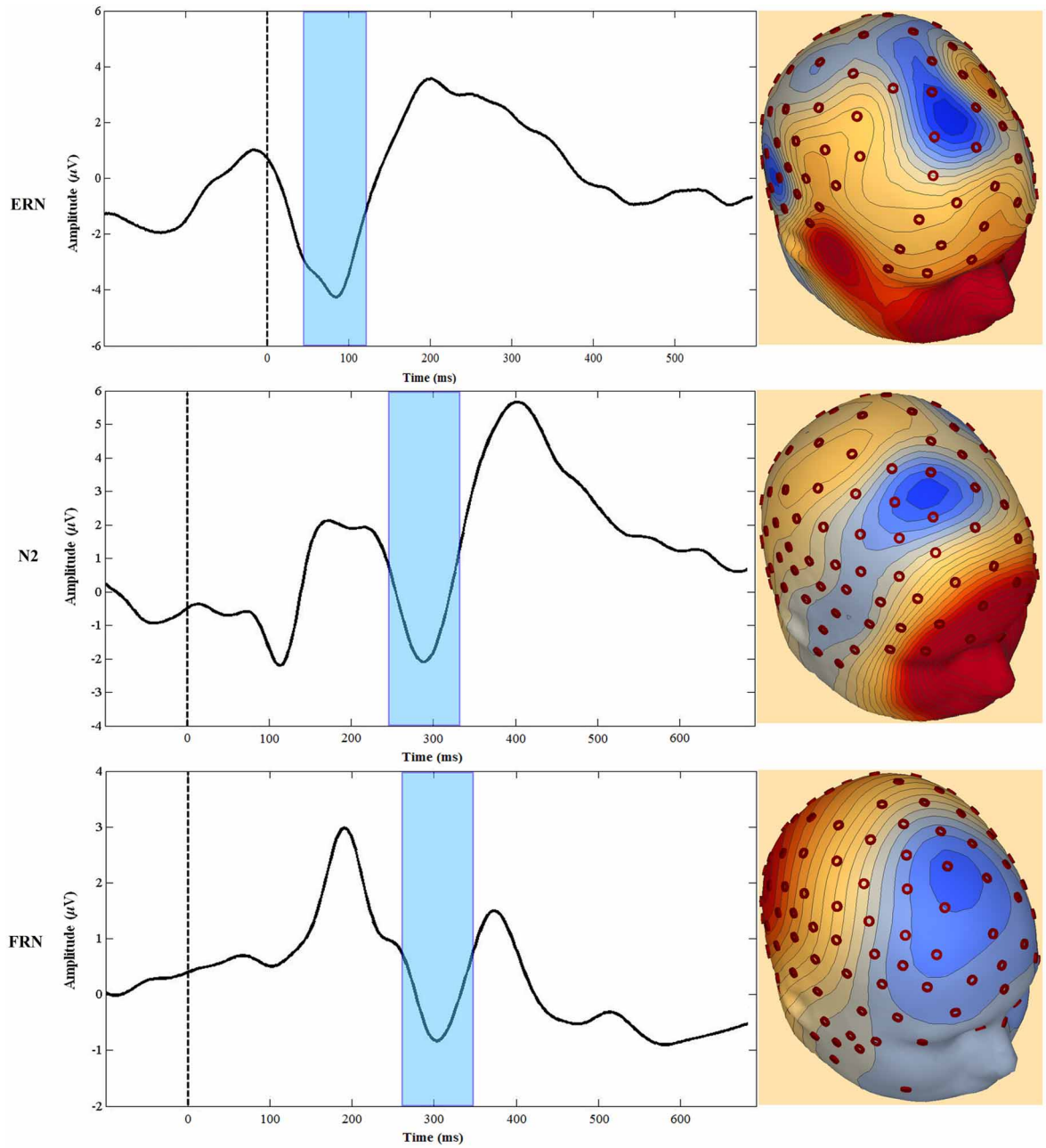


FIGURE 3 | Examples of waveforms and topographic voltage maps illustrating similarities between the ERN, Nogo N2 (N2), and feedback-related negativity (FRN). The FRN waveform is the difference between losing and winning trials in a gambling task (see text). The dashed line at 0 ms represents the onset of the time-locked event, and the shaded area highlights the peak negativity.

and Segalowitz, 2004) or only observing errors made by others (Miltner et al., 2004). Based on data from lesion, functional neuroimaging, and electroencephalography (EEG) source modeling methods, error-related responses have been localized to the dorsomedial prefrontal cortex (DMPFC) (Gehring et al., 1993; Dehaene et al., 1994; Carter et al., 2000; van Veen et al., 2001; van Veen and Carter, 2002a; Luu et al., 2003; Herrmann et al., 2004; Milham and Banich, 2005; Amodio et al., 2007; O'Connell et al., 2007) and, in some studies, ventromedial regions (Kiehl et al., 2000; Menon et al., 2001; Luu and Tucker, 2003; Stemmer et al., 2003; Ridderinkhof et al., 2004a; Swick and Turken, 2004; Taylor et al., 2007). Convergence has been observed across different functional measures, such that error-related scalp potentials correlate with functional magnetic resonance imaging (fMRI)

signals (Mathalon et al., 2003) and current source density (CSD; van Noordt, 2012) in the MPFC.

NOGO N2

When the participant has to withhold a response in the midst of responses that have become habitual (or prepotent), the N2 component of the ERP normally has an increased amplitude (Nieuwenhuis et al., 2003)[2]. Whereas the ERN is time-locked to response onset, the N2 is locked to the stimulus signaling that a response is to be withheld. The N2 has a similar scalp distribution as the ERN, peaking maximally over central and frontal midline sites and, using source analysis, has been shown to share neural generators in the medial frontal cortex (Bokura et al., 2001; Ullsperger and von Cramon, 2001; van Veen and Carter, 2002a,b; Bekker et al., 2005; Jonkman et al., 2007; Amodio et al., 2008; Gründler et al., 2009). As with the ERN, fMRI activation in the MPFC during inhibition of a response to a Nogo cue has also been shown to relate to N2 scalp amplitudes (Mathalon et al., 2003).

FRN

The FRN is similar to the N2 in that it is a stimulus-locked component, is negative in polarity, peaks at a similar latency (approximately 250 ms post-feedback) and therefore may be considered part of the N2 family (see Holroyd, 2004, for a discussion). Compared to the ERN, which reflects the activation of an internal monitoring system, the FRN reflects activity associated with external monitoring (Gentsch et al., 2009) and is time-locked to the external feedback stimulus informing the participant about an environmental (e.g., win or loss) or behavioral (e.g., correct or incorrect) consequence. Typically, the FRN is investigated using gambling (Gehring and Willoughby, 2002) or associative learning paradigms (Nieuwenhuis et al., 2004) in which individuals make choices between stimuli that are characterized by varying features (e.g., riskiness, magnitude, and probability), or attempt to learn action-outcome contingencies on the basis of feedback information (Holroyd and Coles, 2008). The FRN component is not modality specific (Miltner et al., 1997), and considerable evidence suggests that the FRN reflects, to some extent, the evaluation or appraisal of outcomes (Luu et al., 2003; Holroyd et al., 2006), particularly in the context of reinforcement learning (Yeung et al., 2005; Holroyd et al., 2009; Pfabigan et al., 2011). Several groups have reported larger FRN amplitudes to feedback indicating that behavior was incorrect (Miltner et al., 1997; Sato et al., 2005) or that an outcome has resulted in a loss or punishment (e.g., Gehring and Willoughby, 2002; Pfabigan et al., 2011). In addition, FRN amplitude is sensitive to prediction errors (Holroyd and Coles, 2002), unexpected outcome deviations (e.g., false-positive feedback; Oliveira et al., 2007), and predicts future behavioral responses, such as the avoidance of choices which were previously incorrect (Yasuda et al., 2004; Cohen and Ranganath, 2007; van der Helden et al., 2010), or the acceptance of unfair offers from others (Hewig et al., 2011).

Scalp distributions for the FRN suggest that peak activation occurs over sites slightly more anterior to those at which the ERN and N2 are often found to be maximal (Gehring and Willoughby, 2004; Muller et al., 2005; Luu et al., 2009), and are accounted for by source models which often include VMPFC regions (Luu and Posner, 2003; Muller et al., 2005; Nieuwenhuis et al., 2005; Luu et al., 2007; Kamarajan et al., 2009; Polezzi et al., 2010; Segalowitz et al., 2010), suggesting the possibility of additional underlying cortical generators. Nevertheless, similar regions of the perigenual ACC implicated in generating the ERN have been found for the FRN (Bellebaum and Daum, 2008), and results from our lab suggest that activation in both ventral and dorsal medial regions of the PFC at the time of error feedback correlate with FRN scalp amplitudes (Segalowitz et al., 2010, 2012; van Noordt, 2012).

Although not the focus of the present review, it is worth noting that relatively little research has considered the relative regional contribution from areas of the MPFC (e.g., dorsal-ventral) in generating these ERPs, and whether this may vary due to contextual influences because of the impact on arousal and affect. By extension, not much research has focused on how the interactions between personality and task context relate to differences in regional activation.

VARIATION IN MEDIAL PREFRONTAL ACTIVATION ACROSS CONTEXTS

MEDIAL PREFRONTAL ACTIVITY CONTEXTS INFLUENCING MOTIVATION AND AFFECT

Several research groups have manipulated task context and introduced affective content in order to investigate the effects on MPFC activation. Generally, manipulations aimed at influencing arousal and anxiety are associated with increases in MPFC activation, including additional neuronal generators in ventromedial regions. For example, Taylor et al. (2006) recorded fMRI responses while participants completed a modified flanker task involving blocks with different monetary incentives for performance. Their results show that, compared to the non-incentive condition, hemodynamic error responses in VMPFC regions were significantly greater when errors resulted in monetary loss. More recently, Simões-Franklin et al. (2010) employed a similar manipulation in a Go/Nogo task and found that phasic activation in the VMPFC (but not dorsal) to errors was significantly greater in the punishment compared to the neutral condition. In a gender voice decision task, involving neutral and emotional words and either congruent (i.e., auditory presentation and response side matching for gender) or incongruent (i.e., auditory presentation and response side not matching for gender) trials, Kanske and Kotz (2011) found additional recruitment of the VMPFC when participants resolved conflicting stimulus-response mappings, but only when the words were emotional. Together, these functional imaging studies illustrate that the engagement of the MPFC is sensitive to task context and the presence of affective content, and this is particularly the case for ventromedial regions. Similar results have been reported in several EEG studies.

Manipulating the monetary value of errors demonstrates that error-related brain responses are larger when mistakes result in punishment (Potts, 2011), or are associated with high compared to low monetary value (Hajcak et al., 2005). With respect to

[2] An N2 component is generated in many contexts, not necessarily representing the same generator as in the Nogo context. When we refer to the N2, we are doing so only to the inhibitory nogo context within a Go-Nogo task.

sources, VMPFC regions have been implicated when performance monitoring is being executed in arousing contexts, or when individuals are processing emotional information and feedback (Luu et al., 2003). For example, compared to verbal encouragement, derogatory feedback has been shown to increase ERN amplitudes (Wiswede et al., 2009) and, more recently, researchers have shown that verbal admonishment following erroneous responses is associated with additional recruitment of ventromedial sources (Ogawa et al., 2011), at least in females. The sex distribution of the sample may be important when examining the effects of arousing contexts of brain function. For example, in contrast to some of these results, Clayson et al. (2011) found that manipulating state affect had little influence on behavioral or ERP measures; however, forty percent of Clayson et al.'s (2011) sample ($n = 69$) was comprised of males, whereas the samples in the Wiswede et al. (2009; $n = 28$) and Ogawa et al. (2011; $n = 15$) studies were exclusively female.

Inhibitory N2s are also sensitive to arousal manipulations, as revealed by larger amplitudes (Potts, 2011) and ventral sources during conditions of distress and anxiety as compared to neutral and positive conditions (Lewis et al., 2006; Lamm et al., 2011). Overall, contextual demands influence functioning of the MPFC, and additional activation and recruitment of ventral sources of the MPFC may occur when tasks involve responding to emotional content, increased arousal, or motivational pressures.

MEDIAL PREFRONTAL ACTIVITY IN SOCIAL CONTEXTS

Other researchers have focused on the influence of social context on MPFC-related brain responses, introducing social pressures by including performance comparisons or by having participants monitor their performance in the presence of others. For example, Yu and Zhou (2006) found similar FRN effects (loss compared with gain), regardless of whether the feedback was self-relevant or related to another's performance. In other studies, increased FRNs were observed when feedback indicated that someone else had performed better (Boksem et al., 2011, 2012). Similar results were found when the research participant's outcome was yoked to that of another performer. Itagaki and Katayama (2008) collected FRNs in a gambling task to feedback which indicated whether or not the other's performance resulted in the participant winning or losing. Participants produced FRNs not only to the other person's losses, but also when the other person's wins resulted in losses for them. Marco-Pallarés et al. (2010) also found that FRN amplitudes were larger when outcomes resulted in wins for others and losses for the self, as compared to feedback indicating similar outcomes.

Using a social comparison model, Kim et al. (2011) had participants rate faces for attractiveness and then presented them with feedback about how deviant their rating was from an average. Medial frontal responses were found to be larger when feedback indicated that participants' responses differed from the group average.

These studies indicate that the response of the MPFC to evaluative feedback can be found in a wide variety of contexts, including those indicating subtlety of social comparisons. Considering the wide variation in personal responses to this kind of contextual information, these findings reinforce the need to consider individual differences in medial frontal responses.

VARIATION IN MEDIAL PREFRONTAL ACTIVATION ACROSS INDIVIDUALS

MEDIAL FRONTAL NEGATIVITIES, PERSONALITY, AND TEMPERAMENT

We increasingly find studies focused on exploiting individual differences in MPFC activation, particularly with respect to variation in personality and temperament. In general, individuals who score higher on measures of behavioral inhibition, withdrawal, or negative affect produce larger medial frontal responses. Researchers working from Gray's (Gray, 1987, 1989) approach-avoidance model find that larger ERN amplitudes are associated with higher scores on the Behavioral Inhibition Scale (BIS) (Boksem et al., 2006a). Using a Go/Nogo task to collect MFNs, Amodio et al. (2008) reported that higher BIS scores were associated with larger amplitudes (i.e., more negative) of both the ERN and the N2. These effects remained after adjusting for scores on the Behavioral Activation System (BAS) and for the left-right frontal alpha asymmetry, suggesting that it is negative affect and not its associated withdrawal tendency that underlies the increased medial frontal activation in their study (Davidson and Irwin, 1999; Coan and Allen, 2003; Davidson, 2004). We have found similar effects in our lab when investigating medial frontal activation to monetary wins and losses in a gambling task. In line with others' results (Amodio et al., 2008), we showed that the level of punishment sensitivity correlated with FRN amplitude, even after accounting for reward sensitivity and sex differences. Although women demonstrated larger FRNs than men, the gender difference was accounted for by the women's higher levels of sensitivity to punishment. Consistent with the summary on sources outlined above, punishment sensitivity was also associated with greater activation in the VMPFC during the FRN (Santesso et al., 2011).

Similar to the focus on the approach-avoidance dimension, predispositions toward internalizing and externalizing in children relate to increased and decreased activation of the MPFC, respectively. Generally speaking, internalizing is characterized by maladaptive self-focusing on internal negative mood states (e.g., anxiety, depression), whereas externalizing reflects anti-social behavioral tendencies. Results from our lab show that in 10 year olds, poorer socialization (e.g., higher scores of lying and psychoticism) is correlated with smaller ERN amplitudes (Santesso et al., 2005). In a separate study, Stieben et al. (2007) reported that, compared to controls and those co-morbid for externalizing and internalizing tendencies, inhibitory N2 and ERN signals were attenuated in children with pure externalizing symptomatology. Similarly, Moadab et al.'s (2010) examined the N2 and ERN in 9–13 years olds using an emotional Go/Nogo task and found that these MFNs were larger in those scoring higher in internalizing.

Anxiety symptoms have also been related to VMPFC activation in terms of timing rather than amplitude. Lamm et al. (2011) used a Go/Nogo task involving a negative emotion induction (where the participant loses points) while obtaining Nogo N2 amplitudes in anxious aggressive 8–12 year-old children.

During emotion induction, anxious aggressive children showed strong engagement of VMPFC regions during the early stages of inhibitory control (200–300 ms post Nogo stimulus), whereas non-anxious aggressive children showed the dominance of ventral regions during the later stages of behavioral inhibition (400–500 ms post Nogo stimulus). These patterns were interpreted as reflecting an early anxious response due to increased demands on cognitive control in the anxious-aggressive children, versus a later frustration response due to the increased pressure to regulate behavior in the non-anxious aggressive children.

Together, these results indicate the importance of medial prefrontal functioning and regional activation in temperament variation, and support how underlying mechanisms for these differences can be observed early in development.

MEDIAL FRONTAL NEGATIVITIES AND TEMPERAMENT FACTORS IN RISK-TAKING

Investigating the neural correlates of approach-avoidance tendency is particularly relevant to understanding individual differences in risk-taking behaviors. Approach-related behaviors are core to risk-taking, and studies implicate deactivation of the MPFC during performance monitoring as a neural correlate of approach-related dispositions. In a sample of young males, Santesso and Segalowitz (2009) found that individuals scoring higher on sensation-seeking and reward sensitivity produced lower levels of medial frontal activity following erroneous behavioral responses. Similar effects were observed when we used a modified version of the Balloon Analogue Risk Task (BART) in a sample of 28 university students (van Noordt, 2012). In the standard BART, participants inflate a balloon in order to collect points or money, but are faced with the possibility that the balloon could pop, resulting in a loss of the accrued points. Risk-taking is indexed as the number of pumps on those trials on which the balloon did not pop, and is associated with approach-related behaviors (e.g., sensation seeking, impulsivity) and self-reported risk-taking (Lejuez et al., 2003b), as well as self-reports of addiction (Hopko et al., 2006) and detrimental health behaviors (Lejuez et al., 2002, 2003a). In our version, participants decided when to stop the continuous inflation of a balloon in order to collect their points, allowing us to record FRNs to loss feedback (i.e., trials in which the balloon popped which, in our task, also resulted in the participant losing 10% of their previously accumulated winnings). Using standardized low resolution brain electromagnetic tomography (sLORETA; Pascual-Marqui, 2002) to model source activation during the FRN, we found that CSD in the VMPFC correlated with risk-taking (i.e., the amount of time individuals permit the balloon to inflate on win trials), such that lesser VMPFC activation predicted a greater willingness to exhibit behaviors which ultimately become disadvantageous in the BART.

These effects are clarified further by research aimed at disentangling risk-taking profiles across contexts. Polezzi et al. (2010) found that FRN amplitudes did not differentiate between outcomes as long as participants were in their comfort zone for risk-taking (greater for some, less for others). In one condition the gains and losses were of equal magnitude (zero-expected value), whereas in another gains were larger than losses (positive-expected value). Individual brain responses to feedback did not differentiate between gains and losses only in the context in which the participant was more likely to take risks and seemed to be insensitive to the possibility of losing. Thus, individual differences in risk-taking behaviors relate to MFNs in terms of the subjective evaluation of risk.

MEDIAL FRONTAL NEGATIVITIES AND SOCIOPOLITICAL ORIENTATIONS

In addition to the associations described previously, activation of the MPFC has also been related to constructs seemingly more distal from biological temperament, including those reflecting social, political and religious orientations. In a sample of American undergraduates, Amodio et al. (2007) found that students who self-identified as being more liberal showed larger ERN and N2 amplitudes in a Go/Nogo task. We have recently replicated and extended these findings by showing that greater medial frontal activation is associated with a greater predilection for egalitarianism and social change and inversely with traditionalism (Weissflog et al., 2010). Similar to studies showing that conservative orientations relate to reduced engagement of the MPFC during performance monitoring, Inzlicht et al. (2009) found that stronger religious zeal and belief in god were associated with reduced electrocortical activation following error commission.

Given the dynamic developmental relations between activation and cortical growth, one might speculate that the relations should extend to tissue size as well. Unfortunately, there are few studies reporting actual physical size of the ACC and associated medial frontal structures as they relate to personality variables or social attitudes (although see, for example, Whittle et al., 2008a,b, 2009a,b) but one recent report is relevant to the social attitudes research described above. Kanai et al. (2011) reported that greater liberalism is associated with larger ACC size, and that greater conservatism is associated with increased size of the right amygdala. Such anatomical reports, if replicated, lead to intriguing hypotheses concerning how to characterize such differences in temperament, although they do not resolve the issue of cause and effect, considering the degree of plasticity of neural networking in both these structures and their sensitivity to experience (Vyas et al., 2002; Cook and Wellman, 2004; Mitra et al., 2005; Radley et al., 2006a,b; Dias-Ferreira et al., 2009; Liston et al., 2009).

MEDIAL FRONTAL NEGATIVITIES AND STATE-TRAIT MOOD AND AFFECT: CLINICAL SAMPLES

Both clinical and non-clinical levels of anxiety, neuroticism and emotionality relate to medial frontal activation. Generally, greater levels of anxiety (Goldin et al., 2009), worry (Endrass et al., 2010), neuroticism (Pailing and Segalowitz, 2004; Olvet and Hajcak, 2012), social distress (Eisenberger and Lieberman, 2004), negative affect (Luu et al., 2000; Olvet and Hajcak, 2012; Santesso et al., 2011), and emotional reactivity (Fukushima and Hiraki, 2009) are associated with increased activation of the MPFC.

Obsessive–compulsive disorder (OCD) is characterized by negative invasive thought patterns that engender anxiety and worry about subsequent remedial behaviors, and greater neural

responses have been observed in clinical groups (Gehring et al., 2000; Ursu et al., 2003; Fitzgerald et al., 2005; Ruchsow et al., 2005; Endrass et al., 2008; Hajcak and Olvet, 2008) and in non-clinical samples with respect to obsessive-compulsive (OC) behaviors (Santesso et al., 2006), in children (Hajcak and Olvet, 2008) and adults (Endrass et al., 2008, 2010), and during both correct and error trials (Endrass et al., 2008). In addition to group-level effects, medial frontal activation has also been shown to increase as a function of symptom severity (Gehring et al., 2000; Xiao et al., 2011), with activation of ventromedial regions being especially related to symptomatology (Fitzgerald et al., 2005).

Similar to the findings in persons with OCD, hyperactivation of the performance monitoring system has also been observed in persons with depression (Chiu and Deldin, 2007; Holmes and Pizzagalli, 2008; Mies et al., 2011), including those in remitted stages of the disorder (Santesso et al., 2008; Georgiadi et al., 2011). In their study, Santesso et al. (2008) found that, compared to controls, persons with remitted depression had larger FRNs even after controlling for residual symptoms of anxiety and depression. Beyond group differences, higher levels of depressive symptoms are related to larger error-related brain responses (Fitzgerald et al., 2005; Chiu and Deldin, 2007; Weinberg et al., 2010), and the extent to which neural responses differentiate correct (or reward) from error (or loss/non-reward) responses is associated with depression severity (Foti and Hajcak, 2009; Olvet et al., 2010). However, some researchers have reported null or opposite effects (see Ruchsow et al., 2004, 2005; Schrijvers et al., 2008, 2009; Olvet et al., 2010).

Several reviews have focused on the functional significance of MFNs in relation to anxiety and performance monitoring (see Robinson et al., 2010; de Bruijn and Ullsperger, 2011; Lee and Park, 2011; Weinberg et al., 2012). Briefly, similar to the results from studies focusing on OCD or depression, individuals with high levels of generalized anxiety show larger electrocortical MPFC responses following errors (Weinberg et al., 2010; Xiao et al., 2011). Both ERN (Weinberg et al., 2010) and FRN (Gu et al., 2010) amplitudes have been shown to differentiate individuals' anxiety levels, such that more severe symptoms are associated with larger scalp negativities. These findings support the notion that errors provoke defensive responses, and that error-related brain responses may be a marker for individual differences in defensive reactivity (see Weinberg et al., 2012) and susceptibility to anxiety-related psychopathology (Olvet and Hajcak, 2008; Robinson et al., 2010).

The findings described above suggest that MFNs may represent a neurophysiological marker (i.e., endophenotype) for adaptive self-regulation of anxiety and arousal. However, one could also ask whether MFN amplitudes are a result of the person's psychological state and not a trait predisposition. This issue is, of course, difficult to resolve with human research participants because we cannot manipulate the clinical status or personality trait of the individual in order to see how this affects MFNs. It may be the case that a raised level of anxiety increases the reactivity of the medial frontal cortex, or it may be the case that a more reactive medial frontal cortex produces the anxiety symptoms. Evidence in favor of the latter position comes from the finding that persons with OCD produce similar ERNs regardless of punishment associated with their errors (Endrass et al., 2010), and that successful treatment does not attenuate ERN amplitude (Hajcak and Olvet, 2008). This suggests that MFNs may represent an endophenotype (Hajcak and Olvet, 2008; Ullsperger, 2009) of vulnerability for a limited capacity for adaptive self-regulation, and that when attenuation of symptoms result from treatment, this is not done by altering the underlying susceptibility of the person to the illness but by some top-down control over behavior and mental state. To definitively test this hypothesis, one would need to follow patients until complete remission is demonstrated, and then one might find a regularized MFN. However, such studies have not yet been done.

Such studies would be especially important for understanding the associations between brain function and symptomatology, given the evidence suggesting that MPFC functioning relates to treatment effects. For example, there is an extensive literature focused on the relationships between the structural and functional integrity of the VMPFC, particularly subgenual ACC [SGACC; Brodmann Area (BA) 25], and the regulation of mood and affect. Structurally, reduced gray matter volume in or near the SGACC has been found in persons with depression (Boes et al., 2008), in cases of early-onset depression (Botteron et al., 2002), as well as those suffering from other symptoms of mood dysregulation (Drevets et al., 2008). Functionally, the role of the SGACC in mood regulation is reflected in activation patterns. Individuals with family history of mood disorders have been found to exhibit reduced glucose metabolism in SGACC (Drevets et al., 1997), a finding which has also been reported in persons with depression characterized by anhedonia (Pizzagalli et al., 2004). The role of the SGACC, and more broadly the VMPFC, in regulating mood is supported by a growing body of evidence showing that dysregulation in fronto-limbic regions is associated with response to treatment.

In the 1990s, Mayberg and colleagues showed that individual differences in the activation of the cingulate cortex related to treatment efficacy, such that greater activation predicted better response to treatment (Mayberg, 1997; Mayberg et al., 1997). These findings have been extended using high-density EEG recordings in order to model CSD of theta power in the ACC (Pizzagalli et al., 2001). As is the case with hemodynamic measures, Pizzagalli et al. (2001) found that greater activation in multiple anatomical regions of the ACC prior to treatment predicted better outcomes post-treatment. Based on some of these findings, the SGACC has been proposed as an important cortical region which serves as a nexus for supporting processes of self-reference, as well as modulating the functional relationships between other prefrontal areas involved in cognitive control (see Pizzagalli, 2011, for a recent review). Using fMRI, the results from Yoshimura et al.'s (2010) study support the role of the VMPFC as an important cortical region involved in mediating emotional and cognitive self-control. These researchers report that cortical activation near the SGACC mediated the relationship between depressive symptoms and activation of other medial prefrontal regions involved in self-regulation. Taken together, there is good evidence that the functioning of the SGACC, and the VMPFC more generally, supports affective/evaluative processes and is

associated with temperament, personality, and mood, especially in relation to negative affect and anxiety.

MEDIAL FRONTAL NEGATIVITIES AND STATE-TRAIT MOOD AND AFFECT: NON-CLINICAL SAMPLES

In addition to clinical data, associations between cortical activation, personality, and mood are observed in sub/non-clinical samples (Xiao et al., 2011). Similar to those with clinical symptoms, college students who score high on the Obsessive Compulsive Inventory (Hajcak and Simons, 2002), as well as those scoring higher on measures of general anxiety (Hajcak et al., 2003; Xiao et al., 2011) or depression (Xiao et al., 2011), elicit larger ERNs than those scoring lower on these measures. The results of several studies show that factors such as fatigue (Boksem et al., 2006b) task involvement (Yeung et al., 2005; Tops and Boksem, 2010) and perceived responsibility for outcomes (Li et al., 2010, 2011) modulate MFN amplitudes. In addition, greater self-reported negative affect (Luu et al., 2000; Hajcak et al., 2004; Yasuda et al., 2004; Sato et al., 2005; Santesso et al., 2011) and neuroticism (Pailing and Segalowitz, 2004; Eisenberger et al., 2005; Olvet and Hajcak, 2012) also relate to enhanced neuronal activation to error or loss/negative feedback. Even more abstract constructs such as empathy have been found to relate to MFN amplitude, such that persons who are more empathic have larger (i.e., more negative) MFNs (Santesso and Segalowitz, 2009; Larson et al., 2010).

Anxiety in non-clinical samples dissociates physiological responses to error feedback. For example, Santesso et al. (2011) found that healthy adults with higher scores in negative emotionality produce larger FRNs to negative feedback in a monetary incentive task, as well greater activation in VMPFC, possibly reflecting rapid affective processing of negative feedback. In their study, Hajcak et al. (2004) found that, compared to those low in negative affect, individuals high in negative affect produced larger ERNs and greater skin conductance responses following errors. These findings suggest that higher levels of negative affect are associated with a systemic hyperactivation of the nervous system, as reflected by greater responses in both the central and autonomic branches. Similarly, with respect to the FRN, amplitudes have been shown to predict an individual's willingness to reject unfair offers. These decisions are associated with higher levels of negative affect and sympathetic activation (Hewig et al., 2011). Taken together, differences in temperament styles are reflected by the variability in MPFC activity between groups, as well as across individuals. Examining the associations among brain function, temperament, and personality is not only relevant to understanding the neural underpinnings of real-world behaviors, but can also be important for understanding dysfunctional cognitive and affective processes.

INTERACTIONS BETWEEN PERSONALITY AND CONTEXT ON MEDIAL FRONTAL ACTIVATION

Having summarized the effects of task demands and personality on activation of the medial frontal cortex, we should also examine interactions between these broad factors. Such interactions are critical for disentangling mediating and moderating factors in models of performance monitoring.

INTERACTIONS IN CONTEXTS INVOLVING PERFORMANCE-RELATED INCENTIVES

It may be that the degree to which context affects brain responses varies in relation to personality characteristics. For example, we found that individuals who are high in conscientiousness are less sensitive to task manipulations aimed at increasing error significance, as reflected by their larger ERNs for all errors. Conversely, those lower in conscientiousness varied their ERNs as a function of how much their erroneous responses cost, showing larger ERNs when errors were associated with relatively more severe monetary punishments (Pailing and Segalowitz, 2004). Boksem and colleagues have also found interaction effects when investigating personality and temperament. Specifically, persons scoring high in behavioral inhibition not only generate larger ERNs (Boksem et al., 2006a), but this effect is also greater when errors are associated with losing money (Boksem et al., 2008). These data illustrate the interactions between context and personality on brain activation, given that persons who are behaviorally inhibited or have lower self-confidence are more sensitive to being punished for their mistakes, as reflected by their MPFC activity.

With respect to approach behaviors, extraverted individuals are considered to be more approach-oriented and driven by novelty, sensation-seeking, and rewarding outcomes (Campbell et al., 2003; Cohen et al., 2005). Smillie et al. (2010) manipulated feedback frequency (with 80% expected vs 20% unexpected) and outcome type (reward versus non-reward) and found that, compared to those scoring low on extraversion, individuals high in extraversion generated larger FRNs to unexpected reward outcomes, and smaller FRNs to unexpected non-reward outcomes. These results illustrate that those individuals who find novelty and rewards more salient have enhanced MPFC activation to unexpected reward and attenuated MPFC to non-reward outcomes, respectively. Together, these studies highlight how neither individual differences on traits related to performance monitoring nor task demands necessarily act on their own.

INTERACTIONS IN CLINICAL SAMPLES

It is not surprising that interactions among context, personality and brain activation are observed in clinical samples. For example, we reported that incarcerated psychopaths produce attenuated error-related brain responses only when having to deal with affective stimuli that they are known to have difficulty processing (i.e., emotional faces). However, there was no difference in ERN amplitudes between psychopaths and controls when collected in response to errors on a standard letter flanker task (Munro et al., 2007a), suggesting that their performance monitoring system is as sensitive as that of controls when mistakes occur in a non-affective context. In addition, we found no evidence of inhibitory control deficiency in psychopaths, as indexed by N2 amplitudes in a non-affective paradigm, whereas non-psychopathic incarcerated offenders did produce attenuated N2 responses, possibly reflecting lower levels of inhibitory control (Munro et al., 2007b). These studies illustrate that the way the brain responds to performance feedback across contexts varies in relation to personality differences. Moreover, these results caution against treating all

MFNs as reflecting a single construct considering that context can dissociate them.

Interactions between individual differences and context have also been investigated in other clinical samples characterized by mood dysregulation. The difference among the clinical presentations may be reflected in differences in the relative balance of regional activation across the MPFC. For example, symptoms of OCD, neuroticism, anxiety and negative affect may involve a relatively stronger engagement of the VMPFC as compared to dorsal regions (Fitzgerald et al., 2005). Support for this regional differentiation was reported by Gründler et al. (2009) and was explored further by Cavanagh et al. (2010), who found that individual differences in OC symptomatology were characterized by different MPFC activation profiles at rest and during performance monitoring. At rest, OC symptomatology correlated positively and negatively with activity in the VMPFC and DMPFC, respectively. Thus, even when there is no demand for performance monitoring, individuals more prone to experience negative intrusive thoughts and anxiety show an increased activity in medial prefrontal regions involved in saliency appraisal and sympathetic modulation (i.e., VMPFC). Moreover, these individuals show disengagement of regions typically recruited to when cognitive control is needed to regulate behavior (i.e., DMPFC). While monitoring their performance, individuals in the high OC group had hyperactivation of the VMPFC to errors on a flanker task and hypoactivation in the DMPFC to error feedback on a reinforcement learning task. These results suggest that when persons characterized by pathological levels of anxiety and worry make mistakes, they show larger responses in medial prefrontal regions implicated in feedback evaluation and affect regulation. Moreover, when these individuals fail to learn from feedback they produce relatively little activity in prefrontal regions involved in the cognitive control of behavior. Findings such as these not only illustrate the complexity of the interactions between brain activity, context, and individual differences, but also shed light onto how brain-behavior relationships may reflect maladaptive self-regulation.

INTERACTIONS IN NON-CLINICAL SAMPLES

Non-clinical experimental manipulations also reveal that personality and context interact to influence medial prefrontal responses. Olvet and Hajcak (2012) randomly assigned participants to be exposed to either neutral or sad media clips prior to performing a flanker task. They found that, following sad mood induction, greater self-reports of sadness were associated with larger ERNs. In addition, this effect was moderated by neuroticism such that persons higher in neuroticism exhibited a stronger coupling between sad mood and error-related brain responses. Using a different manipulation to punish errors on a flanker task, similar findings were reported by Riesel et al. (2012). In contrast to neutral blocks in which errors were never punished, 50% of errors were followed by an aversive sound in punishment blocks during the first half of the experiment (acquisition phase). Although the aversive sound to errors was removed during the second half of the experiment (extinction phase), participants still generated larger ERNs to these errors compared to errors made in neutral blocks. As would be predicted, the effect of punishment context on medial frontal activation was greater for persons scoring higher on trait anxiety. Thus, individuals who are more prone to worry and experience negative affect are especially sensitive to punishment-related contexts as reflected by electrocortical responses. Although these studies did not examine source activation, other studies focusing on these issues strengthen the association described earlier linking VMPFC regions with negative affect states in clinical populations and their effect on MFNs.

INTERACTIONS IN SOCIAL CONTEXTS

Social factors can also, of course, affect how individual differences in personality relate to medial frontal functioning, which has clinical as well as theoretical implications. From a clinical perspective, such effects may help identify which individuals have a predilection for maladaptive responses when their performance is worse than that of others, or when they are monitoring their performance in competitive situations. For example, Chein et al. (2011) showed that peer presence increased the activation of the incentive system (ventral striatum and orbitofrontal cortex) in adolescents when they were taking risks in a videogame designed to encourage dangerous driving. Peer presence did not influence adults in this way. In a similar research design, we found that peer presence selectively reduced the FRN produced by 15 year-old boys when they lost points due to excessive risk-taking in a similar videogame (Segalowitz et al., 2012). However, we also found that this effect was influenced by individual differences, where higher scores on sensation-seeking, behavioral activation, and sensitivity to reward (summed together as a measure of "surgency") was associated with greater reduction in the FRN (see **Figure 4**). Regional source modeling especially implicated the

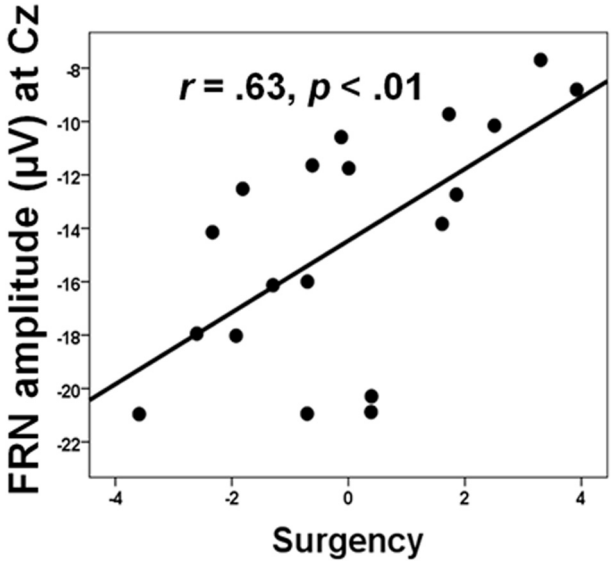


FIGURE 4 | The relation between FRN amplitude and surgency, a composite of BAS, sensation seeking (SSS-V), and positive affect (PANAS) in the "alone" condition. Data from Segalowitz et al. (2012).

VMPFC, although regions of the ACC, including those more dorsal, were also active. However, it is not possible from these data to discriminate between the possibilities that individuals higher in surgency engage in riskier behavior, particularly in the presence of their peers, as a result of hypoactivation in the VMPFC, or that the VMPFC activates less in these individuals because of their personality traits.

Focusing on the nature of the relationship among individuals, Newman-Norlund et al. (2009) examined the influence of friendship on performance monitoring. In their study, participants observed outcomes of virtual penalty kicks for characters labeled as stranger or friend. Even though the stranger-friend associations were established only during the experiment itself, observing a virtual friend's failure engaged performance monitoring networks to a greater extent than witnessing a stranger perform sub-optimally. These effects have been extended by Kang et al. (2010) who recorded FRNs while participants observed an actual friend or a stranger perform the Stroop task. Not only were FRNs larger for a friend's errors compared to those of a stranger, but this effect was mediated by the extent to which the participant considered their friend as part of their self-concept. Thus, watching a friend make mistakes engages performance monitoring networks to a greater degree than witnessing mistakes made by a stranger, and this engagement is larger when individuals perceive themselves to be closer to their friend.

A similar avenue of social neuroscience research focuses on the association between fairness of outcomes and medial frontal activation. In paradigms such as the Ultimatum Game, researchers have reported that highly unfair offers elicits greater MFN activation compared to more fair offers (Van der Veen and Sahibdin, 2011; Wu et al., 2011), and differentially impact peripheral nervous system responses, such as cardiac (Van der Veen and Sahibdin, 2011) and skin conductance responses (Hewig et al., 2011). In two recent studies, outcomes have been shown to interact with individuals' perceptions of fairness. Boksem and De Cremer (2009) collected FRNs to outcomes in the Ultimatum game and found that unfair offers were not only associated with larger amplitudes, but that the effect was strongest for individuals reporting high concerns for fairness. Using a different paradigm, the Dictator Game, Wu et al. (2011) found the FRN to be differentially sensitive to the fairness of outcomes depending on the source of the offer. Specifically, amplitudes were larger to unfair compared to fair outcomes when the offers were made by friends, whereas FRNs did not differ to when offers were made by a stranger.

INTERACTIONS: GENES, NEUROTRANSMITTERS, AND PERSONALITY

Some research has focused on the association between hormones (e.g., cortisol; Tops et al., 2006; Cavanagh and Allen, 2008; Tops and Boksem, 2011) and various neurotransmitters and performance monitoring processes. Several genetic polymorphisms have been shown to affect MFNs (see Jocham and Ullsperger, 2009; Ullsperger, 2009, 2011 for reviews). In the context of performance monitoring, levels of error-related brain activity and corrective behavior are a function of polymorphisms on the catechol-O-methyltransferase (COMT) genotype and, as a result, tonic levels of PFC dopamine (Mueller et al., 2011). Other researchers have focused on allelic differences in genes coding for prefrontal dopamine receptors which are also associated with variations in both error-related brain activity and post-error behavioral adjustments (Kramer et al., 2007).

Serotonin genes have also been associated with MFNs. The variant of 5-HTTLPR which has one or two repeats is associated with lower activity of the serotonergic system, whereas the homozygous long form allele is associated with increased functioning of the 5-HTT system. Fallgatter et al. (2004) found that individuals who have lower 5-HTT function (the short variant) elicit larger ERNs to errors on a letter flanker task. This finding fits well with studies showing that lower levels of serotonin levels are associated with higher levels of anxiety, negative emotionality, and depression (e.g., Karg et al., 2011), all of which are symptoms known to relate to hyperactivation in the MPFC. With respect to depression, Holmes et al. (2010) used fMRI to examine the association between tandem repeats on the 5-HTTLPR gene, medial frontal engagement, and performance on a flanker task. Their findings indicate that persons with low 5-HTT function (the short variant) not only have less conflict-related activation (incongruent correct – congruent correct) in the DMPFC, but also engage the VMPFC to a greater degree following errors (incongruent error – incongruent correct).

Thus, individuals who are more susceptible to mood dysregulation and psychopathology hyperactivate regions thought to be predominantly involved in the modulation of arousal and affect when they make mistakes (i.e., VMPFC). Furthermore, these individuals also show a relative disengagement of prefrontal regions involved in mediating cognitive control (i.e., DMPFC), specifically when there is an increase in the demand to regulate behavior. In addition to these elegant findings, long allele carriers were more accurate following errors, suggesting increased vigilance in performance monitoring after instances of failure in persons who have a higher functioning serotonergic system and are less likely to develop depression.

SUMMARY, CONCLUSIONS AND CAVEATS

As we hope is evident from this review, factors affecting MPFC functioning and performance monitoring are indeed complex. The ERN, N2, and FRN are similar electrocortical responses generated by MPFC neurons, but are functionally distinct and reflect different aspects of performance monitoring. Similarly, although these MFNs have been localized to overlapping regional sources of the MPFC, distinct regions of the MPFC might differentially contribute to the generation of these ERP components. Due to these factors, MFNs, although having some similarities, should not be considered to reflect the same performance monitoring process. The complexity arises from the fact that these differences in brain function vary as a function of personality, task context, and their interactions.

Of course, although the interactions may be significant, caution should exercised when interpreting their complexity until replicated. Furthermore, there has been relatively little focus on the role of other cortical regions with respect to error and performance feedback processing despite consensus that we are seeking to understand the networks associated with performance monitoring, not the activation

of single regions. This is not to say that research aimed at synthesizing our understanding of personality with the role of the MPFC in performance monitoring is unfruitful. On the contrary, the relationship between personality differences and MPFC function is symbiotic at a theoretical level in that individual differences in medial frontal responses can add to our understanding of personality constructs, yet individual differences in personality and temperament that relate to variability in MPFC activation may also provide us with important information concerning the nature of performance monitoring brain responses. In other words, knowledge about a personality construct such as neuroticism is aided by knowing its relation to the structure and functioning of specific MPFC regions, such as the magnitude of response or engagement of dorsal versus ventral MPFC and how the task demands alter these relationships. Similarly, our understanding of the MPFC is aided by seeing to which personality constructs its activation relates. In this sense, this research presents an iterative learning process that supports the formulation, testing, and interpretation of hypotheses focused on the associations between personality, context, and functioning of the MPFC. Note, however, that this iterative process implies a difficulty in attributing a single cause-effect relationship between function and structure. Rather, the MPFC structure may heavily influence how the person responds to the task, with clear implications for how we interpret their personality, yet their personality predispositions may also help shape the structure and functioning of their MPFC over time.

It is important to note that most of the research on individual differences and MPFC functioning rely on cross-sectional, correlational designs. A consequence of this type of research is that causation cannot be inferred from the data, nor does this research directly investigate the mechanisms driving the phenomena of interest. To repeat an example raised earlier, it is not possible to discriminate between the possibility that individuals higher in surgency engage in riskier behavior, particularly in the presence of their peers, as a result of hypoactivation in the VMPFC, from the possibility that the VMPFC activates less in these individuals because of their personality traits (Segalowitz et al., 2012). In addition, although regional source modeling especially implicated the VMPFC, other ventral and dorsal regions of the MPFC were also active. Thus, more sophisticated experimental designs and longitudinal data are needed in order to disentangle issues of cause and effect with respect to personality, task context, and functioning of the MPFC. These studies will be especially important for expanding our clinical understanding of personality and mood disorders, as well as the effectiveness of various treatments.

We should also keep in mind that a MFN represents more than single regional response. There is little debate that information processing in the brain relies on the dynamic coordination of multiple complex neural networks. In order to truly appreciate the neural bases of behavior, an understanding of how various brain networks coordinate their activities to support a given process will be crucial (Pourtois et al., 2010). Specifically, variability in the structural and functional connectivity between regions of the MPFC and subcortical structures might account for individual differences in personality and performance, as well as how these factors interact with task context to impact MFNs (e.g., Cohen, 2011). Another possible research avenue is using Independent Components Analysis (ICA; Makeig et al., 1996) to better isolate independent cortical processes that contribute to variability in performance monitoring and personality. Once identified, these functionally independent components can be source localized to better understand the regional dynamics underlying MFNs. Furthermore, considering how the activation of different independent components or sources varies over time is another way to gain insight about how individual differences in network functioning relate to personality and task context.

ARE MFNs REFLECTING A COMMON GENERATOR AND IF NOT, DOES IT MATTER?

Although the notion of the ERN, FRN and Nogo N2 reflecting a common source generator persists, we think it is clear that it must be the case that they have (at most) something in common and much distinctive. This is partly because the tasks that elicit them are different from each other in fundamental ways, and therefore something reflecting this difference must be coded in the brain signal. However, more importantly, the standard tasks that elicit these components differ in the degree of affect and arousal that they elicit, and there is much evidence that these factors are important. Such empirical support of the components having separate sources is easy to find: Not only do the measures not intercorrelate highly all the time, their variance sometimes maps onto behavior in different ways. For example as mentioned earlier, we found a dissociation between ERNs and the Nogo N2 within a group of violent offenders. Such dissociations indicate that the psychological variables driving at least some of the generator sources may differ for the various MFNs. However, to fully document such differences, studies need to include multiple MFN measures on the same participants, something rarely done. In addition, of course, as illustrated above in terms of LORETA analyzes, the actual regions responsible for the negativity measured at the scalp may differ considerably for the three components either in specific locations or, more likely, in the balance of contribution from the MPFC subregions. We suggest that the relative contribution of the cortical sources underlying the ERN, N2, and FRN may depend on the specific stimuli or context used and the degree of emotional arousal engendered by the task demands.

The use of MFNs as a reflection of MPFC functioning has become well accepted in the research community, a fact well documented by the growth in research literature involving these electrophysiological components. However, the issues raised in this review suggest that despite this relative acceptance, some of the basic assumptions needed for their interpretation remain to be verified by future research.

ACKNOWLEDGMENTS

Stefon van Noordt was supported by the Canadian Institutes of Health Research Graduate Award.

REFERENCES

Amodio, D. M., Jost, J. T., Master, S. L., and Yee, C. M. (2007). Neurocognitive correlates of liberalism and conservatism. *Nat. Neurosci.* 10, 1246–1247.

Amodio, D. M., Master, S. L., Yee, C. M., and Taylor, S. E. (2008). Neurocognitive components of the behavioral inhibition and activation systems: implications for theories of self-regulation. *Psychophysiology* 45, 11–19.

Baumeister, R. F., Vohs, K. D., Nathan, D. C., and Liqing, Z. (2007). How emotion shapes behavior: feedback, anticipation, and reflection, rather than direct causation. *Pers. Soc. Psychol. Rev.* 11, 167–203.

Beckmann, M., Johansen-Berg, H., and Rushworth, M. F. S. (2009). Connectivity-based parcellation of human cingulate cortex and its relation to functional specialization. *J. Neurosci.* 29, 1175–1190.

Bekker, E. M., Kenemans, J. L., and Verbaten, M. N. (2005). Source analysis of the N2 in a cued Go/NoGo task. *Brain Res. Cogn. Brain Res.* 22, 221–231.

Bellebaum, C., and Daum, I. (2008). Learning-related changes in reward expectancy are reflected in the feedback-related negativity. *Eur. J. Neurosci.* 27, 1823–1835.

Boes, A. D., McCormick, L. M., Coryell, W. H., and Nopoulos, P. (2008). Rostral anterior cingulate cortex volume correlates with depressed mood in normal healthy children. *Biol. Psychiatry* 63, 391–397.

Boksem, M. A., and De Cremer, D. (2009). Fairness concerns predict medial frontal negativity amplitude in ultimatum bargaining. *Soc. Neurosci.* 5, 118–128.

Boksem, M. A., Tops, M., Wester, A. E., Meijman, T. F., and Lorist, M. M. (2006a). Error-related ERP components and individual differences in punishment and reward sensitivity. *Brain Res.* 1101, 92–101.

Boksem, M. A. S., Meijman, T. F., and Lorist, M. M. (2006b). Mental fatigue, motivation and action monitoring. *Biol. Psychol.* 72, 123–132.

Boksem, M. A. S., Kostermans, E., and De Cremer, D. (2011). Failing where others have succeeded: Medial Frontal Negativity tracks failure in a social context. *Psychophysiology* 48, 973–979.

Boksem, M. A. S., Kostermans, E., Milivojevic, B., and De Cremer, D. (2012). Social status determines how we monitor and evaluate our performance. *Soc. Cogn. Affect. Neurosci.* 7, 304–313.

Boksem, M. A. S., Tops, M., Kostermans, E., and De Cremer, D. (2008). Sensitivity to punishment and reward omission: evidence from error-related ERP components. *Biol. Psychol.* 79, 185–192.

Bokura, H., Yamaguchi, S., and Kobayashi, S. (2001). Electrophysiological correlates for response inhibition in a Go/NoGo task. *Clin. Neurophysiol.* 112, 2224–2232.

Botteron, K. N., Raichle, M. E., Drevets, W. C., Heath, A. C., and Todd, R. D. (2002). Volumetric reduction in left subgenual prefrontal cortex in early onset depression. *Biol. Psychiatry* 51, 342–344.

Campbell, L., Simpson, J. A., Stewart, M., and Manning, J. (2003). Putting personality in social context: extraversion, emergent leadership, and the availability of rewards. *Pers. Soc. Psychol. Bull.* 29, 1547–1559.

Carter, C. S., MacDonald, A. M., Botvinick, M., Ross, L. L., Stenger, V. A., Noll, D., and Cohen, J. D. (2000). Parsing executive processes: strategic vs. evaluative functions of the anterior cingulate cortex. *Proc. Natl. Acad. Sci. U.S.A.* 97, 1944–1948.

Cavanagh, J. F., and Allen, J. J. B. (2008). Multiple aspects of the stress response under social evaluative threat: an electrophysiological investigation. *Psychoneuroendocrinology* 33, 41–53.

Cavanagh, J. F., Grandler, T. O. J., Frank, M. J., and Allen, J. J. B. (2010). Altered cingulate sub-region activation accounts for task-related dissociation in ERN amplitude as a function of obsessive-compulsive symptoms. *Neuropsychologia* 48, 2098–2109.

Chein, J., Albert, D., O'Brien, L., Uckert, K., and Steinberg, L. (2011). Peers increase adolescent risk taking by enhancing activity in the brain's reward circuitry. *Dev. Sci.* 14, F1–F10.

Chiu, P. H., and Deldin, P. J. (2007). Neural evidence for enhanced error detection in major depressive disorder. *Am. J. Psychiatry* 164, 608–616.

Clayson, P. E., Clawson, A., and Larson, M. J. (2011). The effects of induced state negative affect on performance monitoring processes. *Soc. Cogn. Affect. Neurosci.* doi: 10.1093/scan/nsr040. [Epub ahead of print].

Coan, J. A., and Allen, J. J. B. (2003). Frontal EEG asymmetry and the behavioral activation and inhibition systems. *Psychophysiology* 40, 106–114.

Cohen, M. X. (2011). Error-related medial frontal theta activity predicts cingulate-related structural connectivity. *Neuroimage* 55, 1373–1383.

Cohen, M. X., and Ranganath, C. (2007). Reinforcement learning signals predict future decisions. *J. Neurosci.* 27, 371–378.

Cohen, M. X., Young, J., Baek, J. M., Kessler, C., and Ranganath, C. (2005). Individual differences in extraversion and dopamine genetics predict neural reward responses. *Brain Res. Cogn. Brain Res.* 25, 851–861.

Cook, S. C., and Wellman, C. L. (2004). Chronic stress alters dendritic morphology in rat medial prefrontal cortex. *J. Neurobiol.* 60, 236–248.

Davidson, R. J. (2004). Well-being and affective style: neural substrates and biobehavioural correlates. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 359, 1395–1411.

Davidson, R. J., and Irwin, W. (1999). The functional neuroanatomy of emotion and affective style. *Trends Cogn. Sci.* 3, 11–21.

de Bruijn, E. R. A., and Ullsperger, M. (2011). "Pathological changes in performance monitoring," in *Neural Basis of Motivational and Cognitive Control*, eds R. B. Mars, J. Sallet, M. F. S. Rushworth and N. Yeung (Cambridge, Massachusetts/London, England: The MIT Press), 263–280.

Dehaene, S., Posner, M. I., and Don, M. T. (1994). Localization of a neural system for error detection and compensation. *Psychol. Sci.* 5, 303–305.

Dias-Ferreira, E., Sousa, J. C., Melo, I., Morgado, P., Mesquita, A. R., Cerqueira, J. J., Costa, R. M., and Sousa, N. (2009). Chronic stress causes frontostriatal reorganization and affects decision-making. *Science* 325, 621–625.

Diekhof, E. K., Falkai, P., and Gruber, O. (2011). The orbitofrontal cortex and its role in the assignment of behavioural significance. *Neuropsychologia* 49, 984–991.

Drevets, W. C., Price, J. L., Simpson, J. R., Todd, R. D., Reich, T., Vannier, M., and Raichle, M. R. (1997). Subgenual prefrontal cortex abnormalities in mood disorders. *Nature* 386, 824–827.

Drevets, W. C., Savitz, J., and Trimble, M. (2008). The subgenual anterior cingulate cortex in mood disorders. *CNS Spectr.* 13, 663–681.

Eisenberger, N., Lieberman, M., and Satpute, A. (2005). Personality from a controlled processing perspective: an fMRI study of neuroticism, extraversion, and self-consciousness. *Cogn. Affect. Behav. Neurosci.* 5, 169–181.

Eisenberger, N. I., and Lieberman, M. D. (2004). Why rejection hurts: a common neural alarm system for physical and social pain. *Trends Cogn. Sci.* 8, 294–300.

Endrass, T., Klawohn, J., Schuster, F., and Kathmann, N. (2008). Overactive performance monitoring in obsessive-compulsive disorder: ERP evidence from correct and erroneous reactions. *Neuropsychologia* 46, 1877–1887.

Endrass, T., Schuermann, B., Kaufmann, C., Spielberg, R., Kniesche, R., and Kathmann, N. (2010). Performance monitoring and error significance in patients with obsessive-compulsive disorder. *Biol. Psychol.* 84, 257–263.

Etkin, A., Egner, T., and Kalisch, R. (2011). Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends Cogn. Sci.* 15, 85–93.

Falkenstein, M., Hohnsbein, J., Hoormann, J., and Blanke, L. (1991). Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. *Electroencephalogr. Clin. Neurophysiol.* 78, 447–455.

Fallgatter, A. J., Herrmann, M., Roemmler, J., Ehlis, A. C., Wagener, A., Heidrich, A., Ortega, G., Zeng, Y., and Lesch, K. P. (2004). Allelic variation of serotonin transporter function modulates the brain electrical response for error processing. *Neuropsychopharmacology* 29, 1506–1511.

Fitzgerald, K. D., Welsh, R. C., Gehring, W. J., Abelson, J. L., Himle, J. A., Liberzon, I., and Taylor, S. F. (2005). Error-related hyperactivity of the anterior cingulate cortex in obsessive-compulsive disorder. *Biol. Psychiatry* 57, 287–294.

Foti, D., and Hajcak, G. (2009). Depression and reduced sensitivity to non-rewards versus rewards: evidence from event-related potentials. *Biol. Psychol.* 81, 1–8.

Fukushima, H., and Hiraki, K. (2009). Whose loss is it? Human electrophysiological correlates of non-self reward processing. *Soc. Neurosci.* 4, 261–275.

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., and Donchin, E. (1993). A neural system for error-detection and compensation. *Psychol. Sci.* 4, 385–390.

Gehring, W. J., Himle, J., and Nisenson, L. G. (2000). Action-monitoring dysfunction in obsessive-compulsive disorder. *Psychol. Sci.* 11, 1–6.

Gehring, W. J., and Willoughby, A. R. (2002). The medial frontal cortex

and the rapid processing of monetary gains and losses. *Science* 295, 2279–2282.

Gehring, W. J., and Willoughby, A. R. (2004). "Are all medial frontal negativities created equal? Toward a richer empirical basis for theories of action monitoring," in *Errors, Conflicts, and the Brain. Current Opinions on Performance Monitoring*, ed M. U. M. Falkenstein (Leipzig, Germany: Max Planck Institute of Cognitive Neuroscience), 14–20.

Gentsch, A., Ullsperger, P., and Ullsperger, M. (2009). Dissociable medial frontal negativities from a common monitoring system for self- and externally caused failure of goal achievement. *Neuroimage* 47, 2023–2030.

Georgiadi, E., Liotti, M., Nixon, N. L., and Liddle, P. F. (2011). Electrophysiological evidence for abnormal error monitoring in recurrent major depressive disorder. *Psychophysiology* 48, 1192–1202.

Goldin, P. R., Manber, T., Hakimi, S., Canli, T., and Gross, J. J. (2009). Neural bases of social anxiety disorder: emotional reactivity and cognitive regulation during social and physical threat. *Arch. Gen. Psychiatry* 66, 170–180.

Gray, J. A. (1987). "The neuropsychology of emotion and personality," in *Cognitive Neurochemistry*, eds S. M. Stahl, S. D. Iverson, and E. C. Goodman (Oxford, UK: Oxford), 171–190.

Gray, J. R. (1989). "Fundamental systems of emotion in the mammalian brain," in *Coping with Uncertainty: Behavioral and Developmental Perspectives*, ed D. S. Palermo (Hillsdale, NJ: Lawrence Erlbaum), 173–195.

Gründler, T. O., Cavanagh, J. F., Figueroa, C. M., Frank, M. J., and Allen, J. J. (2009). Task-related dissociation in ERN amplitude as a function of obsessive-compulsive symptoms. *Neuropsychologia* 47, 1978–1987.

Gu, R., Huang, Y.-X., and Luo, Y.-J. (2010). Anxiety and feedback negativity. *Psychophysiology* 47, 961–967.

Hajcak, G., McDonald, N., and Simons, R. F. (2003). Anxiety and error-related brain activity. *Biol. Psychol.* 64, 77–90.

Hajcak, G., McDonald, N., and Simons, R. F. (2004). Error-related psychophysiology and negative affect. *Brain Cogn.* 56, 189–197.

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

Hajcak, G., and Olvet, D. M. (2008). The persistence of attention to emotion: brain potentials during and after picture presentation. *Emotion* 8, 250–255.

Hajcak, G., and Simons, R. F. (2002). Error-related brain activity in obsessive-compulsive undergraduates. *Psychiatry Res.* 110, 63–72.

Herrmann, M. J., Rommler, J., Ehlis, A. C., Heidrich, A., and Fallgatter, A. J. (2004). Source localization (LORETA) of the error-related-negativity (ERN/Ne) and positivity (Pe). *Brain Res. Cogn. Brain Res.* 20, 294–299.

Hewig, J., Kretschmer, N., Trippe, R. H., Hecht, H., Coles, M. G. H., Holryod, C. B., and Miltner, W. H. R. (2011). Why humans deviate from rational choice. *Psychophysiology* 48, 507–514.

Holmes, A. J., Bogdan, R., and Pizzagalli, D. A. (2010). Serotonin transporter genotype and action monitoring dysfunction: a possible substrate underlying increased vulnerability to depression. *Neuropsychopharmacology* 35, 1186–1197.

Holmes, A. J., and Pizzagalli, D. A. (2008). Spatiotemporal dynamics of error processing dysfunctions in major depressive disorder. *Arch. Gen. Psychiatry* 65, 179–188.

Holroyd, C. B. (2004). "A note on the oddball N200 and feedbak ERN," in *Errors, Conflicts, and the Brain: Current Opinions on Performance Monitoring*, ed M. U. M. Falkenstein (Leipzig, Germany: MPI of Cognitive Neuroscience), 211–218.

Holroyd, C. B., and Coles, M. G. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychol. Rev.* 109, 679–709.

Holroyd, C. B., and Coles, M. G. (2008). Dorsal anterior cingulate cortex integrates reinforcement history to guide voluntary behavior. *Cortex* 44, 548–559.

Holroyd, C. B., Dien, J., and Coles, M. G. (1998). Error-related scalp potentials elicited by hand and foot movements: evidence for an output-independent error-processing system in humans. *Neurosci. Lett.* 242, 65–68.

Holroyd, C. B., Hajcak, G., and Larsen, J. T. (2006). The good, the bad and the neutral: electrophysiological responses to feedback stimuli. *Brain Res.* 1105, 93–101.

Holroyd, C. B., Krigolson, O. E., Baker, R., Lee, S., and Gibson, J. (2009). When is an error not a prediction error? An electrophysiological investigation. *J. Cogn. Affect. Behav. Neurosci.* 9, 59–70.

Holroyd, C. B., and Yeung, N. (2012). Motivation of extended behaviors by anterior cingulate cortex. *Trends Cogn. Sci.* 16, 122–128.

Hopko, D. R., Lejuez, C. W., Daughters, S. B., Aklin, W. M., Osborne, A., Simmons, B. L., and Strong, D. R. (2006). Construct validity of the balloon analogue risk task (BART); relationship with mdma use by inner-city drug users in residential treatment. *J. Psychopathol. Behav. Assess.* 28, 95–101.

Inzlicht, M., McGregor, I., Hirsh, J. B., and Nash, K. (2009). Neural markers of religious conviction. *Psychol. Sci.* 20, 385–392.

Itagaki, S., and Katayama, J. (2008). Self-relevant criteria determine the evaluation of outcomes induced by others. *Neuroreport* 19, 383–387.

Jocham, G., and Ullsperger, M. (2009). Neuropharmacology of performance monitoring. *Neurosci. Biobehav. Rev.* 33, 48–60.

Jonkman, L. M., Sniedt, F. L., and Kemner, C. (2007). Source localization of the Nogo-N2, a developmental study. *Clin. Neurophysiol.* 118, 1069–1077.

Kamarajan, C., Porjesz, B., Rangaswamy, M., Tang, Y., Chorlian, D. B., Padmanabhapillai, A., Saunders, R., Pandey, A. K., Roopesh, B. N., Manz, N., Stimus, A. T., and Begleiter, H. (2009). Brain signatures of monetary loss and gain: outcome-related potentials in a single outcome gambling task. *Behav. Brain Res.* 197, 62–76.

Kanai, R., Feilden, T., Firth, C., and Rees, G. (2011). Political orientations are correlated with brain structure in young adults. *Curr. Biol.* 21, 677–680.

Kang, S. K., Hirsh, J. B., and Chasteen, A. L. (2010). Your mistakes are mine: self-other overlap predicts neural response to observed errors. *J. Exp. Soc. Psychol.* 46, 229–232.

Kanske, P., and Kotz, S. A. (2011). Emotion speeds up conflict resolution: a new role for the ventral anterior cingulate cortex? *Cereb. Cortex* 21, 911–919.

Karg, K., Burmeister, M., Shedden, K., and Sen, S. (2011). The serotonin transporter promoter variant (5-httlpr), stress, and depression meta-analysis revisited: evidence of genetic moderation. *Arch. Gen. Psychiatry* 68, 444–454.

Kiehl, K. A., Liddle, P. F., and Hopfinger, J. B. (2000). Error processing and the rostral anterior cingulate: an event-related fMRI study. *Psychophysiology* 37, 216–223.

Kim, M. J., Gee, D. G., Loucks, R. A., Davis, F. C., and Whalen, P. J. (2011). Anxiety dissociates dorsal and ventral medial prefrontal cortex functional connectivity with the amygdala at rest. *Cereb. Cortex* 21, 1667–1673.

Krämer, U. M., Cunillera, T., Càmara, E., Marco-Pallarés, J., Cucurell, D., Nager, W., Bauer, P., Schüle, R., Schöls, L., Rodriguez-Fornells, A., and Münte, T. F. (2007). The impact of catechol-o-methyltransferase and dopamine D4 receptor genotypes on neurophysiological markers of performance monitoring. *J. Neurosci.* 27, 14190–14198.

Lamm, C., Granic, I., Zelazo, P. D., and Lewis, M. D. (2011). Magnitude and chronometry of neural mechanisms of emotion regulation in subtypes of aggressive children. *Brain Cogn.* 77, 159–169.

Larson, M. J., Fair, J. E., Good, D. A., and Baldwin, S. A. (2010). Empathy and error processing. *Psychophysiology* 47, 415–424.

Lee, S. H., and Park, G. H. (2011). "Psychophysiological markers of anxiety disorders and anxiety symptoms," in *Anxiety Disorders*, ed V. V. Kalinin (Rijeka, Croatia: InTech).

Lejuez, C. W., Aklin, W. M., Jones, H. A., Richards, J. B., Strong, D. R., Kahler, C. W., and Read, J. P. (2003a). The Balloon Analogue Risk Task (BART) differentiates smokers and nonsmokers. *Exp. Clin. Psychopharmacol.* 11, 26–33.

Lejuez, C. W., Aklin, W. M., Zvolensky, M. J., and Pedulla, C. M. (2003b). Evaluation of the Balloon Analogue Risk Task (BART) as a predictor of adolescent real-world risk-taking behaviours. *J. Adolesc.* 26, 475–479.

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., Strong, D. R., and Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: the Balloon Analogue Risk Task (BART). *J. Exp. Psychol. Appl.* 8, 75–84.

Lewis, M. D., Lamm, C., Segalowitz, S. J., Stieben, J., and Zelazo, P. D. (2006). Neurophysiological correlates of emotion regulation in children and adolescents. *J. Cogn. Neurosci.* 18, 430–443.

Li, P., Han, C., Lei, Y., Holroyd, C. B., and Li, H. (2011). Responsibility modulates neural mechanisms of outcome processing: an ERP study. *Psychophysiology* 48, 1129–1133.

Li, P., Jia, S., Feng, T., Liu, Q., Suo, T., and Li, H. (2010). The influence of the diffusion of responsibility effect

on outcome evaluations: electrophysiological evidence from an ERP study. *Neuroimage* 52, 1727–1733.

Liston, C., McEwen, B. S., and Casey, B. J. (2009). Psychosocial stress reversibly disrupts prefrontal processing and attentional control. *Proc. Natl. Acad. Sci. U.S.A.* 106, 912–917.

Luck, S. J. (2005). *An Introduction to the Event-Related Potential Technique.* Cambridge, MA: MIT Press.

Luu, P., Collins, P., and Tucker, D. M. (2000). Mood, personality, and self-monitoring: negative affect and emotionality in relation to frontal lobe mechanisms of error monitoring. *J. Exp. Psychol. Gen.* 129, 43–60.

Luu, P., and Posner, M. I. (2003). Anterior cingulate cortex regulation of sympathetic activity. *Brain* 126, 2119–2120.

Luu, P., Shane, M., Pratt, N., and Tucker, D. (2009). Corticolimbic mechanisms in the control of trial and error learning. *Brain Res.* 1247, 100–113.

Luu, P., and Tucker, D. M. (2003). "Self-regulation and the executive functions: electrophysiological cues," in *The Cognitive Electrophysiology of Mind and Brain*, eds A. Zani and A. M. Proverbio (San Diego, CA: Academic), 199–223.

Luu, P., Tucker, D. M., Derryberry, D., Reed, M., and Poulsen, C. (2003). Electrophysiological responses to errors and feedback in the process of action regulation. *Psychol. Sci.* 14, 47–53.

Luu, P., Tucker, D. M., and Stripling, R. (2007). Neural mechanisms for learning actions in context. *Brain Res.* 1179, 89–105.

Makeig, S., Bell, A., Jung, T.-P., and Sejnowski, T. (1996). "Independent component analysis of electroencephalographic data," in *Advances in Neural Information Processing Systems*, Vol. 8, eds M. Mozer, D. Touretzky, and M Hasselmo (Cambridge, MA: MIT Press), 145–151.

Marco-Pallarés, J., Krämer, U., Strehl, S., Schröder, A., and Münte, T. (2010). When decisions of others matter to me: an electrophysiological analysis. *BMC Neurosci.* 11, 1–8.

Masaki, H., and Segalowitz, S. J. (2004). "Error negativity: a test of the response conflict versus error detection hypotheses," in *Errors, Conflicts, and the Brain Current Opinions on Performance Monitoring*, eds M. Ullsperger and M. Falkenstein (Leipzig, Germany: MPI), 76–83.

Masaki, H., Tanaka, H., Takasawa, N., and Yamazaki, K. (2001). Error-related brain potentials elicited by vocal errors. *Neuroreport* 12, 1851–1855.

Mathalon, D. H., Whitfield, S. L., and Ford, J. M. (2003). Anatomy of an error: ERP and fMRI. *Biol. Psychol.* 64, 119–141.

Mayberg, H. S. (1997). "Limbic-cortical dysregulation: a proposed model of depression," in *The Neuropsychiatry of Limbic and Subcortical Disorders*, eds S. Salloway, P. Malloy, and L. J. Cummings (Washington, DC: American Psychiatric Press Inc).

Mayberg, H. S., Brannan, S. K., Mahurin, R. K., Jerabek, P. A., Brickman, J. S., Tekell, J. L., Silva, J. A., McGinnis, S., Glass, T. G., Martin, C. C., and Fox, P. T. (1997). Cingulate function in depression: a potential predictor of treatment response. *Neuroreport* 8, 1057–1061.

Menon, V., Adleman, N. E., White, C. D., Glover, G. H., and Reiss, A. L. (2001). Error-related brain activation during a Go/NoGo response inhibition task. *Hum. Brain Mapp.* 12, 131–143.

Mies, G. W., van der Veen, F. M., Tulen, J. H., Birkenhäger, T. K., Hengeveld, M. W., and van der Molen, M. W. (2011). Drug-free patients with major depression show an increased electrophysiological response to valid and invalid feedback. *Psychol. Med.* 1–11 [Epub ahead of print].

Milham, M., and Banich, M. (2005). Anterior cingulate cortex: an fMRI analysis of conflict specificity and functional differentiation. *Hum. Brain Mapp.* 25, 328–335.

Miltner, W. H. R., Brauer, J., Hecht, H., Trippe, R., and Coles, M. G. H. (2004). "Parallel brain activity for self-generated and observed errors," in *Errors, Conflicts, and the Brain. Current Opinions on Performance Monitoring*, eds M. Ullsperger and M. Falkenstein (Leipzig, Germany: Max Planck Institute for Human Cognitive and Brain Sciences), 228.

Miltner, W. H. R., Braun, C. H., and Coles, M. G. H. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task: evidence for a "generic" neural system for error detection. *J. Cogn. Neurosci.* 9, 788–798.

Mitra, R., Jadhav, S., McEwen, B. S., Vyas, A., and Chattarji, S. (2005). Stress duration modulates the spatiotemporal patterns of spine formation in the basolateral amygdala. *Proc. Natl. Acad. Sci. U.S.A.* 102, 9371–9376.

Moadab, I., Gilbert, T., Dishion, T. J., and Tucker, D. M. (2010). Frontolimbic activity in a frustrating task: covariation between patterns of coping and individual differences in externalizing and internalizing symptoms. *Dev. Psychopathol.* 22, 391–404.

Mueller, E. M., Makeig, S., Stemmler, G., Hennig, J., and Wacker, J. (2011). Dopamine effects on human error processing depend on catechol-O-methyltransferase VAL158MET genotype. *J. Neurosci.* 31, 15818–15825.

Muller, S. V., Moller, J., Rodriguez-Fornells, A., and Munte, T. F. (2005). Brain potentials related to self-generated and external information used for performance monitoring. *Clin. Neurophysiol.* 116, 63–74.

Munro, G., Dywan, J., Harris, G. T., McKee, S., Unsal, A., and Segalowitz, S. J. (2007a). ERN varies with degree of psychopathy in an emotion discrimination task. *Biol. Psychol.* 76, 31–42.

Munro, G., Dywan, J., Harris, G. T., McKee, S., Unsal, A., and Segalowitz, S. J. (2007b). Response inhibition in psychopathy: the frontal N2 and P3. *Neurosci. Lett.* 418, 149–153.

Murphy, T., Richard, M., Masaki, H., and Segalowitz, S. J. (2006). The effect of sleepiness on performance monitoring: I know what I am doing, but do I care? *J. Sleep Res.* 15, 15–21.

Newman-Norlund, R. D., Ganesh, S., van Schie, H. T., De Bruijn, E. R., and Bekkering, H. (2009). Self-identification and empathy modulate error-related brain activity during the observation of penalty shots between friend and foe. *Soc. Cogn. Affect. Neurosci.* 4, 10–22.

Nieuwenhuis, S., Holroyd, C. B., Mol, N., Coles, M. G. (2004). Reinforcement-related brain potentials from medial frontal cortex: origins and functional significance. *Neurosci. Biobehav. Rev.* 28, 441–448.

Nieuwenhuis, S., Slagter, H. A., von Geusau, N. J., Heslenfeld, D. J., and Holroyd, C. B. (2005). Knowing good from bad: differential activation of human cortical areas by positive and negative outcomes. *Eur. J. Neurosci.* 21, 3161–3168.

Nieuwenhuis, S., Yeung, N., van den Wildenberg, W., and Ridderinkhof, K. R. (2003). Electrophysiological correlates of anterior cingulate function in a go/no-go task: effects of response conflict and trial type frequency. *Cogn. Affect. Behav. Neurosci.* 3, 17–26.

O'Connell, R. G., Dockree, P. M., Bellgrove, M. A., Kelly, S. P., Hester, R., Garavan, H., Robertson, I. H., and Foxe, J. J. (2007). The role of cingulate cortex in the detection of errors with and without awareness: a high-density electrical mapping study. *Eur. J. Neurosci.* 25, 2571–2579.

Ogawa, K., Masaki, H., Yamazaki, K., and Sommer, W. (2011). The influence of emotions due to verbal admonishment and encouragement on performance monitoring. *Neuroreport* 22, 313–318.

Oliveira, F. T., McDonald, J. J., and Goodman, D. (2007). Performance monitoring in the anterior cingulate is not all error related: expectancy deviation and the representation of action-outcome associations. *J. Cogn. Neurosci.* 19, 1994–2004.

Olvet, D. M., and Hajcak, G. (2008). The error-related negativity (ERN) and psychopathology: toward an endophenotype. *Clin. Psychol. Rev.* 28, 1343–1354.

Olvet, D. M., and Hajcak, G. (2012). The error-related negativity relates to sadness following mood induction among individuals with high neuroticism. *Soc. Cogn. Affect. Neurosci.* 7, 289–295.

Olvet, D. M., Klein, D. N., and Hajcak, G. (2010). Depression symptom severity and error-related brain activity. *Psychiatry Res.* 179, 30–37.

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

Palomero-Gallagher, N., Bidmon, H. J., Cremer, M., Schleicher, A., Kircheis, G., Reifenberger, G., Kostopoulos, G., Häussinger, D., and Zilles, K. (2009). Neurotransmitter receptor imbalances in motor cortex and basal ganglia in hepatic encephalopathy. *Cell. Physiol. Biochem.* 24, 291–306.

Pascual-Marqui, R. D. (2002). Standardized low resolution brain electromagnetic tomography (sLORETA): technical details. *Methods Find. Exp. Clin. Pharmacol.* 24(Suppl. D), 5–12.

Pfabigan, D. M., Alexopoulos, J., Bauer, H., Lamm, C., and Sailer, U. (2011). All about the money? External performance monitoring is affected by monetary, but not by socially conveyed feedback cues in more antisocial individuals. [Original Research]. *Front. Hum.*

Neurosci. 5:100. doi: 10.3389/fnhum.2011.00100

Pizzagalli, D., Pascual-Marqui, R. D., Nitschke, J. B., Oakes, T. R., Larson, C. L., Abercrombie, H. C., Schaefer, S. M., Koger, J. V., Benca, R. M., and Davidson, R. J. (2001). Anterior cingulate activity as a predictor of degree of treatment response in major depression: evidence from brain electrical tomography analysis. *Am. J. Psychiatry* 158, 405–415.

Pizzagalli, D. A. (2011). frontocingulate dysfunction in depression: toward biomarkers of treatment response. *Neuropsychopharmacology* 36, 183–206.

Pizzgalli, D. A., Oakes, T. R., Fox, A. S., Chung, M. K., Larson, C. L., Abercrombie, H. C., Chaefer, S. M., Benca, R. M., and Davidson, R. J. (2004). Functional but not structural subgenual prefrontal cortex abnormalities in melancholia. *Mol. Psychiatry* 9, 325–325.

Polezzi, D., Sartori, G., Rumiati, R., Vidotto, G., and Daum, I. (2010). Brain correlates of risky decision-making. *Neuroimage* 49, 1886–1894.

Polli, F. E., Barton, J. J. S., Cain, M. S., Thakkar, K. N., Rauch, S. L., and Manoach, D. S. (2005). Rostral and dorsal anterior cingulate cortex make dissociable contributions during antisaccade error commission. *Proc. Natl. Acad. Sci. U.S.A.* 102, 15700–15705.

Potts, G. F. (2011). Impact of reward and punishment motivation on behavior monitoring as indexed by the error-related negativity. *Int. J. Psychophysiol.* 81, 324–331.

Pourtois, G., Vocat, R., N'Diaye, K., Spinelli, L., Seeck, M., and Vuilleumier, P. (2010). Errors recruit both cognitive and emotional monitoring systems: simultaneous intracranial recordings in the dorsal anterior cingulate gyrus and amygdala combined with fMRI. *Neuropsychologia* 48, 1144–1159.

Radley, J. J., Arias, C. M., and Sawchenko, P. E. (2006a). Regional differentiation of the medial prefrontal cortex in regulating adaptive responses to acute emotional stress. *J. Neurosci.* 26, 12967–12976.

Radley, J. J., Rocher, A. B., Miller, M., Janssen, W. G. M., Liston, C., Hof, P. R., McEwen, B. S., and Morrison, J. H. (2006b). Repeated stress induces dendritic spine loss in the rat medial prefrontal cortex. *Cereb. Cortex* 16, 313–320.

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

Ridderinkhof, K. R., van den Wildenberg, W. P., Segalowitz, S. J., and Carter, C. S. (2004b). Neurocognitive mechanisms of cognitive control: the role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. *Brain Cogn.* 56, 129–140.

Riesel, A., Weinberg, A., Endrass, T., Kathmann, N., and Hajcak, G. (2012). Punishment has a lasting impact on error-related brain activity. *Psychophysiology* 49, 239–247.

Robinson, M. D., Moeller, S. K., and Fetterman, A. K. (2010). Neuroticism and responsiveness to error feedback: adaptive self-regulation versus affective reactivity. *J. Pers.* 78, 1469–1496.

Ruchsow, M., Grön, G., Reuter, K., Spitzer, M., Hermle, L., and Kiefer, M. (2005). Error-related brain activity in patients with obsessive-compulsive disorder and in healthy controls. *J. Psychophysiol.* 19, 298–304.

Ruchsow, M., Herrnberger, B., Wiesend, C., Grön, G., Spitzer, M., and Kiefer, M. (2004). The effect of erroneous responses on response monitoring in patients with major depressive disorder: a study with event-related potentials. *Psychophysiology* 41, 833–840.

Sallet, J., Mars, R., Quilodran, R., Procyk, E., Petrides, M., and Rushworth, M. (2011). "Neuroanatomical basis of motivational and cognitive control: a focus on the medial and lateral prefrontal cortex," in *Neural Basis of Motivational and Cognitive Control*, eds J. Sallet, R. B. Mars, N. Yeung, and M. F. Rushworth (Cambridge, MA: MIT), 5–20.

Santesso, D. L., Bogdan, R., Birk, J. L., Goetz, E. L., Holmes, A. J., and Pizzagalli, D. A. (2011). Neural responses to negative feedback are related to negative emotionality in healthy adults. *Soc. Cogn. Affect. Neurosci.* doi: 10.1093/scan/nsr054. [Epub ahead of print].

Santesso, D. L., and Segalowitz, S. J. (2009). The error-related negativity is related to risk taking and empathy in young men. *Psychophysiology* 46, 143–152.

Santesso, D. L., Segalowitz, S. J., and Schmidt, L. A. (2005). ERP correlates of error monitoring in 10-year olds are related to socialization. *Biol. Psychol.* 70, 79–87.

Santesso, D. L., Segalowitz, S. J., and Schmidt, L. A. (2006). Error-related electrocortical responses are enhanced in children with obsessive-compulsive behaviors. *Dev. Neuropsychol.* 29, 431–445.

Santesso, D. L., Steele, K. T., Bogdan, R., Holmes, A. J., Deveney, C. M., Meites, T. M., and Pizzgalli, D. A. (2008). Enhanced negative feedback responses in remitted depression. *Neuroreport* 19, 1045–1048.

Sato, A., Yasuda, A., Ohira, H., Miyawaki, K., Nishikawa, M., Kumano, H., and Kuboki, T. (2005). Effects of value and reward magnitude on feedback negativity and P300. *Neuroreport* 16, 407–411.

Schnider, A., Treyer, V., and Buck, A. (2005). The human orbitofrontal cortex monitors outcomes even when no reward is at stake. *Neuropsychologia* 43, 316–323.

Schrijvers, D., De Bruijn, E. R., Maas, Y. J., Vancoillie, P., Hulstijn, W., and Sabbe, B. G. (2009). Action monitoring and depressive symptom reduction in major depressive disorder. *Int. J. Psychophysiol.* 71, 218–224.

Schrijvers, D., de Bruijn, E. R. A., Maas, Y., De Grave, C., Sabbe, B. G. C., and Hulstijn, W. (2008). Action monitoring in major depressive disorder with psychomotor retardation. *Cortex* 44, 569–579.

Segalowitz, S. J., Santesso, D. L., Murphy, T. I., Homan, D., Chantziantoniou, D. K., and Khan, S. (2010). Retest reliability of medial frontal negativities during performance monitoring. *Psychophysiology* 47, 260–270.

Segalowitz, S. J., Santesso, D. L., Willoughby, T., Reker, D. L., Campbell, K., Chalmers, H., and Rose-Krasnor, L. (2012). Adolescent peer interaction and trait surgency weaken medial prefrontal cortex responses to failure. *Soc. Cogn. Affect. Neurosci.* 7, 115–124.

Shackman, A. J., Salomons, T. V., Slagter, H. A., Fox, A. S., Winter, J. J., and Davidson, R. J. (2011). The integration of negative affect, pain and cognitive control in the cingulate cortex. *Nat. Rev. Neurosci.* 12, 154–167.

Simões-Franklin, C., Hester, R., Shpaner, M., Foxe, J. J., and Garavan, H. (2010). Executive function and error detection: the effect of motivation on cingulate and ventral striatum activity. *Hum. Brain Mapp.* 31, 458–469.

Smillie, L. D., Cooper, A. J., and Pickering, A. D. (2010). Individual differences in reward prediction error: extraversion and feedback-related negativity. *Soc. Cogn. Affect. Neurosci.* 6, 646–652.

Stemmer, B., Segalowitz, S. J., Witzke, W., and Schönle, P. W. (2003). Error detection in patients with lesions to the medial prefrontal cortex: an ERP study. *Neuropsychologia* 42, 118–130.

Stieben, J., Lewis, M. D., Granic, I., Zelazo, P. D., Segalowitz, S. J., and Pepler, D. (2007). Neurophysiological mechanisms of emotion regulation for subtypes of externalizing children. *Dev. Psychopathol.* 19, 455–480.

Swick, D., and Turken, A. U. (2004). "Focusing on the anterior cingulate cortex: effects of focal lesions on cognitive performance," in *Cognitive Neuroscience of Attention*, ed M. I. Posner (New York, NY: Guilford Press), 393–406.

Taylor, S. F., Martis, B., Fitzgerald, K. D., Welsh, R. C., Abelson, J. L., Liberzon, I., Himle, J. A., and Gehring, W. J. (2006). Medial frontal cortex activity and loss-related responses to errors. *J. Neurosci.* 26, 4063–4070.

Taylor, S. F., Stern, E. R., and Gehring, W. J. (2007). Neural systems for error monitoring: Recent findings and theoretical perspectives. *Neuroscientist* 13, 160–172.

Tops, M., and Boksem, M. A. (2010). Absorbed in the task: Personality measures predict engagement during task performance as tracked by error negativity and asymmetrical frontal activity. *Cogn. Affect. Behav. Neurosci.* 10, 441–453.

Tops, M., and Boksem, M. A. (2011). Cortisol involvement in mechanisms of behavioral inhibition. *Psychophysiology* 48, 723–732.

Tops, M., Boksem, M. A., Wester, A. E., Lorist, M. M., and Meijman, T. F. (2006). Task engagement and the relationships between the error-related negativity, agreeableness, behavioral shame proneness and cortisol. *Psychoneuroendocrinology* 31, 847–858.

Ullsperger, M. (2009). Genetic association studies of performance monitoring and learning from feedback: the role of dopamine and serotonin. *Neurosci. Biobehav. Rev.* 34, 649–659.

Ullsperger, M. (2011). "Neurochemistry of performance monitoring," in *Neural Basis of Motivational and Cognitive Control*, eds J. Sallet, R. B. Mars, M. F. S. Rushworth, and N. Yeung (Cambridge, MA: MIT), 37–49.

Ullsperger, M., and von Cramon, D. Y. (2001). Subprocesses of

performance monitoring: a dissociation of error processing and response competition revealed by event-related fMRI and ERPs. *Neuroimage* 14, 1387–1401.

Ursu, S., Stenger, V. A., Shear, M. K., Jones, M. R., and Carter, C. S. (2003). Overactive action monitoring in obsessive-compulsive disorder: evidence from functional magnetic resonance imaging. *Psychol. Sci.* 14, 347–353.

van der Helden, J., Boksem, M. A. S., and Blom, J. H. G. (2010). The importance of failure: feedback-related negativity predicts motor learning efficiency. *Cereb. Cortex* 20, 1596–1603.

Van der Veen, F. M., and Sahibdin, P. P. (2011). Dissociation between medial frontal negativity and cardiac responses in the ultimatum game: effects of offer size and fairness. *Cogn. Affect. Behav. Neurosci.* 11, 516–525.

van Noordt, S. (2012). *Hot Cognition and Activation of the Medial Prefrontal Cortex: Self-Regulating in Contexts Involving Motivational Pressures,* (St. Catharines, ON: Brock University).

Van 't Ent, D., and Apkarian, P. (1999). Motoric response inhibition in finger movement and saccadic eye movement: a comparative study. *Clin. Neurophysiol.* 110, 1058–1072.

van Veen, V., and Carter, C. S. (2002a). The anterior cingulate as a conflict monitor: fMRI and ERP studies. *Physiol. Behav.* 77, 477–482.

van Veen, V., and Carter, C. S. (2002b). The timing of action-monitoring processes in the anterior cingulate cortex. *J. Cogn. Neurosci.* 14, 593–602.

van Veen, V., Cohen, J., Botvinick, M., Stenger, V., and Carter, C. (2001). Anterior cingulate cortex, conflict monitoring, and levels of processing. *Neuroimage* 14, 1302–1308.

Vidal, F., Hasbroucq, T., Grapperon, J., and Bonnet, M. (2000). Is the 'error negativity' specific to errors? *Biol. Psychol.* 51, 109–128.

Vogt, B. A. (2009). "Regions and subregions of the cingulate cortex," in *Cingulate Neurobiology and Disease*, ed B. A. Vogt (New York, NY: Oxford University Press).

Vyas, A., Mitra, R., Shankaranarayana Rao, B. S., and Chattarji, S. (2002). Chronic stress induces contrasting patterns of dendritic remodeling in hippocampal and amygdaloid neurons. *J. Neurosci.* 22, 6810–6818.

Weinberg, A., Olvet, D. M., and Hajcak, G. (2010). Increased error-related brain activity in generalized anxiety disorder. *Biol. Psychol.* 85, 472–480.

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

Weissflog, M., van Noordt, S., Choma, B., Dywan, J., and Segalowitz, S. J. (2010). Machiavellianism, sociopolitical ideologies and the ERN. *Paper presented at the 50th Annual Society for Psychophysiological Research Meeting*. (Portland, OR)

Whittle, S., Allen, N. B., Fornito, A., Lubman, D. I., Simmons, J. G., Pantelis, C., and Yücel, M. (2009a). Variations in cortical folding patterns are related to individual differences in temperament. *Psychiatry Res.* 172, 68–74.

Whittle, S., Chanen, A. M., Fornito, A., McGorry, P. D., Pantelis, C., and Yücel, M. (2009b). Anterior cingulate volume in adolescents with first-presentation borderline personality disorder. *Psychiatry Res.* 172, 155–160.

Whittle, S., Yap, M. B. H., Yücel, M., Fortino, A., Simmons, J. G., Barrett, A., Sheeber, L., and Allen, N. B. (2008a). Prefrontal and amygdala volumes are related to adolescents' affective behaviors during parent-adolescent interactions. *Proc. Natl. Acad. Sci. U.S.A.* 105, 3652–3657.

Whittle, S., Yücel, M., Fortino, A., Barrett, A., Wood, S. J., Lubman, D. I., Simmons, J., Pantelis, C., and Allen, N. B. (2008b). Neuroanatomical correlates of temperament in early adolescents. *J. Am. Acad. Child Adolesc. Psychiatry* 47, 682–693.

Wiswede, D., Münte, T. F., and Rüsseler, J. (2009). Negative affect induced by derogatory verbal feedback modulates the neural signature of error detection. *Soc. Cogn. Affect. Neurosci.* 4, 227–237.

Wu, Y., Leliveld, M. C., and Zhou, X. (2011). Social distance modulates recipient's fairness consideration in the dictator game: an ERP study. *Biol. Psychol.* 88, 253–262.

Wu, Y., Zhou, Y., van Dijk, E., Leliveld, M. C., and Zhou, X. (2011). Social comparison affects brain responses to fairness in asset division: an ERP study with the ultimatum game. [Original Research]. *Front. Hum. Neurosci.* 5:131. doi: 10.3389/fnhum.2011.00131

Xiao, Z., Wang, J., Zhang, M., Li, H., Tang, Y., Wang, Y., Fan, Q., and Fromson, J. A. (2011). Error-related negativity abnormalities in generalized anxiety disorder and obsessive - compulsive disorder. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 35, 265–272.

Yasuda, A., Sato, A., Miyawaki, K., Kumano, H., and Kuboki, T. (2004). Error-related negativity reflects detection of negative reward prediction error. *Neuroreport* 15, 2561–2565.

Yeung, N., Holroyd, C. B., and Cohen, J. D. (2005). ERP correlates of feedback and reward processing in the presence and absence of response choice. *Cereb. Cortex* 15, 535–544.

Yoshimura, S., Okamoto, Y., Onoda, K., Matsunaga, M., Ueda, K., Suzuki, S., and Yamawaki, S. (2010). Rostral anterior cingulate cortex activity mediates the relationship between the depressive symptoms and the medial prefrontal cortex activity. *J. Affect. Disord.* 122, 76–85.

Yu, R., and Zhou, X. (2006). Brain responses to outcomes of one's own and other's performance in a gambling task. *Neuroreport* 17, 1747–1751.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 09 March 2012; paper pending published: 31 March 2012; accepted: 17 June 2012; published online: 11 July 2012.

*Citation: van Noordt SJR and Segalowitz SJ (2012) Performance monitoring and the medial prefrontal cortex: a review of individual differences and context effects as a window on self-regulation. Front. Hum. Neurosci. **6**:197. doi: 10.3389/fnhum.2012.00197*

frontiers in
HUMAN NEUROSCIENCE

REVIEW ARTICLE
published: 24 May 2012
doi: 10.3389/fnhum.2012.00147

The role of prediction in social neuroscience

***Elliot C. Brown*[1,2]* *and Martin Brüne*[1,2]**

[1] *Research Department of Cognitive Neuropsychiatry and Preventative Medicine, LWL University Hospital Bochum, Bochum, Germany*
[2] *International Graduate School of Neuroscience, Ruhr University Bochum, Bochum, Germany*

Edited by:
Christian Bellebaum, Ruhr University Bochum, Germany

Reviewed by:
Xiaolin Zhou, Peking University, China
Inti Brazil, Donders Institute for Brain Cognition and Behaviour, Netherlands

***Correspondence:**
Elliot C. Brown, Research Department of Cognitive Neuropsychiatry and Preventative Medicine, LWL University Hospital of Ruhr University Bochum, Alexandrinenstrasse 1-3, Bochum 44791, Germany.
e-mail: elliot.c.brown@gmail.com

Research has shown that the brain is constantly making predictions about future events. Theories of prediction in perception, action and learning suggest that the brain serves to reduce the discrepancies between expectation and actual experience, i.e., by reducing the prediction error. Forward models of action and perception propose the generation of a predictive internal representation of the expected sensory outcome, which is matched to the actual sensory feedback. Shared neural representations have been found when experiencing one's own and observing other's actions, rewards, errors, and emotions such as fear and pain. These general principles of the "predictive brain" are well established and have already begun to be applied to social aspects of cognition. The application and relevance of these predictive principles to social cognition are discussed in this article. Evidence is presented to argue that simple non-social cognitive processes can be extended to explain complex cognitive processes required for social interaction, with common neural activity seen for both social and non-social cognitions. A number of studies are included which demonstrate that bottom-up sensory input and top-down expectancies can be modulated by social information. The concept of competing social forward models and a partially distinct category of social prediction errors are introduced. The evolutionary implications of a "social predictive brain" are also mentioned, along with the implications on psychopathology. The review presents a number of testable hypotheses and novel comparisons that aim to stimulate further discussion and integration between currently disparate fields of research, with regard to computational models, behavioral and neurophysiological data. This promotes a relatively new platform for inquiry in social neuroscience with implications in social learning, theory of mind, empathy, the evolution of the social brain, and potential strategies for treating social cognitive deficits.

Keywords: predictive coding, social interaction, forward models, prediction error, sensorimotor control, social learning, imitation, social decision-making

INTRODUCTION

> "*Engage people with what they expect; it is what they are able to discern and confirms their projections. It settles them into predictable patterns of response, occupying their minds while you wait for the extraordinary moment—that which they cannot anticipate.*"
>
> Sun Tzu, The Art of War.

It has long been known by military strategists, psychologists, and neuroscientists that surprise and uncertainty can occur at a high cost. We live in an uncertain world full of ambiguous stimuli and events of which we are not sure of. Preparedness promotes efficiency, and this is true not only of behavior, but is also reflected in representations of an optimized neural system. For the brain to be energetically efficient and for our behavior to be optimal and adaptive, we utilize knowledge from our previous experiences to make predictions about the future and minimize the cost of surprise (Friston et al., 2006). In the animal kingdom, such previous experiences are often pre-programmed or innate through the forces of natural and sexual selection, that is, the flexibility of the prediction process is small. In "higher" mammals such as primates, pre-programmed patterns also exist (Tinbergen, 1951). However, due to the complexity and variability of the environment, the prediction mechanisms are much more plastic, and the programmes more open to personal (ontogenetically acquired) experience (Mayr, 1974). The brain's attempts to minimize the discrepancy between expectations or predictions and actual experience, addresses the problem of uncertainty and optimality, while also providing a common fundamental principle for processing incoming sensory information in our environment (Friston, 2005). The ideas just expressed here illustrate a generalized description of the role of prediction in cognition and the brain, and is becoming the consensus on a general and universal principle of how the brain works. Expectation, prediction, inference, anticipation, foresight, prospection, forecasting, and preparation are all terms that have been used to refer to different types of predictive processes that occur in the brain, in cognition and are evident in overt behavior. Predictive processing can refer to any psychological or neural process that utilizes estimations about the future. The proposal of the "predictive brain" broadly states that we are constantly generating mental representations to predict future states of the world around us, and about our own future internal mental state (e.g., Bar, 2007). These predictions of the future can include short-term estimations about upcoming

events in one's current situation, and also long-term prospections about the likelihood of events occurring in the distant future, outside of our currently situated environment. It is thought that these predictive internal representations of the future are constantly compared with the actual perceived outcome of internal mental and external events. We must be able to process our errors to learn from our mistakes, and consequently update internal representations of the predicted future. This allows us to learn from our previous experiences. Many authors have implied that predictive and inferential processes underlie a wide range of cognitive processes, including, most prominently, motor control (Wolpert and Miall, 1996), perceptual inference (Friston and Kiebel, 2009) and reward-based associative learning (Schultz and Dickinson, 2000).

Social neuroscience seeks to find the underlying neural mechanisms responsible for social behavior. To successfully navigate ourselves through the social world, we must be able to understand socially relevant stimuli, make interpretations about their meaning and behave according to decisions that are accurate, optimal, and adaptive. This includes understanding others' minds, and their intentions and beliefs. Inference and estimation is essential to social understanding, particularly as social scenarios are riddled with complex circumstances filled with ambiguity and uncertainty. The central role of predictive mechanisms in motor control, perception, and learning are clearly evident and well-documented, though a fast-growing body of experimental and theoretical work is providing increasing evidence of how such predictive mechanisms are also embedded in social cognitive processes. Shared neural representations of one's own experience and the experience of others' lies at the heart of this work (Decety and Sommerville, 2003). Similar neural activity has been revealed in the observer when both experiencing and observing actions (Gallese et al., 1996), rewards (Marco-Pallares et al., 2010), and emotions such as fear (Olsson et al., 2007) and pain (Cheng et al., 2008). The interpersonal connection of shared experience, as also represented by shared neural representations, is likely to form the basis for high-level social cognitions such as empathy (Decety and Ickes, 2009). Although it is still not as clear as to what degree predictive mechanisms may underlie social cognitions and how predictive mechanisms drive processing during social perception, social understanding, social interaction, and social learning. Numerous conceptualizations of predictive mechanisms have been made to account for different psychological and neural phenomena, each with their own specification of what prediction is, and how it is generated and utilized. The increasing popularity and support for the concept of a "predictive brain" calls for further discussion about whether this framework can be extended to other domains such as social cognition, and especially where conceptual boundaries between specifications may lie.

Given the broad scope of the implications of prediction in fundamental cognitive and neural processes, and in light of the evidence that in primates the complexity of the social environment has been a major driving force of cognitive evolution, it follows that predictive mechanisms are likely to both underlie and modulate processes involved in social cognition and social interaction. The aim of this article is to highlight recent theoretical and empirical work that is at the interface of the "predictive brain" and the "social brain." Our article is intended to propose extensions of predictive mechanisms of fundamental cognitive processes into the context of social cognitions. The goal is to highlight the potential for operationalizing social cognitive processes in predictive coding frameworks and related predictive mechanisms. The hope is that this will promote further discussion on the potential for extending current predictive frameworks in action, perception, and learning to cognitions required for socially relevant cognitive processes.

THE PREDICTIVE BRAIN

Before discussing the relationship between the predictive brain and the social brain, some crucial points will be introduced, which are particularly relevant to the issues related to social cognition. The terminology used to label different predictive processes can determine the definition of the mechanisms being referred to, and therefore it is important, for this article and for future work exploring this topic, to provide operational definitions of the terms being used. The mathematical methods on which computational models representing neural activity and behavior are built upon provide the basis for the application of the principles of predictive coding. The basics of the role of prediction in perception, action, and learning will also be introduced here.

PREDICTION, INFERENCE, AND SIMULATION

We can roughly distinguish between three main predictive concepts that are relevant to different aspects of the predictive social brain; these are inference, prediction, and simulation. Inference can refer to deterministic short-term processes that are largely situated in current behavior and are probabilistic estimations about the state of the world, and are most relevant to prediction errors and concepts modeled with Bayesian statistics (Friston et al., 2009). In contrast, the term "prediction" is generally more relevant to long-term prospections made about the potential for distant future events to occur. Simulation can be stated as a constructed internal representation of imagined events (Gilbert and Wilson, 2007) based on episodic memory (Williams et al., 1996). This can include processes such as scene construction that retrieves and integrates previous experiences to form a coherent event or mental image (Schacter and Addis, 2007), and is somewhat autobiographical in nature (Buckner and Carroll, 2007). Related to this is the framework of the predictive brain from Barsalou (2009) who describes simulation as a "re-enactment of perceptual, motor, and introspective states (e.g., affect, motivation, intentions, metacognition, etc.)". For the sake of integration, but without intended overgeneralization, the use of the term "predictive mechanisms" in this article does not only refer to higher-level long-term prospections about the future, but are also inclusive of the more fundamental low-level short-term predictive processes of inference, such as those conceptualized in Bayesian predictive coding formulations of action and perception. The frequent lack of consistency in the previous use of these terms reflects a need for further specification of what they imply, though the use of these terms are clarified operationally in this article according to the context discussed.

BAYESIAN INFERENCE

To introduce the concept of the predictive brain, it is important to mention the basic principles behind Bayesian statistics, as this underlies many of the predictive coding frameworks in perception, action, and learning. Bayesian statistical inference is a mathematical method of inference which incorporates priors, or prior beliefs learned from previous experiences that generate internal models of a predicted outcome, and consequently act as top-down modulators of bottom-up sensory input. This statistical method can be used to determine the probability of a certain outcome, given a predetermined assumption (i.e., the prior "belief" and the likelihood), which can then be subsequently updated according to the actual outcome. This is thought to be comparable to how the brain makes predictions about future outcomes in learning and motor control. An example of Bayesian inference states that we already have a prior belief of the probability of seeing either a white van or a white polar bear in the street, and thus the probability of receiving the visual input of a white van will be much higher than that of a white polar bear. While this example clearly relates to one's acquired knowledge, priors can also be innate. For example, visual perception experiments have shown that the recognition pattern of an object changes just by turning the image of the object by 180°, because animals have the built-in expectation that natural light comes from above (Schober and Rentschler, 1972). The Bayesian inferential approach can be used to create simulated models of neural activity. Substantial experimental evidence demonstrates that motor learning is performed by subjects in a Bayesian fashion (Kording and Wolpert, 2004) with probabilistic calculations being performed in the brain to predict future sensory consequences. This Bayesian predictive coding scheme has been applied to perceptual inference (Kiebel et al., 2009), perceptual learning (Friston, 2008), reinforcement learning (Friston et al., 2009), active inference (Friston et al., 2010), attentional processing (Feldman and Friston, 2010), and sensorimotor control (Wolpert and Miall, 1996).

INFERENCE AND PREDICTION IN PERCEPTION AND ACTION

It is clear that perception is not just a merely reactive process in which sensory input is received and registered, but is more likely a construction of predicted representations of the environment. The concept of generative internal models in perception was proposed many years ago, and represents another common principle that is at the basis of some predictive coding frameworks. The first experimental evidence found to support the proposal of predictive internal models of perception came from Sperry (1950), referring to a "corollary discharge," and von Holst and Mittelstaedt (1950) referring to an "efference copy." These were both set out to explain how we perceive a stable world despite rapid and frequent eye movements. In both perception and action, the efference copy creates a predicted internal ("forward") model of the sensory feedback according to the corresponding visual percept or motor action. The corollary discharge is then generated from this forward model and is then compared to the actual incoming sensory feedback, or afference, once transmitted from the peripheral sensory receptor to the brain. The matching process that occurs between the generative forward model, or corollary discharge, and the corresponding sensory feedback is thought to maintain the experience of stability in the ever-changing visual scene during eye movement. During motor control, the matching process between the generative forward model of the planned action and the sensory consequence of the action is thought to be the neural basis for distinguishing between self-generated and externally generated motor actions (Blakemore et al., 1998). Predictive coding in visual processing has been implemented in various hierarchical models (Rao and Ballard, 1999) that generally propose an integration of top-down expectancies and bottom-up sensory input reflecting stimulus information. Feedforward connections carry error signals, but in addition, cortical feedback connections transmit expectancy biases or predictions. A similar matching process is thought to occur during the execution of motor actions, and has been argued to be the underlying mechanism that maintains a sense of agency or ownership over our own motor actions. A mismatch, i.e., a discrepancy between the predictive forward model and the actual sensory feedback, can have pathological consequences, particularly relevant to schizophrenia and pathological abnormalities in ownership of action (Feinberg, 1978). This discrepancy between the expected and actual sensory outcome of a visual percept or motor action can also be referred to as a prediction error.

The coupling of action and perception is well-established, with the ideomotor framework even considering action and perception as inseparable, and fundamentally the same (Prinz, 1997). The ideomotor theory of action proposes that internal representations or images of actions are coupled to the execution of actions themselves, and that perceptual events associated with an action also initiates an internal representation of that action. [e.g., the theory of event coding (TEC) (Hommel et al., 2001)]. Frameworks for forward models of action have been formulated and operationalized in computational models, such as the MOSAIC model (Haruno et al., 2001), which rely on Bayesian inferential statistics. In such frameworks, "forward" refers to the causal direction from motor command into the corresponding sensory consequence, with the forward model ("predictor") being generated from the efference copy. An inverse model ("controller") represents the opposite direction, whereby desired sensory consequences are transformed into motor commands. Computational representations of forward models of action have been formulated with Bayesian statistics, with priors and their likelihood acting as predictive elements for the outcome of an action.

PREDICTION IN LEARNING AND DECISION-MAKING

Prediction also crucially guides learning through the updating of future estimations about the state of the world and probabilities of the likelihood of potential future events. Computing probabilistic predictions about the outcomes of one's own and others' actions is based on previously learned action-effect contingencies, i.e., the associative mapping between the action and the outcome. Reward prediction errors generated in dopaminergic neurons, are thought to encode the magnitude of the discrepancy between expected reward and experienced reward (Schultz and Dickinson, 2000) and therefore, drive decision-making. This reward prediction error acts as a teaching signal for updating expected reward value and is the neural basis for learning. Error processing, response monitoring, and cognitive control are also

intrinsic to predictive processing (Hoffmann and Falkenstein, 2011), whereby errors are generated by the discrepancy between our predictions and the actual outcome, and allow us to learn from our mistakes. The neural substrate underlying the valuation of positive and negative feedback is largely founded upon the reward processing system in the brain.

THE PREDICTIVE SOCIAL BRAIN

Social interaction and social functioning involves a multitude of socially relevant cognitive processes including, to name a few, social perception, understanding others' actions, observational social learning, social decision-making, and empathy. Top-down influences of social information can directly drive how we process visual information. More evidence is emerging which suggests that a similar mechanism, or "shared neural representation" is used for understanding others' actions, whereby an internal model of others' actions allows us to make predictions about the consequence and outcome of an observed action, and consequently understand and interpret the goals and intentions of the action. Many authors have suggested how conceptualizations of fundamental cognitions such as learning, could be extended to explain mental processes required for social understanding, social interaction, and social learning (e.g., Rushworth et al., 2009). There is also substantial work to indicate that there is top-down influence of social information and social interaction on fundamental error processing, learning, and decision-making processes.

We now present neurophysiological and behavioral findings, and computational principles that point to an essential relevance of predictive mechanisms in a broad variety of social cognitive processes that may be intrinsic to motor, perceptual, and learning processes, and permeate different levels of processing. Numerous parallels are also drawn to illustrate how the basic principles in prediction, inference, and simulation in non-social contexts can be applied to social aspects of cognition.

SOCIAL PERCEPTION AND SEEING OTHERS' ACTIONS

Person perception can be described as the impressions or mental representations we form of others based on socially constructed information, for which the perception of actions and faces act as crucial cues, and which predictive mechanisms are likely to also play a central role. The superior temporal sulcus (STS) is involved in the perception of biological motion and in inferring the intentions or goals from biological motion (Perrett and Emery, 1994; Allison et al., 2000; Jellema et al., 2000), and has been implicated in the mirroring network (Molenberghs et al., 2010). When we observe others' actions, we can see activity in the STS and it is, therefore, likely related to the mirror system and possibly in determining whether movements have social relevance. A recent study used an fMRI repetition suppression paradigm to measure activity in the action observation network while watching a robot, an android, and a human move (Saygin et al., 2012). This study interestingly found neural activity that was distinctive for the mismatch between (human versus robotic) appearance and motion, which was proposed to reflect prediction error activity, possibly as an index of an expectancy violation. They also suggested that this mismatch prediction error signal could account for the "uncanny valley" in which androids are seen as strange and disconcerting if they are too human-like (Mori, 1970).

The STS also appears to have a role in face perception, and the perception of the dynamic features of a face (Haxby et al., 2000; Gobbini and Haxby, 2007; Ishai, 2008). An MEG study from Furl et al. (2007) found that evoked neuromagnetic fields, originating from the fusiform face area (FFA) and the STS, were modulated by adaptation to facial expressions, and that these predicted behavioral after-effects. They propose that this can be explained by experience-dependent coding, according to a predictive coding account, which consequently creates top-down biases in face perception. Another phenomenon of face perception, in which low level visual processes may be modulated by socially relevant factors, is the "other-race effect." People have been shown to be better at recognizing faces of their own race as opposed to other races (O'Toole et al., 1994; Meissner and Brigham, 2001), which appears to occur at the visual encoding stage of face processing (Walker and Tanaka, 2003). This effect could be accounted for by after-effects from visual adaptation to facial race categories (Webster et al., 2004) that is likely to be based on long-term expertise and learning processes (Rhodes et al., 1989; Stahl et al., 2010), and has been represented by hierarchical generative models in a predictive coding framework (Furl et al., 2007).

Social perception can refer broadly to high-level visual processing of socially relevant stimuli, though social factors can also influence low-level visual processing performance. One major challenge for theories of forward models of action is to demonstrate an inverse relationship in which motor behavior directly influences perception. It has been shown that synchronized and communicative interaction can influence visual discrimination performance (Neri et al., 2006), and improve visual detection of biological motion (Manera et al., 2011), respectively. Manera et al. (2011) explain their finding in terms of predictive coding in that one's own communicative gestures can predict the other's expected action. Bortoletto et al. (2011) found, with EEG event-related potentials (ERPs), that action plans and intentions of observed hand gestures can modulate ERPs associated with early visual processing of the observed actions. Motor training has been shown to directly modulate activity in the occipital lobe (Engel et al., 2008), and TMS over the ventral premotor cortex, but not the primary cortex suppressed a visual after-effect when categorizing others' actions (Cattaneo et al., 2011). These studies demonstrate an early effect of social interaction on low-level visual processing, at an early stage of processing before awareness, therefore confirming the inverse relationship. Research on the neural processing associated with observing others' actions has received widespread interest in a broad range of research areas, particularly in the last 15 years, though there still seems to be some divergence in theoretical standpoints, which could potentially be bridged with a common dialog of predictive mechanisms.

PREDICTION AND THE MIRROR SYSTEM

The discovery of the activation of apparently functionally-specific "mirror neurons" in monkey premotor cortex during both action execution and action observation (Gallese et al., 1996) has led to broad speculations about their role in social cognition through

action understanding. This hypothesis is compatible with simulation theories of theory of mind (e.g., Davies and Stone, 1995), which in general argues that individuals utilize simulations of their own actions, and consequently their own thoughts, intentions, beliefs, and emotions to predict the mental state of others and therefore, ascertain knowledge of other minds. It is thought that this then ultimately provides the fundamental elements for the ability of an individual to understand, and empathize with, the social behavior of others. Naturally, this has also revealed a number of controversies questioning the functional specificity of the mirror system (e.g., Hickok, 2009), and the anatomical validity of a human mirror system as originally specified (e.g., Molenberghs et al., 2009; Mukamel et al., 2010).

Some alternative models of the mirror neuron system have been put forward to try to deal with some of these issues and controversies. One, which is most relevant here, is a predictive coding account of the mirror neuron system (Kilner et al., 2007a,b) that uses a Bayesian framework for its implementation. It argues that an internal model is generated during action observation, which in turn transfers an action prediction through backwards connections, from frontal areas implicated in the mirror system, to action representations in the STS and parietal mirror areas, which then produces an action prediction error. As with other predictive systems, the brain seeks to minimize the prediction error. This has been demonstrated with simulations of handwriting that artificially produce electrophysiological responses to movement expectation violations (Friston et al., 2011). Another alternative account of the mirror system is based on associative learning (Heyes and Ray, 2000; Heyes, 2001), and argues that learned sensorimotor experiences, from self-observation and the observation of others, actually promote the formation and emergence of a mirror system, which is acquired and refined throughout development. The learned associations of action contingencies are thought to provide the basis for action understanding. This associative learning account is supported by findings related to expertise and familiarity of actions in motor cortex activity during action observation, and by studies showing neural activity outside of the mirror neuron system during observation of actions that are unfamiliar or difficult to understand (Brass et al., 2007; Kilner and Frith, 2008). Greater expertise and familiarity of observed actions induces greater activity in the action observation/mirror-neuron network in the brain (Calvo-Merino et al., 2005, 2006; Orgs et al., 2008). This is evident from both practicing a particular motor sequence and from passively observing actions (Cross et al., 2009). Automatic imitation and motor interference also appears to be influenced by previous sensorimotor experience (Capa et al., 2011). These findings lend themselves to an associative learning account of imitation and the mirror system (Catmur et al., 2009), whereby motor representations can be learned through observation (Hayes et al., 2010). Although it is likely that the coding of motor sequences for observed and practiced actions differs (Gruetzmacher et al., 2011), though this is still an elusive, but crucial issue in conceptualizations of imitation and action observation.

The only known single neuron recordings of the proposed mirror system in the human brain comes from Mukamel et al. (2010), who intriguingly found activity in the hippocampus, an area never before included in the classical mirror system. The involvement of the hippocampus in a mirror neuron network could potentially be accounted for by Bar's (2009) proposal of the predictive brain with memory "scripts" as predictions, and by Barsalou's (2009) suggestion of the involvement of long-term memory in simulation and perceptual prediction, which may not have been detected previously with fMRI techniques. Bar presents an integrated framework of perception and cognition that argues that memory "scripts," generated through learned associations from previously real and imagined experiences, form the basis for predictions of what is about to come next in our environment. It is also suggested that this association-based prediction framework can be applied to prediction in social interactions (Bar, 2007; Bar et al., 2007). By taking an inference-based account of the mirror neuron system, this allows for the integration of Bar and Barsalou's frameworks into the realm of social cognition, action understanding, and the mirror neuron hypothesis.

These accounts of the mirror neuron system highlight the potential role of predictive mechanisms, particularly simulation, and inference with the predictive coding, and associative learning accounts, in social interaction. Consequently, these accounts could legitimately be extended to highlight the role of prediction, simulation and inference in other non-motor social cognitions associated with mirror neuron activity. Inference-based accounts of the mirror neuron system could potentially apply to some examples of work in social neuroscience showing that mirror neuron activity has been implicated in the distinction between self and other (Sinigaglia and Rizzolatti, 2011), mentalizing (De Lange et al., 2008; Centelles et al., 2011) and simulation of emotions (Bastiaansen et al., 2009). Even though the mirror neuron hypothesis provides a very appealing explanation for the processing of others' actions, there are other theories also related to predictive mechanisms that propose integrative frameworks for sensorimotor control and social interaction.

FORWARD MODELS OF ACTION AND SOCIAL INTERACTION

Forward models of action and the corollary discharge are thought to be crucial in determining ownership of action, or sense of agency, and being able to distinguish between self and other by distinguishing between self-generated actions and movements generated by external forces (Fourneret et al., 2001; Franck et al., 2001; Knoblich et al., 2004; Yomogida et al., 2010). Numerous studies have shown that our sense of agency for our actions can be disturbed if there is a discrepancy in visuomotor perception between expected and intended actions (Daprati et al., 1997; Franck et al., 2001; Van den Bos and Jeannerod, 2002). One recent study demonstrating this found that a pre-reflective or implicit sense of agency can be influenced by the accuracy of sensorimotor predictions (Gentsch and Schutz-Bosbach, 2011). The ability to distinguish between self and other is a fundamental prerequisite for many social cognitive processes required for understanding others.

A corollary discharge has also been proposed to be present in the speech system, and is therefore, suggested to be responsible for attributing self-generated speech as one's own (Ford and Mathalon, 2004). Evidence mostly comes from ERP work

on occasions when a disturbance in the corollary discharge occurs, which is relevant to symptoms seen in schizophrenia, particularly with auditory hallucinations. This auditory corollary discharge may also therefore, contribute to establishing the distinction between self and other in verbal communication. A recent study used MEG to compare valid and invalid predictions made between visual speech input and auditory speech signals (Arnal et al., 2011). From their results, they inferred that top down predictions were coded by slower frequencies of neural activity, whereas prediction errors in audiovisual speech were reflected by high frequency ranges. In a social interactive setting, i.e., during natural verbal communication, Stephens et al. (2010) found that spatiotemporal brain activity of the speaker and the listener became synchronized, and the greater this coupling, the greater the understanding. The findings also revealed anticipatory neural responses in the listener, particularly in the striatum, medial prefrontal cortex (MPFC) and dorsolateral prefrontal cortex (DLPFC), areas that also encode the reward prediction error and value representation.

An extension of one forward model framework of action, the MOSAIC model, has been applied to explain social interaction (Wolpert et al., 2003). The model parallels the sensorimotor loop between the forward model and the incoming sensory information, with the social interactive loop being between self-generated and observed communicative actions. Communicative actions are thought to be generated from the motor commands observed by a confederate, which consequently causes changes in the observer's mental state, which in turn initiates communicative actions from the other person, which are perceived by the observer. This forward model of social interaction is proposed to allow us to make predictions and learn about the likely behavior of another person in response to our own communicative behavior. An inverse model of action in social interaction is proposed to be used to access the hidden mental states of others, and consequently predict their behavior. The internal models of other people are considered to be decoded and learned through the mappings between our own actions and our own mental states as *a priori* information, thereby using our own motor system to compute the internal mental states of others, and are consequently suggested to form a basis for theory of mind.

There are crucial differences between the hypotheses of the mirror neuron system and forward models of action. Internal forward models of action are likely to be coded in the cerebellum (Wolpert et al., 1998). Consequently, neuroimaging studies have suggested that some of the characteristics of internal models of action, seen from cerebellar activity, can be extended to understanding higher-level cognitions including optimization of behavior toward long-term goals and social interaction, particularly in predicting and understanding of others' actions, theory of mind and language processing (Imamizu and Kawato, 2009). Separate mechanisms in the cerebellum may underlie different processes for switching internal models, with predictive switching being based on changes in context, and postdictive switching being based on the sensorimotor prediction error (Haruno et al., 2001). Though interestingly, activity associated with the prediction error, used for the postdictive switch, was found in the inferior parietal lobule (IPL) (Imamizu and Kawato, 2008), an area implicated in the human mirror neuron network (Chong et al., 2008).

It is evident that predictive mechanisms of simulation and inference are likely to be central to both the mirror neuron system and social forward models of action, and may underlie fundamental processes recruited in social interaction. Predictive forward models of action generated from efference copies also may provide the basis for being able to dissociate ourselves from others, on different levels of processing and in different sensory and cognitive modalities. Novel comparisons can be established if forward models of action and the mirror neuron hypothesis are framed in a predictive coding scheme, and consequently stimulating more dialog between the mirror neuron work and work on forward models, while also having implications for social cognition. A crucial issue in making such comparisons is the degree to which neural activity associated with simulated/imagined actions or forward models of planned actions constitutes the same activity as that seen during the execution of a motor action.

PREPARING, PREDICTING, AND IMAGINING ACTIONS

The dynamic changes in neural activity during preparation, online control, and imagination of one's own movements are likely to correspond with, and be embedded in, the neural processes recruited in the prediction of action kinematics and action understanding, during observation of others during social interaction (Grezes and Decety, 2001). One crucial and unresolved issue when discussing the role of prediction in social cognition and motor actions is to what degree preparatory, imagined, predictive, and observational motor responses overlap in terms of neural activity and cognitive function. For example, it may be the case that preparing for an action recruits a forward model, and therefore, the associated neural activity could in part reflect the generation of the forward model and the corollary discharge. It is also not clear as to whether imagined actions also recruit a forward model, but without the matching process of incoming sensory feedback, which could also apply to the observation of others' actions in social interaction. To further clarify the role for motor-related neural activity in social interaction and social cognition, these issues need to be first resolved.

An ERP that has been found to be associated with motor preparation is the contingent negative variation (CNV) (Walter et al., 1964). The CNV partly overlaps with the lateralized readiness potential (LRP), another similar motor preparatory response. Kilner et al. (2004) have found that a CNV is also evoked for observed actions, reflecting a preparatory or predictive response to others' actions. The LRP is thought to reflect choice response (Coles, 1989), whereby lateralized motor cortex activity is seen according to the hand used for response, before the response is made. The LRP could be another ERP to use for future explorations of how these preparatory motor responses interact with social cognitions and social contexts, such as task-sharing and action co-representation (Hollander et al., 2011). If forward models are involved in motor preparation, then such ERPs most likely reflect the neural processing of the efference copy or corollary discharge for both one's own and for others' actions.

It is quite possible that the neural activity seen in preparatory motor responses substantially overlaps with the neural activity

during the prediction of one's own and of others' forthcoming actions. Predictable stimuli lead to faster reaction times, for which the temporoparietal junction (TPJ) has been implicated in terms of predictive motor coding (Jakobs et al., 2009), an area also crucial to the mentalizing network. Prediction and simulation of an observed action in real-time is most relevant to everyday action observation and action understanding in social interaction. Graf et al. (2007) showed subjects actions where part of the movement sequence was occluded, demonstrating better predictive performance when the timing of the occluder duration fit with the predicted movement, therefore suggesting that predictive mechanisms involved in the observation of others' actions uses real-time simulations. An intriguing study from Miles et al. (2010) found that mental time-travel, i.e., imagining the past and the future, correlated with the direction of subjects' movements, with subjects swaying forward when thinking about the future and swaying backwards when thinking of the past, suggesting an embodied representation of time and space. Interestingly, Mitchell (2009) highlights overlapping brain areas responsible for mental state inference and remembering the past, imagining the future, and spatial navigation to argue that internal self-projections are central to theory of mind processes.

The difference between the underlying neural processing involved in imagining and observing actions has relevance to the ideomotor theory of action. Recently, numerous confirmations of ideomotor principles have been revealed with neuroimaging techniques, particularly with studies demonstrating motor cortex activation for imagined actions (e.g., Decety, 1996). The ideomotor principle has also been used to explain imitation in an attempt to overcome the correspondence problem of imitation, in that movement specifics are not directly observable by the observer, and therefore, there is no direct way to match sensory input of another's actions onto our own sensorimotor system (Iacoboni, 2009; Massen and Prinz, 2009). Imagined actions and events have also been found to influence self-monitoring (De Lange et al., 2007; Turner et al., 2008), inferring a possible role in self-referential processing and consequently also in dissociating between self and other. An intriguing fMRI study has revealed that the prediction of sequential patterns can evoke activity in areas of the premotor cortex that are related to motor properties of the context of the prediction (Schubotz and von Cramon, 2002), without the execution of an action. This suggests that there may be a somatotopic mapping during the prediction of upcoming sequential events on corresponding motor cortex. The specificity of neural activity and dynamic changes involved in action execution, observation, and imagination are yet to be fully clarified. Paradigms investigating neural activity and behavior in more ecologically-valid social interactive scenarios, such as those using cooperative actions, are likely to shed more light on these questions.

PREDICTING AND MOVING TOGETHER

Studies investigating coordinated and cooperative actions are particularly relevant to social interaction and everyday social scenarios, in addition to passively observing actions. Joint action can be defined as a social interaction whereby two people coordinate their actions, often with a shared goal in mind (Sebanz et al., 2006), in other words, a co-representation of the action and its goal (Wenke et al., 2011). Given the implied role of the mirror system in imitation, co-representation, and coordinated actions, similar predictive mechanisms of prediction and simulation recruited during action observation in the mirror system could also be extended to apply to joint action, imitative, and synchronous behavior.

Imitation and synchronization of action with another person may reflect preparatory or anticipatory offline mechanisms during action observation and online real-time prediction of action (Konvalinka et al., 2010). Both may rely on similar processes of motor simulation in the brain that directly relate to inferential and predictive processes, in terms of prediction of forthcoming action and forward models of action, whereby an internal representation may guide imitation and synchronization facilitating matching of the other's actions. Individual differences in the ability to make temporal predictions for forthcoming events have been found during interpersonal sensorimotor synchronization (Pecenka and Keller, 2011), suggesting that temporal predictions could be trained through observation (Scully and Newell, 1985), and are also a necessary precursor to causal predictions, and action-effect contingencies. Therefore, the ability to make temporal action predictions may also be directly related to the ability to make more high-level, non-motor causal associations, inferences, and interpretations in social scenarios, such as during the process of mentalizing.

Automatic imitation and mimicry are thought to reflect underlying shared neural representations of action and mirror system related activity (Brass and Heyes, 2005). Imitative performance can be modulated by the social context of the action such as whether the performer is a human or not, the degree to which the observer relates to the performer of the action (Kühn et al., 2011), the level of self-focus (Spengler et al., 2010), the strategic context (Cook and Bird, 2011) and social attitudes (Cook and Bird, 2011). Synchronized movement promotes cooperative behavior (Wiltermuth and Heath, 2009) and the degree to which we perceive others as similar to ourselves (Valdesolo and Desteno, 2011) and the ability to pursue mutual goals together (Valdesolo et al., 2010), thereby also likely encouraging social cohesion. Joint action and interpersonal synchrony can also be influenced by social context, including perceived group membership (Chartrand and Bargh, 1999; Miles et al., 2011). Muller et al. (2011) found that ethnically white participants only showed a joint compatibility effect when observing a white hand, but not for a black hand, though this was eliminated when subjects were asked to take the perspective of the performer. Differences in group relations were also found to influence the tendency to co-represent remembered items of the co-actor (He et al., 2011). In addition to this, Humphreys and Bedford (2011) used neurological patients to infer that theory of mind and joint action may have some common neural substrate.

It is clear that much work has already been done to investigate the interdependency between high-level social cognitive processing and low-level motor processes. The top-down influence of social information on bottom-up neural motor activity and the apparent embededness of social cognitive processing in the processing of both one's own and others' motor actions demonstrates

the potential coupling of movement to social cognition. It is also evident that predictive mechanisms of simulation and inference, and predictive coding frameworks, provide a fruitful foundation on which to build further common dialogs between currently disparate research disciplines and theoretical viewpoints. However, it is not only the motor response associated with the observation of others' actions that is represented in the observer's brain, but also includes the consequence of the outcome and the implications of the observed action in terms of error, feedback, and reward, and therefore consequently influencing decision-making and learning. Predictive mechanisms also lie at the core of the processes of evaluation of the outcomes of others' actions, and can be applied to both non-social and social contexts.

COGNITIVE CONTROL AND ERROR MONITORING IN A SOCIAL CONTEXT

The ability to accurately detect and process errors is crucial for learning. Certain EEG ERPs are thought to be indices of error-processing and the reward prediction error. The feedback-related negativity (FRN) is evoked when negative or positive feedback is given following response choice and is considered to be an index of reward prediction and expectancy violation (Holroyd and Coles, 2002). An error-related negativity (ERN) is seen following the onset of muscle activation during an erroneous response in a forced choice reaction time task (Falkenstein et al., 1990). The ERN is an index of error-processing and response monitoring, when the intended response is different from the executed response (Baker and Holroyd, 2011), and has been found to originate from the anterior cingulate cortex (ACC) (Dehaene et al., 1994). Both the ERN and FRN are intrinsically linked to each other and are mediated by the mesencephalic dopamine system and projections to the ACC (Holroyd and Coles, 2002).

Some studies have recently shown that corresponding brain activity involved in error and feedback processing can also be evoked by the observation of others' performance. An ERN and FRN is evoked when watching other people's mistakes (observational ERN or oERN) (Van Schie et al., 2004) and when observing feedback from other people's response choices (observational FRN or oFRN), respectively, with the oERN and oFRN both also thought to originate from the ACC (Yu and Zhou, 2006). Shane et al. (2008) have confirmed the activation in the ACC, in the dorsal region, during one's own and observation of a confederate's errors, with additional activity also being found in orbitofrontal areas and premotor cortex. Though, interestingly, a dedicated network appears to be active only when observing others' errors, which includes the inferior parietal cortex (IPC) and the rostral and ventral parts of the ACC (r/vACC), with the IPC correlating with measures of perspective-taking and the r/vACC correlating with self-reported empathetic concern (Shane et al., 2008, 2009). Another recent study found activity in the MPFC, an area associated with the mentalizing network, specifically activated for errors that affected others (Radke et al., 2011).

Observational error and feedback processing also seems to be influenced by the degree of self-relatedness and the interpersonal relationship between the observer and the performer, i.e., if the performer is a friend or a stranger, with differences seen in activity in error-related brain areas (Newman-Norlund et al., 2009), and in error-related (Carp et al., 2009) and feedback-related ERPs (Kang et al., 2010; Ma et al., 2011). Competition and cooperation appear to modulate processing of observed errors to the degree that they influence performance monitoring and even modify performance adjustments. For example, when observing someone else's errors, it appears that a post-error slowing occurs for one's own errors in a cooperative scenario, although there is a post-error speeding in the competitive scenario (De Bruijn et al., 2012; Nunez Castellar et al., 2011). An ERN has also been found to be evoked by observed errors performed by cooperators, whereas observed correct responses of competitors evoked a later ERN (Koban et al., 2010). It has been confirmed that this activity is likely to be not just associated with self-reward, but is a reflection of performance monitoring and updating of expected outcomes based on others' actions (De Bruijn et al., 2009). The FRN and oFRN have also been shown to be modulated by competition and cooperation (Itagaki and Katayama, 2008; Rigoni et al., 2010; Van Meel and Van Heijningen, 2010), suggesting that this neural response is influenced by both the benefit or loss to oneself, and the benefit or loss of others (Marco-Pallares et al., 2010).

These studies all demonstrate how the neural processing of both one's own and others' errors and feedback can be directly influenced by social context and by differences in the social relationships between confederates involved in a social scenario. Therefore, the central role of error and feedback in predictive mechanisms of inference and learning provides a fundamental link between prediction and social cognition. However, an important note to make here is that it is not clear as to how others' gains interact with our own processing and valuation of reward, i.e., from the observed choices of others. This is a crucial issue, as it addresses the degree to which others' gains can be rewarding for us. Differences in neural activity may be wholly reflecting some form of "empathetic" response to others' experience, or, though not mutually exclusive, may be an index of the relevance of the reward to oneself, as the outcome of others' choices may be indirectly associated with a reward for us.

SOCIAL LEARNING AND SOCIAL REWARDS

Observational learning is acquired through making associations between actions and their outcomes, and the value associated with that action and the predicted outcome. It is becoming more apparent that there are some common cognitive and neural processes driving both active experiential learning and observational social learning. In particular, social learning has been proposed to be based on the same simple processes recruited in associative learning. Heyes (2011) compares learning across different species suggesting that learning only becomes social through adaptation to interactions with conspecifics, and "tuning in" of perceptual, attentional, and motivational information channels to other social agents. She convincingly argues that social learning does not involve mechanisms that are different from those used in non-social learning, and therefore do not have special "social" properties. In support of this, Jones et al. (2011) found neural activity in areas associated with basic reinforcement learning during a task involving acceptance from peers. One fMRI study also revealed similar underlying neural mechanisms during social valuation and non-social associative reward-based learning, finding a "social prediction error" (Behrens et al., 2008). Computational

modeling has also been used to relate the brain network responsible for reward-related processing with the theory of mind network (Behrens et al., 2009). Therefore, the principles underlying associative learning can also be extrapolated to explore the role of predictive mechanisms in observational social learning.

It is already known that the processing of reward is dependent on the context in which the reward is presented (Nieuwenhuis et al., 2005). Although, there is much evidence to suggest that there is something special about social contexts (e.g., cooperation versus competition) and social relations (e.g., ingroup versus outgroup) that modulate the computation of value. Differences in activity can be seen in brain areas associated with motivated behavior and reward evaluation when a social betting task is compared to non-social betting (Nawa et al., 2008), namely the amygdala, the right DLPFC and the ventral striatum. The processing of feedback and one's own experience of reward, for others' gains or losses interacts with how the observer views the other person, in terms of the opinion or social evaluation of them. Our own valuation of objects can be influenced by others' opinions, as Campbell-Meiklejohn et al. (2010) demonstrated differences in activity in the ventral striatum, an area thought to code prediction error-related activity, depending upon the opinion of an "expert" reviewer. Ratings of subjective value and the associated neural activity have also been shown to be affected by the valuations made by one's peers, particularly in the nucleus accumbens and the orbitofrontal cortex (OFC) (Zaki et al., 2011). In addition, the dorsal striatum has been found to encode reward prediction errors in both one's own experiential instrumental conditioning and the observation of others' (Cooper et al., 2012). There is some evidence to suggest that there may be a common underlying neural network related to one's own valuation of rewards and the valuation of others' action outcomes during observational learning, which may culminate in the ventromedial prefrontal cortex (VMPFC) (Behrens et al., 2008; Hare et al., 2010). A review of social preferences collates numerous fMRI studies to find common activation in the dorsal and ventral striatum for the processing of social rewards, with these areas substantially overlapping with areas related to reinforcement learning and anticipation of monetary reward (Fehr and Camerer, 2007), further adding to the argument for shared neural representations for one's own and others' rewards.

Sescousse et al. (2010) found prediction error-related activity in the ventral striatum, anterior insula, and the ACC for monetary reward, and from the presentation of erotic stimuli, suggesting some common coding of prediction errors regardless of the type of reward, social or not. However, a more recent study found distinctions between brain areas activated during the processing of financial reward feedback, and the valuation of social stimuli, suggesting some separability between the brain's classical reward circuit and the network responsible for the valuation of social stimuli (Evans et al., 2011). Furthermore, a distinction between action prediction errors and outcome prediction errors have been made in neural areas associated with observational learning (Burke et al., 2010). The action prediction error is proposed to reflect the discrepancy between expected and observed action choices, coded in the DLPFC, and the outcome prediction error is thought to represent the discrepancy between the expected and observed outcome received by others, coded in the VMPFC.

In social learning, it may be the case that social context and our opinion of others induces different motivational states that correspond to different utility functions, in terms of reinforcement learning theory and expected utility theory, which consequently dictate social decisions and future social judgments. The motivational states in learning theory (Niv et al., 2006) are mappings of the utility onto the outcome, whereby valuation is driven by both external factors (i.e., the probability of the occurrence) and the internal context (the motivational, emotional, and cognitive state). This could be paralleled in social contexts in which the internal state is driven by predetermined judgments and opinions of the other person and our intrinsic social needs (the internal context), that is weighed up against a statistical probability calculation based on prior experience, learned through socially relevant stimuli and cues (the external factor).

There is conflicting evidence to argue for and against a distinction between social observational and non-social/active learning. However, it appears that more weight is given to the side that proposes a lack of distinction, in that both largely share some common neural substrates, with both also utilizing a form of prediction error, associated with both the valuation of one's own and of others' outcomes in non-social and social scenarios, respectively. Although it is clear that social learning involves an additional dimension in which the social context can directly influence the valuation of an outcome. The social context created by the external environmental situation (e.g., competition or cooperation) or by internal motivational states (including that created from prejudice or through the social relationship with a confederate) can determine the valuation of rewards from others' outcomes. Consequently, social contextual factors will contribute to the formation of social judgments and as a result could also drive decision-making in social situations.

SOCIAL DECISION-MAKING AND ECONOMIC GAMES

Social decision-making deals with high-level computations based on complex socially relevant information such as fairness, trust, social norms, and social preference. Economic decision-making in social contexts, though apparently recruiting additional processes, is still rooted in reward processing and cognitive control, and can also be framed in terms of probabilistic predictive computations, and consequently has been shown to involve similar neural structures in both social and non-social decision-making. This has been largely explored with economic games that include a social component, often with some form of social interaction.

Feedback indicating a violation of a social norm and social expectation has been shown to evoke an FRN, suggesting that the brain treats social deviance in a similar way to a prediction error (Harris and Fiske, 2010; Kim et al., 2011). Klucharev et al. (2009) also confirmed this with activity seen in the ACC and supplementary motor area when there is conflict with a social norm. It is likely that cooperative behavior and biding by social norms is based on observational learning and inference-based processes (Boyd and Richerson, 1988; Seymour et al., 2009; Yoshida et al., 2010). These findings relate closely to studies showing the effect of others' opinions on our own valuation of objects, as previously

mentioned. Popularity ratings have been shown to influence the valuation of adolescents' ratings of music, and interestingly, the tendency to change one's opinion of a song positively correlated with activity in the anterior insula and the ACC (Berns et al., 2010). Activity in the DLPFC and ACC, both crucially involved in cognitive control and error processing, have consistently been found to be activated when making moral decisions, in particular, when making decisions between fair and unfair offers (Sanfey et al., 2003; Greene et al., 2004). A study using a social comparison scenario that induced self-reported envy found that activity in the dorsal ACC was positively correlated with levels of envy (Takahashi et al., 2009). The ACC and DLPFC have also been shown to be activated when one breaks a promise, as compared to fulfilling that promise (Baumgartner et al., 2009).

The overlap between reward prediction error-related neural activity and activity utilized in social judgments implies an underlying role of prediction in more complex, higher-level socially relevant psychological processes, such as empathy, trust, judgments of fairness, envy, shame, and guilt. Activity in the ventral striatum has been found during the experience of mutual human cooperation, as opposed to cooperating with a computer (Rilling et al., 2002, 2004), with two other reward-related areas implicated in reciprocated cooperation, namely the caudate nucleus (Rilling et al., 2002, 2004; Delgado et al., 2005) and the OFC (Rilling et al., 2002, 2004). It is likely that people experience some hedonic pleasure when acting altruistically (Thibaut and Kelley, 1959), which outweighs the potential financial cost. This is confirmed in studies showing activity in the reward circuit when giving charitable donations (Moll et al., 2005; Harbaugh et al., 2007). A proposed computational model of decision-making demonstrates that the application of reinforcement learning theory in game-theoretic social interactions and imitative or inference based observational learning can be used to generate altruistic behavior (Seymour et al., 2009). The evaluation of fairness and the social comparison of monetary rewards have been associated with activity in the ventral striatum (Fliessbach et al., 2007), with fairer offers also inducing greater activation in the VMPFC and higher subjective ratings of happiness (Tabibnia et al., 2008). The VMPFC has also been found to be implicated in judgments of trust (Krajbich et al., 2009) and being trusted by others (Li et al., 2009). The DMPFC, caudate nucleus, and the striatum have been shown to be activated when learning the trustworthiness of another person (King-Casas et al., 2005). Findings of activity in the VMPFC, medial OFC, and DLPFC in emotional synchrony of another person also relate to this (Kühn et al., 2011). Implicit judgments of trustworthiness from facial cues influence social decision-making (Van 'T Wout and Sanfey, 2008; Schlicht et al., 2010), and the use of reinforcement learning models of trustworthiness also suggests that the evaluation of trust is based on probabilistic beliefs that are dynamically updated according to the proceeding experience and prediction error (Chang et al., 2010).

The substantial overlap between areas encoding prediction errors, error monitoring, and those implicated in social decision-making tasks implies a common neural basis for social and non-social decision-making processes. This therefore, also highlights the central role of predictive mechanisms of inference in social decision-making and the formation of complex social judgments. In addition to shared neural representations of others' motor actions, outcomes of others' actions, and the implications of others' actions on the observer, there can also be shared sensory and emotional experiences when watching others. This then brings us closer to a conceptualization of empathy in which we not only experience the cold cognitive processes of others, but also experience others' emotional state during observation.

PREDICTING OTHERS' FEELINGS

An interesting study has recently shown synchronized arousal, reflected in heart rate data, among spectators and fire-walkers during observation of a collective ritual (Konvalinka et al., 2011). Just as sensorimotor matching or motor resonance can occur during action observation, it appears that other people's sensations and emotions can also be contagious and therefore, has consequently been linked with the mirror system. A recent fMRI meta-analysis of areas implicated in the human mirror system found significant overlap with areas involved in emotional processing (Molenberghs et al., 2012). Observed tactile stimulation can induce shared experiential and neural representations of the others' somatosensation, including another's pain. Threat detection is involved in evaluations of trustworthiness and social decision-making. This clearly has adaptive advantages for survival and has evolved from the ability to efficiently perceive fear-related stimuli, and can also be transmitted through social interaction. Fear-conditioning is likely to be based on similar predictive mechanisms as in reinforcement and associative learning, and these principles could legitimately be extended to explain the social transmission of fear by observational learning processes. Emotional contagion forms the basis for affective empathetic responses (Decety and Ickes, 2009), which may be founded on internal predictive or anticipatory emotional representations (Gilbert and Wilson, 2007). Simulations of emotions, or "pre-feelings," may not only be used to imagine future emotional states, but may also be used to simulate others' emotional states in social interactive scenarios.

Observing actions, tactile, and painful stimulation in others all invoke activity in the brain of the observer in secondary somatosensory cortex (SII) and BA2 (Brodmann Area 2—posterior primary somatosensory cortex), which is adjacent to SII (Keysers et al., 2010). Some authors have pointed out a lack of distinction between motor and somatosensory activation during the observation of others' actions, and consequently argue for a lack of distinction between somatosensation and motor processes in the hypothesized mirror neuron system (De Vignemont and Haggard, 2008). Numerous fMRI studies have confirmed this overlap showing activation of SII, but more significant activation of BA2 during action observation (Grezes et al., 2003; Dinstein et al., 2007; Gazzola et al., 2007; Gazzola and Keysers, 2009; Turella et al., 2009). Empathy and self-versus-other-related processing can also influence somatosensory perception (Jackson et al., 2006; Lawrence et al., 2006). Serino et al. (2009) found that tactile somatosensation on one's own face, while observing another person's face being touched, was enhanced when the observed face was of the same ethnic or political group. Serino et al. (2008) also previously discovered that viewing one's own face can enhance tactile sensitivity, which is

also reflected by enhanced neural activity in a ventral parietal area in a later fMRI study (Cardini et al., 2011). This also appears to work in the other causal direction, in which somatosensory stimulation of one's own face can improve self-face recognition (Tsakiris, 2008).

The prediction error signal has a crucial role in fear-conditioning and avoidance behavior, achieved by learning relationships between harmful events and environmental stimuli (Delgado et al., 2008; McNally et al., 2011; Spoormaker et al., 2011). Many different animals can evidently learn fear from the observed fear-related behaviors of a conspecific (e.g., John et al., 1968; Kavaliers et al., 2001; Munksgaard et al., 2001; Knapska et al., 2010). Aversive learning can be communicated by primates through fearful face expressions, with some studies providing support for the suggestion of common processes in fear conditioning and observational fear learning (Mineka et al., 1984; Mineka and Cook, 1993). Facial expressions are also one of the main ways for socially transmitting fear in humans, with the expression of another's response to stimuli serving as the Pavlovian aversive US (unconditioned stimulus) for the observer (Gerull and Rapee, 2002). Other physical cues can also lead to learned fear responses through observation (Berber, 1962), and even just abstracted information about a fearful response can lead to social transmission of fear (Hygge and Ohman, 1978), though this is also determined by context (Lanzetta and Englis, 1989; Singer et al., 2006). Neuroimaging studies reveal similar networks involved in both fear conditioning and observational fear learning. Primarily, the amygdala is central to the processing of fear, from both one's own experience and from others', though additional areas have been implicated exclusively in observational fear learning, including the anterior insula and ACC, possibly reflecting anticipation, and parts of the MPFC likely to be involved in some mental state inferential processing of the observed person (Olsson et al., 2007).

Fear and pain are directly related to one another, and in a social context, observing someone else's pain can induce a representation of that pain in the observer, with activations seen in the observer's somatosensory cortex (Cheng et al., 2008). Observation of others' pain is also directly related to the social transmission of fear. Neural responses induced by empathy for others' pain have also been shown to be modulated by perceived fairness (Singer et al., 2006), group membership (Forgiarini et al., 2011), emotional closeness (Beeney et al., 2011), emotional context (Han et al., 2009), self-relatedness (Perry et al., 2010), the identity of the person being observed and personality differences of the observer (Mazzola et al., 2010; Goubert et al., 2011). Goubert et al. (2011) have presented an intriguing account of the observation of pain and pain-related fear from an observational learning perspective, with added recent experimental evidence (Helsen et al., 2011). In support of this, Meulders et al. (2011) have demonstrated that pain-relevant fear conditioning is driven by associative learning mechanisms. It may also be the case that learned aversive behavior is directly linked to reward processing, in that it has been shown to be modulated by monetary reward (Guo et al., 2011), and the avoidance of aversive outcomes may in itself be rewarding, and therefore reinforcing aversion avoiding behavior (Kim et al., 2006).

In sum, it is evident that anticipatory neural responses, and predictive coding in the context of learning, are crucial to empathetic somatosensory representations of others' experiences and consequently have a central role in observational learning of fear and pain and emotional contagion. The learning of aversive behavior, transmitted socially by others, will have substantial mediating effects on social decision-making and social behavior. Both fundamental predictive inferential mechanisms and high-level expectations are likely to be at the root of such processes, with interaction and interdependence between processing levels forming the basis for fear-related social decision-making. Predictive mechanisms of simulation and inference are likely to form the underlying processes that allow us to have empathy for others' pain and to learn about aversive stimuli through observation. The ultimate function of shared representations of others' actions, errors, rewards, sensations, and emotions is likely to be the basis for understanding others' minds in social interaction. Arguably, at the highest level of understanding others' minds is the ability to make inferences about others' mental states, which may be founded upon many of the principles already discussed.

PREDICTING OTHERS' MINDS

In social neuroscience, "mentalizing" or "theory of mind" refers to the ability to infer the mental states of others, ultimately to predict another person's behaviors, and is a central topic of discussion. Daunizeau et al. (2010) present a meta-Bayesian model for solving the Inverse Bayesian Decision Theory problem, which is the problem of inferring the hidden causes of sensory signals under a prior assumption about the causes. These signals are hidden both in our own experience of sensory input and also hidden from the observer when observing others' behavior, as we do not have perceptual access to these sensory signals. When observing others, the problem is that we are required to determine someone else's prior beliefs and goals with only their behavior as the information available to infer this. This meta-Bayesian solution has been suggested to explain processes such as metacognition, mental state inference, and theory of mind. A recent predictive model of theory of mind has been proposed by Baker et al. (2011), which relies on Bayesian inferential statistics to model belief and goal-dependent action, which is mediated by the state of the environment and perceptual access to the belief state, and by general knowledge of the world and by general preferences. Bayes' rule is used to model mental state inference, in which action understanding is acquired from integrating "bottom-up information from observed actions and top-down constraints from the prior to infer the goal, given observed actions and the environment." This inverse planning model is described to account for goal-based predictions of future actions in new situations, according to predictions formed from similar previous experiences. Comparable to this, the mental state inference model (MSI), another computational model of mental state inference, uses forward models of action in a prediction circuit to incorporate visual feedback as the control mechanism for inferring the goals and intentions of others through mental simulation or motor imagery (Oztop et al., 2005). Mental state inference and theory of mind may represent, and be achieved by, the culmination of many fundamental predictive inferential, and simulation processes related to the processing

of others' actions, errors, rewards, and emotional cues. It is also evident that the models of mental state inference that incorporate some predictive principles, such as Bayesian inferential statistics and forward models of action, can accurately simulate behavior.

DISCUSSION

SUMMARY

In the present article, we sought to review findings from a huge body of literature on the role of prediction in perception, action, and learning. Specifically, our aim was to highlight the relevance of prediction in the realm of social neuroscience. Accordingly, prediction, in its many forms, plays a central and fundamental role in social cognition. In light of the interaction and interdependence of predictive processes at different levels of processing and inference, such processes are best represented by hierarchical models of the brain, as this also reflects the hierarchical structure of the cortex, with top-down predictions being conveyed through backward connections, and prediction errors being propagated forward across the cortical hierarchy. Previous experience shapes and guides our current behavior and the choices we make. Predictive coding models that include generative forward models and Bayesian statistical inference present a plausible mechanistic platform to integrate a broad range of topics. Adaptive social behaviors and social learning relies on updating states based on prediction errors. Experience-dependent development of sensorimotor coupling may be crucial to the development of the sensorimotor representation of the embodied self and a sense of agency. Among these, this article draws numerous parallels between fundamental principles of cognition, previously thought to be non-social, with cognitive processes required for social interaction.

Three main points are proposed in our review that link predictive mechanisms in the brain to processes encompassed in social cognition. The first point is that both bottom-up sensory input and top-down expectancies for non-social cognitive processes, founded upon predictive principles of inference and simulation, are responsible for social cognitive processes, and are essential for the processing of social information. It is suggested that the development of such non-social predictive mechanisms are crucial in the development of socially relevant neural and psychological processes. Secondly, but not unrelated to the first point, this review highlights a number of studies which imply that there may not be a categorical difference between related social and non-social cognitive processes, particularly also in terms of the neural substrates of such processes. The similarities between neural activity recruited for our own experiences and during the observation of others' appear to outweigh the differences. Thirdly, it is clear that the framing of stimuli in a social context can have a modulatory influence on non-social predictive neural activity and behavioral processes. Social contextual information evidently has a substantial additive contribution to top-down modulations of bottom-up sensory input. Generative forward models of perception and action may integrate social information to facilitate mental state inference and understanding of others' actions and intentions. It is also legitimate to propose the existence of "social" forward models that compete against each other, with other non-social forward models and with bottom-up sensory input. It is also evident that there is a bidirectional relationship between social information and sensory processing, i.e., both a forward and inverse relationship.

COMPETING "SOCIAL" INTERNAL MODELS

Multiple bottom-up and top-down predictive systems are likely to be at play in the brain, working together, and competing against one another in different levels of processing, being both hierarchical and parallel in nature. Evidence for this can be seen in the induction of interference in action-perception coupling (e.g., Zwickel et al., 2010a,b), whereby predictive systems compete with each other. Anscombe (1957) states that there an almost infinite number of possible inferential descriptions of a single action. Comparatively, there are also an infinite number of inferential interpretations of social cues, socially relevant stimuli, and social scenarios. There may be internal "social" models that compete with other top-down internal models, and bottom-up sensory information, consequently causing the modulatory effect of social information and social context on action, perception, learning, and other cognitive processes. These generative "social" internal models would be acquired through previous social experience and social knowledge. This suggestion of competing social predictive systems working in synchrony is supported by the MOSAIC model for sensorimotor control and social interaction (Wolpert et al., 2003), which proposes distributed cooperation and competition of internal models whereby forward and inverse models compete for overall control. Competition between predictive systems is also likely at different levels of information processing, whereby bottom-up prediction in terms of perception and action compete with likelihood estimations and top-down prior beliefs (i.e., knowledge established from social experience) in addition to the top-down priors proposed in current Bayesian predictive models of perception and action. Top-down modulations of prediction in social contexts are likely to be multidimensional, including both cognitive and affective dimensions. Evidence of competing predictive systems from the modulation of sensorimotor control and action planning by social information (Ferri et al., 2011; Sartori et al., 2009a,b) and emotional context is clearly illustrated with numerous studies mentioned in this review.

SOCIAL PREDICTION ERRORS

A category of social prediction errors are likely to be present in social learning, social action and interaction, and social expectancy violations, reflecting the discrepancy between expectation and actual experience in social situations. These subsequently may be founded upon similar cognitive processes and neural substrates as prediction errors coded in non-social contexts. The difference between a social and non-social prediction error would be that the social prediction error is mediated by additional modulating factors, including social knowledge and expectations of others' behavior. For example the coding of the social reward prediction error, when observing the outcomes of others' choices, may produce a prediction error similar to that produced when one observes the outcome of one's own choice, but rather the value becomes relative to the effect of the observed outcome on the observer, i.e., how rewarding other's rewards are to the observer. As already mentioned, this can be modulated by

cooperation and competition, and by the relationship between the observer and the performer. A social action prediction error could be generated when predicting, simulating, or inferring others' actions in a social interaction, as described by the MOSAIC model of social interaction (Wolpert et al., 2003). Another category of prediction error that may be specific to social interaction could be a higher-level complex prediction error that is founded upon previously learnt social information whereby expectancies about others' social behavior are formed. This could be produced in circumstances such as when social norms are violated or when promises are broken. Expectations, or predictions, for this specific social behavior prediction error could come from cultural knowledge, or may be created by contextual information. Contextual information for social expectancies could come externally from environmental cues, and particularly the context of the social situation, or could be generated from internal cues such as an individual's affective state, or from a cognitive bias, such as an attributional bias.

THE "PREDICTIVE" BRAIN AS A FOUNDATION FOR THE "SOCIAL" BRAIN

Due to the coupling of action, perception, and learning, it may not be valid to consider these domains as separable or independent, and consequently predictive coding models dealing with these domains may overlap to the degree unto which they also operate under some common processes. This may also be the case for socially relevant cognitions in that the fundamentals of perception and action are embedded in processes considered to be specially "social," but instead could be structural and functional extrapolations of these fundamental principles that link action, perception, and learning, i.e., fundamental predictive mechanisms. Therefore, it also follows that it is improbable that distinct neural structures or networks can be found as responsible for predictive mechanisms, which may also parallel the ubiquity and omnipresence of the "social brain" and the neural correlates of social cognition and social interaction. As evident from many of the studies mentioned in this article, there is substantial overlap between the "predictive" brain (see Bubic et al., 2010) and the "social" brain (e.g., Abu-Akel and Shamay-Tsoory, 2011), including areas such as the DLPFC, MPFC, DMPFC, ACC, TPJ, OFC, medial temporal areas, precuneus, ventral striatum, amygdala, lateral parietal areas and the motor, premotor and somatosensory cortices.

The modulation of predictive sensory, motor and learning processes, and associated neural activity by social information, could be explained merely by an enhanced attentional orientation toward stimuli that have saliency created by the social relevance. Social information may not be of a categorically different type, as compared to non-social information, but instead could just have a higher level of priority in terms of information processing in the brain, and therefore be more pronounced in terms of saliency, and consequently capturing attention and utilizing greater cognitive resources. This would be in line with work on social cue orienting; findings of automatic attentional orientation to social cues (Driver et al., 1999), gaze cues and joint attention (Friesen and Kingstone, 1998). Tipper et al. (2008) found little difference between brain areas activated with social cues and non-social cues, though this is also contrasted by other conflicting studies (Greene et al., 2009). Prioritizing the processing of social information over non-social information, in terms of attentional orientation and cognitive resources, may have had some adaptive evolutionary benefits for survival and reproductive success.

EVOLUTION OF THE "PREDICTIVE" AND THE "SOCIAL" BRAIN

The ability to accurately predict future events is always a major challenge for all living creatures, particularly because a failure to learn from errors may have fatal consequences. This suggests that natural selection has strongly operated on predictive mechanisms, however, in very different ways. Hibernators, for example, can "predict" future food shortages, even if they have never experienced such a situation before. In other words, their predictive knowledge is innate. In the case of our own species, future scenarios are much more diverse, and therefore, even more unpredictable. Consequently, prediction mechanisms need to be more flexible, are more dependent on feedback (to adjust goal-setting), and are perhaps more independent on actual sensory input. For example, humans readily learn that there may be a food shortage in the future, independent of the current food supply. In addition, the complexity of social encounters in ancestral and contemporary societies has shaped predictive mechanisms to become more sophisticated in social matters. Successfully predicting the behavior of conspecifics has almost certainly paid-off reproductively.

From an evolutionary point of view, it is conceivable that predicting and anticipating the sources of potential threat, but also the availability of food resources, is largely independent of specific "social" contexts. Arguably, however, in primates and humans, the increasing complexity of social environments may have left its mark on how predictive mechanisms operate in the brain. Put another way, predictive mechanisms have been constantly shaped by environmental contingencies that, in the case of primates, consequently became more social, and less non-social. Living in complex social groups or troops certainly poses different demands on predictive mechanisms compared to living a solitary life. For example, while social group living provides protection from predation, it also increases intersexual competition for mates. Consistent with this assumption, Sallet et al. (2011) have shown that social network size and social status in Macaques was correlated with gray matter volume in the superior temporal cortex and rostral prefrontal cortex and with increased connectivity between frontal and temporal areas, suggesting that the size of the social group an individual is able to cope with impacts on both brain structure and function (Sallet et al., 2011). Arguably it should also be advantageous to be able to correctly predict the behavior of conspecifics. Therefore, individuals who have been better in predicting others' behavior may have been those whose genes benefitted from a greater reproductive success. If correct, this strongly suggests that selection pressures operating on predictive mechanisms have made them more malleable by social matters. A hypothesis derived from these considerations would be that the brain areas concerned with "social" prediction are similarly moulded by the complexity of the social environment (i.e., network size).

WHEN THE "PREDICTIVE" BRAIN GOES WRONG

The relationship between predictive mechanisms in cognition and the brain, and social cognition also has implications in psychopathology. If the development of predictive mechanisms have evolved with increasing complexity, with the adaptation of social cognitions as a result of this, then it is also probable that deficits of social cognition (such as theory of mind, empathy, imitation, self-recognition) are manifested by underlying disturbances in the "predictive brain" in pathologies like psychopathy (Hare, 1980), autism (Mundy et al., 1986), and schizophrenia (Brüne, 2005). Brazil et al. (2011) found a reduced ERN in psychopaths only when observing others' actions, when compared to non-psychopathic subjects. However, they also found a reduction in the ERP for correct responses, therefore reflecting a more general deficit in the processing of others' action outcomes in psychopathy. These findings suggests that deficits in higher level social cognitions, such as empathy, may also be related to processing of others' actions and action understanding, and would therefore be reflected in associated neural activity. This is also suggested by studies that have found abnormalities in mirror-neuron related activity in people with autism (Oberman et al., 2005) and schizophrenia (Singh et al., 2011) as indexed by EEG and the mu rhythm suppression, although conflicting results make this still unclear (Raymaekers et al., 2009). It has been suggested that psychosis may be partly caused by disturbances and dysfunctions of predictive mechanisms in different domains and in different levels of predictive processing. A deficit in fundamental predictive mechanisms of cognition may be expressed in certain psychotic symptoms, including passivity symptoms such as disturbances in sense of agency, auditory hallucinations, and delusions of control (Fletcher and Frith, 2009; Corlett et al., 2010), which are intrinsic to both social cognition and predictive mechanisms. Inherent of this may also be the abnormalities seen in error-processing in schizophrenia and the concurrent effects on learning, cognitive control, and self-monitoring (Carter et al., 2001).

CONCLUSIONS AND FUTURE DIRECTIONS

Future studies may aim to directly compare social and non-social predictive processes in the brain and in behavior to further elucidate how prediction is linked with social cognitive processes. This article encourages a number of questions, with testable hypotheses, which could be addressed by future work. Firstly it is suggested that neural structures and networks involved in non-social predictive processing, such as those coding for the prediction error and the efference copy, are also essentially utilized in social decision-making, social learning, and social interaction. Meta-analyses of functional imaging studies comparing analogous social and non-social cognitive processes will give more insight into this. It is also proposed that further work be done to investigate the interaction between top-down social information and expectations, non-social top-down priors and bottom-up sensory input, and the competition between these. Behavioral experiments comparing interference between levels of social and non-social processing, and computational hierarchical models would be useful to investigate such questions. This may also reveal different degrees of prioritization of information processing according to social versus non-social categorization of information. Pathological conditions associated with impairments in social functioning may serve as appropriate models to explore the independence of social cognition from predictive processing. Fletcher and Frith (2009) present a Bayesian framework for explaining the positive symptoms of schizophrenia and the potential for dysfunctions of underlying predictive mechanisms, such as the prediction error, corollary discharge, and efference copy, being at the core of such symptoms. If distinct patterns of impairment in predictive cognitive processes exist in such psychopathologies, it can be hoped that these can be therapeutically targeted to concurrently improve social functioning, and also to reduce other pathological symptoms related to predictive processing. Social interaction requires the understanding of others' beliefs, intentions, and emotions, which are formed from our own internal representations, and predictions, of others' mental states, though it is still under debate as to where and how internal representations of other social agents are represented in the brain. The use of a predictive coding framework provides conceptual scaffolding to bridge different domains of cognition and different research disciplines. In exploring social neuroscience under the guise of prediction, a more integrated and inclusive approach is permitted to understand the brain as a whole and not just a sum of its parts.

ACKNOWLEDGMENTS

The author Elliot C. Brown would like to thank Cumhur Tas and Burak Erdeniz for their invaluable comments and inspiration for this article. The authors would also like to thank the reviewers for their helpful, constructive, and supportive comments that allowed the authors to bring more clarity and focus to the article.

REFERENCES

Abu-Akel, A., and Shamay-Tsoory, S. (2011). Neuroanatomical and neurochemical bases of theory of mind. *Neuropsychologia* 49, 2971–2984.

Allison, T., Puce, A., and McCarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends Cogn. Sci.* 4, 267–278.

Anscombe, G. E. M. (1957). *Intention.* Oxford: Blackwell.

Arnal, L. H., Wyart, V., and Giraud, A. L. (2011). Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. *Nat. Neurosci.* 14, 797–801.

Baker, C. L., Saxe, R. R., and Tenenbaum, J. B. (2011). "Bayesian theory of mind: modeling joint belief-desire attribution," in *Proceedings of the Thirty-Third Annual Conference of the Cognitive Science Society,* (Austin, TX), 2469–2474.

Baker, T. E., and Holroyd, C. B. (2011). Dissociated roles of the anterior cingulate cortex in reward and conflict processing as revealed by the feedback error-related negativity and N200. *Biol. Psychol.* 87, 25–34.

Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends Cogn. Sci.* 11, 280–289.

Bar, M., Aminoff, E., Mason, M., and Fenske, M. (2007). The units of thought. *Hippocampus* 17, 420–428.

Bar, M. (2009). The proactive brain: memory for predictions. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 1235–1243.

Barrett, L. F., and Bar, M. (2009). See it with feeling: affective predictions during object perception. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 1325–1334.

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 1281–1289.

Bastiaansen, J. A., Thioux, M., and Keysers, C. (2009). Evidence for mirror systems in emotions. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 2391–2404.

Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U., and Fehr, E. (2008). Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. *Neuron* 58, 639–650.

Baumgartner, T., Fischbacher, U., Feierabend, A., Lutz, K., and Fehr, E. (2009). The neural circuitry of a broken promise. *Neuron* 64, 756–770.

Beeney, J. E., Franklin, R. G. Jr., Levy, K. N., and Adams, R. B. Jr. (2011). I feel your pain: emotional closeness modulates neural responses to empathically experienced rejection. *Soc. Neurosci.* 6, 369–376.

Behrens, T. E., Hunt, L. T., Woolrich, M. W., and Rushworth, M. F. (2008). Associative learning of social value. *Nature* 456, 245–249.

Behrens, T. E., Hunt, L. T., and Rushworth, M. F. (2009). The computation of social behavior. *Science* 324, 1160–1164.

Berber, S. M. (1962). Conditioning through vicarious instigation. *Psychol. Rev.* 69, 450–466.

Berns, G. S., Capra, C. M., Moore, S., and Noussair, C. (2010). Neural mechanisms of the influence of popularity on adolescent ratings of music. *Neuroimage* 49, 2687–2696.

Blakemore, S. J., Goodbody, S. J., and Wolpert, D. M. (1998). Predicting the consequences of our own actions: the role of sensorimotor context estimation. *J. Neurosci.* 18, 7511–7518.

Bortoletto, M., Mattingley, J. B., and Cunnington, R. (2011). Action intentions modulate visual processing during action perception. *Neuropsychologia* 49, 2097–2104.

Boyd, R., and Richerson, P. J. (1988). The evolution of reciprocity in sizable groups. *J. Theor. Biol.* 132, 337–356.

Brass, M., and Heyes, C. (2005). Imitation: is cognitive neuroscience solving the correspondence problem? *Trends Cogn. Sci.* 9, 489–495.

Brass, M., Schmitt, R. M., Spengler, S., and Gergely, G. (2007). Investigating action understanding: inferential processes versus action simulation. *Curr. Biol.* 17, 2117–2121.

Brazil, I. A., Mars, R. B., Bulten, B. H., Buitelaar, J. K., Verkes, R. J., and De Bruijn, E. R. (2011). A neurophysiological dissociation between monitoring one's own and others' actions in psychopathy. *Biol. Psychiatry* 69, 693–699.

Brüne, M. (2005). "Theory of mind" in schizophrenia: a review of the literature. *Schizophr. Bull.* 31, 21–42.

Bubic, A., von Cramon, D. Y., and Schubotz, R. I. (2010). Prediction, cognition and the brain. *Front. Hum. Neurosci.* 4, 25. doi: 10.3389/fnhum.2010.00025

Buckner, R. L., and Carroll, D. C. (2007). Self-projection and the brain. *Trends Cogn. Sci.* 11, 49–57.

Burke, C. J., Tobler, P. N., Baddeley, M., and Schultz, W. (2010). Neural mechanisms of observational learning. *Proc. Natl. Acad. Sci. U.S.A.* 107, 14431–14436.

Calvo-Merino, B., Glaser, D. E., Grezes, J., Passingham, R. E., and Haggard, P. (2005). Action observation and acquired motor skills: an fMRI study with expert dancers. *Cereb. Cortex* 15, 1243–1249.

Calvo-Merino, B., Grezes, J., Glaser, D. E., Passingham, R. E., and Haggard, P. (2006). Seeing or doing? Influence of visual and motor familiarity in action observation. *Curr. Biol.* 16, 1905–1910.

Campbell-Meiklejohn, D. K., Bach, D. R., Roepstorff, A., Dolan, R. J., and Frith, C. D. (2010). How the opinion of others affects our valuation of objects. *Curr. Biol.* 20, 1165–1170.

Capa, R. L., Marshall, P. J., Shipley, T. F., Salesse, R. N., and Bouquet, C. A. (2011). Does motor interference arise from mirror system activation? the effect of prior visuo-motor practice on automatic imitation. *Psychol. Res.* 75, 152–157.

Cardini, F., Costantini, M., Galati, G., Romani, G. L., Ladavas, E., and Serino, A. (2011). Viewing one's own face being touched modulates tactile perception: an fMRI study. *J. Cogn. Neurosci.* 23, 503–513.

Carp, J., Halenar, M. J., Quandt, L. C., Sklar, A., and Compton, R. J. (2009). Perceived similarity and neural mirroring: evidence from vicarious error processing. *Soc. Neurosci.* 4, 85–96.

Carter, C. S., Macdonald, A. W. 3rd., Ross, L. L., and Stenger, V. A. (2001). Anterior cingulate cortex activity and impaired self-monitoring of performance in patients with schizophrenia: an event-related fMRI study. *Am. J. Psychiatry* 158, 1423–1428.

Catmur, C., Walsh, V., and Heyes, C. (2009). Associative sequence learning: the role of experience in the development of imitation and the mirror system. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 2369–2380.

Cattaneo, L., Barchiesi, G., Tabarelli, D., Arfeller, C., Sato, M., and Glenberg, A. M. (2011). One's motor performance predictably modulates the understanding of others' actions through adaptation of premotor visuo-motor neurons. *Soc. Cogn. Affect. Neurosci.* 6, 301–310.

Centelles, L., Assaiante, C., Nazarian, B., Anton, J. L., and Schmitz, C. (2011). Recruitment of both the mirror and the mentalizing networks when observing social interactions depicted by point-lights: a neuroimaging study. *PLoS ONE* 6:e15749. doi: 10.1371/journal.pone.0015749

Chang, L. J., Doll, B. B., Van 'T Wout, M., Frank, M. J., and Sanfey, A. G. (2010). Seeing is believing: trustworthiness as a dynamic belief. *Cogn. Psychol.* 61, 87–105.

Chartrand, T. L., and Bargh, J. A. (1999). The chameleon effect: the perception-behavior link and social interaction. *J. Pers. Soc. Psychol.* 76, 893–910.

Cheng, Y., Yang, C. Y., Lin, C. P., Lee, P. L., and Decety, J. (2008). The perception of pain in others suppresses somatosensory oscillations: a magnetoencephalography study. *Neuroimage* 40, 1833–1840.

Chong, T. T., Cunnington, R., Williams, M. A., Kanwisher, N., and Mattingley, J. B. (2008). fMRI adaptation reveals mirror neurons in human inferior parietal cortex. *Curr. Biol.* 18, 1576–1580.

Coles, M. G. (1989). Modern mind-brain reading: psychophysiology, physiology, and cognition. *Psychophysiology* 26, 251–269.

Cook, R., Bird, G., Lunser, G., Huck, S., and Heyes, C. (2012). Automatic imitation in a strategic context: players of rock-paper-scissors imitate opponents' gestures. *Proc. Biol. Sci.* 279, 780–786.

Cook, J., and Bird, G. (2011). Social attitudes differentially modulate imitation in adolescents and adults. *Exp. Brain Res.* 211, 601–612.

Cooper, J. C., Dunne, S., Furey, T., and O'Doherty, J. P. (2012). Human dorsal striatum encodes prediction errors during observational learning of instrumental actions. *J. Cogn. Neurosci.* 24, 106–118.

Corlett, P. R., Taylor, J. R., Wang, X. J., Fletcher, P. C., and Krystal, J. H. (2010). Toward a neurobiology of delusions. *Prog. Neurobiol.* 92, 345–369.

Cross, E. S., Kraemer, D. J., Hamilton, A. F., Kelley, W. M., and Grafton, S. T. (2009). Sensitivity of the action observation network to physical and observational learning. *Cereb. Cortex* 19, 315–326.

Daprati, E., Franck, N., Georgieff, N., Proust, J., Pacherie, E., Dalery, J., and Jeannerod, M. (1997). Looking for the agent: an investigation into consciousness of action and self-consciousness in schizophrenic patients. *Cognition* 65, 71–86.

Daunizeau, J., Den Ouden, H. E., Pessiglione, M., Kiebel, S. J., Stephan, K. E., and Friston, K. J. (2010). Observing the observer (I): meta-bayesian models of learning and decision-making. *PLoS ONE* 5:e15554. doi: 10.1371/journal.pone.0015554

Davies, M., and Stone, T. (1995). *Mental Simulation: Evaluations and Applications.* Oxford: Blackwell Publishers.

De Bruijn, E. R., Schubotz, R. I., and Ullsperger, M. (2007). An event-related potential study on the observation of erroneous everyday actions. *Cogn. Affect. Behav. Neurosci.* 7, 278–285.

De Bruijn, E. R., De Lange, F. P., Von Cramon, D. Y., and Ullsperger, M. (2009). When errors are rewarding. *J. Neurosci.* 29, 12183–12186.

De Bruijn, E. R., Mars, R. B., Bekkering, H., and Coles, M. G. (2012). Your mistake is my mistake ... or is it? behavioural adjustments following own and observed actions in cooperative and competitive contexts. *Q. J. Exp. Psychol. (Hove)* 65, 317–325.

De Lange, F. P., Roelofs, K., and Toni, I. (2007). Increased self-monitoring during imagined movements in conversion paralysis. *Neuropsychologia* 45, 2051–2058.

De Lange, F. P., Spronk, M., Willems, R. M., Toni, I., and Bekkering, H. (2008). Complementary systems for understanding action intentions. *Curr. Biol.* 18, 454–457.

De Vignemont, F., and Haggard, P. (2008). Action observation and execution: what is shared? *Soc. Neurosci.* 3, 421–433.

Decety, J. (1996). Do imagined and executed actions share the same neural substrate? *Brain Res. Cogn. Brain Res.* 3, 87–93.

Decety, J., and Ickes, W. (2009). *The Social Neuroscience of Empathy.* Cambridge, MA: MIT Press.

Decety, J., and Sommerville, J. A. (2003). Shared representations between self and other: a social cognitive neuroscience view. *Trends Cogn. Sci.* 7, 527–533.

Dehaene, S., Posner, M. I., and Tucker, D. M. (1994). Localization of a neural system for error detection and compensation. *Psychol. Sci.* 3, 303–305.

Delgado, M. R., Frank, R. H., and Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during

the trust game. *Nat. Neurosci.* 8, 1611–1618.

Delgado, M. R., Li, J., Schiller, D., and Phelps, E. A. (2008). The role of the striatum in aversive learning and aversive prediction errors. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 3787–3800.

Dinstein, I., Hasson, U., Rubin, N., and Heeger, D. J. (2007). Brain areas selective for both observed and executed movements. *J. Neurophysiol.* 98, 1415–1427.

Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., and Baron-Cohen, S. (1999). Gaze perception triggers reflexive visuospatial orienting. *Vis. Cogn.* 6, 509–540.

Engel, A., Burke, M., Fiehler, K., Bien, S., and Rosler, F. (2008). Motor learning affects visual movement perception. *Eur. J. Neurosci.* 27, 2294–2302.

Evans, S., Fleming, S. M., Dolan, R. J., and Averbeck, B. B. (2011). Effects of emotional preferences on value-based decision-making are mediated by mentalizing and not reward networks. *J. Cogn. Neurosci.* 23, 2197–2210.

Falkenstein, M., Hohnsbein, J., Hoormann, J., and Blanke, L. (1990). "Effects of errors in choice reaction tasks on the ERP under focused and divided attention," in *Psychophysiological Brain Research*, eds. C. H. M. Brunia, A. W. K. Gaillard, and A. Kok (Tilburg: Tilburg University Press), 192–195.

Fehr, E., and Camerer, C. F. (2007). Social neuroeconomics: the neural circuitry of social preferences. *Trends Cogn. Sci.* 11, 419–427.

Feinberg, I. (1978). Efference copy and corollary discharge: implications for thinking and its disorders. *Schizophr. Bull.* 4, 636–640.

Feldman, H., and Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Front. Hum. Neurosci.* 4:215. doi: 10.3389/fnhum.2010.00215

Ferri, F., Campione, G. C., Dalla Volta, R., Gianelli, C., and Gentilucci, M. (2011). Social requests and social affordances: how they affect the kinematics of motor sequences during interactions between conspecifics. *PLoS ONE* 6:e15855. doi: 10.1371/journal.pone.0015855

Fletcher, P. C., and Frith, C. (2009). Perceiving is believing: a bayesian approach to explaining the positive symptoms of schizophrenia. *Nat. Rev. Neurosci.* 10, 48–58.

Fliessbach, K., Weber, B., Trautner, P., Dohmen, T., Sunde, U., Elger, C. E., and Falk, A. (2007). Social comparison affects reward-related brain activity in the human ventral striatum. *Science* 318, 1305–1308.

Ford, J. M., and Mathalon, D. H. (2004). Electrophysiological evidence of corollary discharge dysfunction in schizophrenia during talking and thinking. *J. Psychiatr. Res.* 38, 37–46.

Forgiarini, M., Gallucci, M., and Maravita, A. (2011). Racism and the empathy for pain on our skin. *Front. Psychol.* 2:108. doi: 10.3389/fpsyg.2011.00108

Fourneret, P., Franck, N., Slachevsky, A., and Jeannerod, M. (2001). Self-monitoring in schizophrenia revisited. *Neuroreport* 12, 1203–1208.

Franck, N., Farrer, C., Georgieff, N., Marie-Cardine, M., Dalery, J., D'Amato, T., and Jeannerod, M. (2001). Defective recognition of one's own actions in patients with schizophrenia. *Am. J. Psychiatry* 158, 454–459.

Friesen, C. K., and Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychon. B Rev.* 5, 490–495.

Friston, K. (2005). A theory of cortical responses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 815–836.

Friston, K., Kilner, J., and Harrison, L. (2006). A free energy principle for the brain. *J. Physiol. Paris* 100, 70–87.

Friston, K. (2008). Hierarchical models in the brain. *PLoS Comput. Biol.* 4:e1000211. doi: 10.1371/journal.pcbi.1000211

Friston, K. J., Daunizeau, J., and Kiebel, S. J. (2009). Reinforcement learning or active inference? *PLoS ONE* 4:e6421. doi: 10.1371/journal.pone.0006421

Friston, K., and Kiebel, S. (2009). Cortical circuits for perceptual inference. *Neural Netw.* 22, 1093–1104.

Friston, K. J., Daunizeau, J., Kilner, J., and Kiebel, S. J. (2010). Action and behavior: a free-energy formulation. *Biol. Cybern.* 102, 227–260.

Friston, K., Mattout, J., and Kilner, J. (2011). Action understanding and active inference. *Biol. Cybern.* 104, 137–160.

Furl, N., Van Rijsbergen, N. J., Treves, A., Friston, K. J., and Dolan, R. J. (2007). Experience-dependent coding of facial expression in superior temporal sulcus. *Proc. Natl. Acad. Sci. U.S.A.* 104, 13485–13489.

Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain* 119(Pt 2), 593–609.

Gazzola, V., and Keysers, C. (2009). The observation and execution of actions share motor and somatosensory voxels in all tested subjects: single-subject analyses of unsmoothed fMRI data. *Cereb. Cortex* 19, 1239–1255.

Gazzola, V., Rizzolatti, G., Wicker, B., and Keysers, C. (2007). The anthropomorphic brain: the mirror neuron system responds to human and robotic actions. *Neuroimage* 35, 1674–1684.

Gentsch, A., and Schutz-Bosbach, S. (2011). I did it: unconscious expectation of sensory consequences modulates the experience of self-agency and its functional signature. *J. Cogn. Neurosci.* 23, 3817–3828.

Gerull, F. C., and Rapee, R. M. (2002). Mother knows best: effects of maternal modelling on the acquisition of fear and avoidance behaviour in toddlers. *Behav. Res. Ther.* 40, 279–287.

Gilbert, D. T., and Wilson, T. D. (2007). Prospection: experiencing the future. *Science* 317, 1351–1354.

Gobbini, M. I., and Haxby, J. V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia* 45, 32–41.

Guo, X., Zheng, L., Zhang, W., Zhu, L., Li, J., Wang, Q., Dienes, Z., and Yang, Z. (2011). Empathic neural responses to others' pain depend on monetary reward. *Soc. Cogn. Affect. Neurosci.* doi: 10.1093/scan/nsr034. [Epub ahead of print].

Goubert, L., Vlaeyen, J. W., Crombez, G., and Craig, K. D. (2011). Learning about pain from others: an observational learning account. *J. Pain* 12, 167–174.

Graf, M., Reitzner, B., Corves, C., Casile, A., Giese, M., and Prinz, W. (2007). Predicting point-light actions in real-time. *Neuroimage* 36(Suppl. 2), T22–T32.

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., and Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron* 44, 389–400.

Greene, D. J., Mooshagian, E., Kaplan, J. T., Zaidel, E., and Iacoboni, M. (2009). The neural correlates of social attention: automatic orienting to social and nonsocial cues. *Psychol. Res.* 73, 499–511.

Grezes, J., and Decety, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: a meta-analysis. *Hum. Brain Mapp.* 12, 1–19.

Grezes, J., Armony, J. L., Rowe, J., and Passingham, R. E. (2003). Activations related to "mirror" and "canonical" neurones in the human brain: an fMRI study. *Neuroimage* 18, 928–937.

Gruetzmacher, N., Panzer, S., Blandin, Y., and Shea, C. H. (2011). Observation and physical practice: coding of simple motor sequences. *Q. J. Exp. Psychol. (Hove)* 64, 1111–1123.

Han, S., Fan, Y., Xu, X., Qin, J., Wu, B., Wang, X., Aglioti, S. M., and Mao, L. (2009). Empathic neural responses to others' pain are modulated by emotional contexts. *Hum. Brain Mapp.* 30, 3227–3237.

Harbaugh, W. T., Mayr, U., and Burghart, D. R. (2007). Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science* 316, 1622–1625.

Hare, R. D. (1980). A research scale for the assessment of psychopathy in criminal populations. *Pers. Indiv. Differ.* 1, 111–119.

Hare, T. A., Camerer, C. F., Knoepfle, D. T., and Rangel, A. (2010). Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. *J. Neurosci.* 30, 583–590.

Harris, L. T., and Fiske, S. T. (2010). Neural regions that underlie reinforcement learning are also active for social expectancy violations. *Soc. Neurosci.* 5, 76–91.

Haruno, M., Wolpert, D. M., and Kawato, M. (2001). Mosaic model for sensorimotor learning and control. *Neural Comput.* 13, 2201–2220.

Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends Cogn. Sci.* 4, 223–233.

Hayes, S. J., Elliott, D., and Bennett, S. J. (2010). General motor representations are developed during action-observation. *Exp. Brain Res.* 204, 199–206.

He, X., Lever, A. G., and Humphreys, G. W. (2011). Interpersonal memory-based guidance of attention is reduced for ingroup members. *Exp. Brain Res.* 211, 429–438.

Helsen, K., Goubert, L., Peters, M. L., and Vlaeyen, J. W. (2011). Observational learning and pain-related fear: an experimental study with colored cold pressor tasks. *J. Pain* 12, 1230–1239.

Heyes, C. (2001). Causes and consequences of imitation. *Trends Cogn. Sci.* 5, 253–261.

Heyes, C. (2011). What's social about social learning? *J. Comp. Psychol.* doi: 10.1037/a0025180. [Epub ahead of print].

Heyes, C. M., and Ray, E. D. (2000). What is the significance of imitation in animals? *Adv. Stud. Behav.* 29, 215–245.

Hickok, G. (2009). Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *J. Cogn. Neurosci.* 21, 1229–1243.

Hoffmann, S., and Falkenstein, M. (2011). Predictive information processing in the brain: errors and response monitoring. *Int. J. Psychophysiol.* 83, 208–212.

Hollander, A., Jung, C., and Prinz, W. (2011). Covert motor activity on NoGo trials in a task sharing paradigm: evidence from the lateralized readiness potential. *Exp. Brain Res.* 211, 345–356.

Holroyd, C. B., and Coles, M. G. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychol. Rev.* 109, 679–709.

Holroyd, C. B., Nieuwenhuis, S., Yeung, N., and Cohen, J. D. (2003). Errors in reward prediction are reflected in the event-related brain potential. *Neuroreport* 14, 2481–2484.

Hommel, B., Musseler, J., Aschersleben, G., and Prinz, W. (2001). The Theory of event coding (TEC): a framework for perception and action planning. *Behav. Brain Sci.* 24, 849–878. discussion 878–937.

Hosp, J. A., Pekanovic, A., Rioult-Pedotti, M. S., and Luft, A. R. (2011). Dopaminergic projections from midbrain to primary motor cortex mediate motor skill learning. *J. Neurosci.* 31, 2481–2487.

Humphreys, G. W., and Bedford, J. (2011). The relations between joint action and theory of mind: a neuropsychological analysis. *Exp. Brain Res.* 211, 357–369.

Hygge, S., and Ohman, A. (1978). Modeling processes in the acquisition of fears: vicarious electrodermal conditioning to fear-relevant stimuli. *J. Pers. Soc. Psychol.* 36, 271–279.

Iacoboni, M. (2009). Imitation, empathy, and mirror neurons. *Annu. Rev. Psychol.* 60, 653–670.

Imamizu, H., and Kawato, M. (2008). Neural correlates of predictive and postdictive switching mechanisms for internal models. *J. Neurosci.* 28, 10751–10765.

Imamizu, H., and Kawato, M. (2009). Brain mechanisms for predictive control by switching internal models: implications for higher-order cognitive functions. *Psychol. Res.* 73, 527–544.

Ishai, A. (2008). Let's face it: it's a cortical network. *Neuroimage* 40, 415–419.

Itagaki, S., and Katayama, J. (2008). Self-relevant criteria determine the evaluation of outcomes induced by others. *Neuroreport* 19, 383–387.

Jackson, P. L., Brunet, E., Meltzoff, A. N., and Decety, J. (2006). Empathy examined through the neural mechanisms involved in imagining how I feel versus how you feel pain. *Neuropsychologia* 44, 752–761.

Jakobs, O., Wang, L. E., Dafotakis, M., Grefkes, C., Zilles, K., and Eickhoff, S. B. (2009). Effects of timing and movement uncertainty implicate the temporo-parietal junction in the prediction of forthcoming motor actions. *Neuroimage* 47, 667–677.

Jellema, T., Baker, C. I., Wicker, B., and Perrett, D. I. (2000). Neural representation for the perception of the intentionality of actions. *Brain Cogn.* 44, 280–302.

John, E. R., Chesler, P., Bartlett, F., and Victor, I. (1968). Observation learning in cats. *Science* 159, 1489–1491.

Jones, R. M., Somerville, L. H., Li, J., Ruberry, E. J., Libby, V., Glover, G., Voss, H. U., Ballon, D. J., and Casey, B. J. (2011). Behavioral and neural properties of social reinforcement learning. *J. Neurosci.* 31, 13039–13045.

Kang, S. K., Hirsh, J. B., and Chastee, A. L. (2010). Your mistakes are mine: self-other overlap predicts neural response to observed errors. *J. Exp. Soc. Psychol.* 4, 229–232.

Kapogiannis, D., Campion, P., Grafman, J., and Wassermann, E. M. (2008). Reward-related activity in the human motor cortex. *Eur. J. Neurosci.* 27, 1836–1842.

Kavaliers, M., Choleris, E., and Colwell, D. D. (2001). Learning from others to cope with biting flies: social learning of fear-induced conditioned analgesia and active avoidance. *Behav. Neurosci.* 115, 661–674.

Keysers, C., Kaas, J. H., and Gazzola, V. (2010). Somatosensation in social perception. *Nat. Rev. Neurosci.* 11, 417–428.

Kiebel, S. J., Von Kriegstein, K., Daunizeau, J., and Friston, K. J. (2009). Recognizing sequences of sequences. *PLoS Comput. Biol.* 5:e1000464. doi: 10.1371/journal.pcbi.1000464

Kilner, J. M., Vargas, C., Duval, S., Blakemore, S. J., and Sirigu, A. (2004). Motor activation prior to observation of a predicted movement. *Nat. Neurosci.* 7, 1299–1301.

Kilner, J. M., Friston, K. J., and Frith, C. D. (2007a). The mirror-neuron system: a bayesian perspective. *Neuroreport* 18, 619–623.

Kilner, J. M., Friston, K. J., and Frith, C. D. (2007b). Predictive coding: an account of the mirror neuron system. *Cogn. Process.* 8, 159–166.

Kilner, J. M., and Frith, C. D. (2008). Action observation: inferring intentions without mirror neurons. *Curr. Biol.* 18, R32–R33.

King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., and Montague, P. R. (2005). Getting to know you: reputation and trust in a two-person economic exchange. *Science* 308, 78–83.

Kim, B. R., Liss, A., Rao, M., Singer, Z., and Compton, R. J. (2011). Social deviance activates the brain's error-monitoring system. *Cogn. Affect. Behav. Neurosci.* 12, 65–73.

Kim, H., Shimojo, S., and O'Doherty, J. P. (2006). Is avoiding an aversive outcome rewarding? neural substrates of avoidance learning in the human brain. *PLoS Biol.* 4:e233. doi: 10.1371/journal.pbio.0040233

Klucharev, V., Hytonen, K., Rijpkema, M., Smidts, A., and Fernandez, G. (2009). Reinforcement learning signal predicts social conformity. *Neuron* 61, 140–151.

Koban, L., Pourtois, G., Vocat, R., and Vuilleumier, P. (2010). When your errors make me lose or win: event-related potentials to observed errors of cooperators and competitors. *Soc. Neurosci.* 5, 360–374.

Koelewijn, T., Van Schie, H. T., Bekkering, H., Oostenveld, R., and Jensen, O. (2008). Motor-cortical beta oscillations are modulated by correctness of observed action. *Neuroimage* 40, 767–775.

Konvalinka, I., Vuust, P., Roepstorff, A., and Frith, C. D. (2010). Follow you, follow me: continuous mutual prediction and adaptation in joint tapping. *Q. J. Exp. Psychol. (Hove)* 63, 2220–2230.

Konvalinka, I., Xygalatas, D., Bulbulia, J., Schjodt, U., Jegindo, E. M., Wallot, S., Van Orden, G., and Roepstorff, A. (2011). Synchronized arousal between performers and related spectators in a fire-walking ritual. *Proc. Natl. Acad. Sci. U.S.A.* 108, 8514–8519.

Kording, K. P., and Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature* 427, 244–247.

Knapska, E., Mikosz, M., Werka, T., and Maren, S. (2010). Social modulation of learning in rats. *Learn. Mem.* 17, 35–42.

Knoblich, G., Stottmeister, F., and Kircher, T. (2004). Self-monitoring in patients with schizophrenia. *Psychol. Med.* 34, 1561–1569.

Krajbich, I., Adolphs, R., Tranel, D., Denburg, N. L., and Camerer, C. F. (2009). Economic games quantify diminished sense of guilt in patients with damage to the prefrontal cortex. *J. Neurosci.* 29, 2188–2192.

Kühn, S., Muller, B. C., Van Der Leij, A., Dijksterhuis, A., Brass, M., and Van Baaren, R. B. (2011). Neural correlates of emotional synchrony. *Soc. Cogn. Affect. Neurosci.* 6, 368–374.

Lanzetta, J. T., and Englis, B. G. (1989). Expectations of cooperation and competition and their effects on observers' vicarious emotional responses. *J. Pers. Soc. Psychol.* 56, 543–554.

Lawrence, E. J., Shaw, P., Giampietro, V. P., Surguladze, S., Brammer, M. J., and David, A. S. (2006). The role of 'shared representations' in social perception and empathy: an fMRI study. *Neuroimage* 29, 1173–1184.

Li, J., Xiao, E., Houser, D., and Montague, P. R. (2009). Neural responses to sanction threats in two-party economic exchange. *Proc. Natl. Acad. Sci. U.S.A.* 106, 16835–16840.

Luft, A. R., and Schwarz, S. (2009). Dopaminergic signals in primary motor cortex. *Int. J. Dev. Neurosci.* 27, 415–421.

Ma, Q., Shen, Q., Xu, Q., Li, D., Shu, L., and Weber, B. (2011). Empathic responses to others' gains and losses: an electrophysiological investigation. *Neuroimage* 54, 2472–2480.

Manera, V., Becchio, C., Schouten, B., Bara, B. G., and Verfaillie, K. (2011). Communicative interactions improve visual detection of biological motion. *PLoS ONE* 6:e14594. doi: 10.1371/journal.pone.0014594

Marco-Pallares, J., Kramer, U. M., Strehl, S., Schroder, A., and Munte, T. F. (2010). When decisions of others matter to me: an electrophysiological analysis. *BMC Neurosci.* 11, 86.

Massen, C., and Prinz, W. (2009). Movements, actions and tool-use actions: an ideomotor approach to imitation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 2349–2358.

Mayr, E. (1974). Behavior programs and evolutionary strategies. *Am. Sci.* 62, 650–659.

Mazzola, V., Latorre, V., Petito, A., Gentili, N., Fazio, L., Popolizio, T., Blasi, G., Arciero, G., and Bondolfi, G. (2010). Affective response to a loved one's pain: insula activity as a function of individual

differences. *PLoS ONE* 5:e15268. doi: 10.1371/journal.pone.0015268
McNally, G. P., Johansen, J. P., and Blair, H. T. (2011). Placing prediction into the fear circuit. *Trends Neurosci.* 34, 283–292.
Meissner, C. A., and Brigham, J. C. (2001). Thirty years of investigating the own-race advantage in memory for faces: a meta-analytic review. *Psychol. Public Policy Law* 7, 3–35.
Meulders, A., Vansteenwegen, D., and Vlaeyen, J. W. (2011). The acquisition of fear of movement-related pain and associative learning: a novel pain-relevant human fear conditioning paradigm. *Pain* 152, 2460–2469.
Miles, L. K., Nind, L. K., and Macrae, C. N. (2010). Moving through time. *Psychol. Sci.* 21, 222–223.
Miles, L. K., Lumsden, J., Richardson, M. J., and Neil Macrae, C. (2011). Do birds of a feather move together? group membership and behavioral synchrony. *Exp. Brain Res.* 211, 495–503.
Mineka, S., Davidson, M., Cook, M., and Keir, R. (1984). Observational conditioning of snake fear in rhesus monkeys. *J. Abnorm. Psychol.* 93, 355–372.
Mineka, S., and Cook, M. (1993). Mechanisms involved in the observational conditioning of fear. *J. Exp. Psychol. Gen.* 122, 23–38.
Mitchell, J. P. (2009). Inferences about mental states. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 1309–1316.
Molenberghs, P., Brander, C., Mattingley, J. B., and Cunnington, R. (2010). The role of the superior temporal sulcus and the mirror neuron system in imitation. *Hum. Brain Mapp.* 31, 1316–1326.
Molenberghs, P., Cunnington, R., and Mattingley, J. B. (2009). Is the mirror neuron system involved in imitation? a short review and meta-analysis. *Neurosci. Biobehav. Rev.* 33, 975–980.
Molenberghs, P., Cunnington, R., and Mattingley, J. B. (2012). Brain regions with mirror properties: a meta-analysis of 125 human fMRI studies. *Neurosci. Biobehav. Rev.* 36, 341–349.
Moll, J., Zahn, R., De Oliveira-Souza, R., Krueger, F., and Grafman, J. (2005). Opinion: the neural basis of human moral cognition. *Nat. Rev. Neurosci.* 6, 799–809.
Mori, M. (1970). The uncanny valley. *Energy* 7, 33–35.
Mukamel, R., Ekstrom, A. D., Kaplan, J., Iacoboni, M., and Fried, I. (2010). Single-neuron responses in humans during execution and observation of actions. *Curr. Biol.* 20, 750–756.
Muller, B. C., Kuhn, S., Van Baaren, R. B., Dotsch, R., Brass, M., and Dijksterhuis, A. (2011). Perspective taking eliminates differences in co-representation of out-group members' actions. *Exp. Brain Res.* 211, 423–428.
Mundy, P., Sigman, M., Ungerer, J., and Sherman, T. (1986). Defining the social deficits of autism: the contribution of non-verbal communication measures. *J. Child Psychol. Psychiatry* 27, 657–669.
Munksgaard, L., Depassille, A. M., Rushen, J., Herskin, M. S., and Kristensen, A. M. (2001). Dairy cows' fear of people: social learning, milk yield and behaviour at milking. *Appl. Anim. Behav. Sci.* 73, 15–26.
Nawa, N. E., Nelson, E. E., Pine, D. S., and Ernst, M. (2008). Do you make a difference? social context in a betting task. *Soc. Cogn. Affect. Neurosci.* 3, 367–376.
Neri, P., Luu, J. Y., and Levi, D. M. (2006). Meaningful interactions can enhance visual discrimination of human agents. *Nat. Neurosci.* 9, 1186–1192.
Newman-Norlund, R. D., Ganesh, S., Van Schie, H. T., De Bruijn, E. R., and Bekkering, H. (2009). Self-identification and empathy modulate error-related brain activity during the observation of penalty shots between friend and foe. *Soc. Cogn. Affect. Neurosci.* 4, 10–22.
Nieuwenhuis, S., Heslenfeld, D. J., Von Geusau, N. J., Mars, R. B., Holroyd, C. B., and Yeung, N. (2005). Activity in human reward-sensitive brain areas is strongly context dependent. *Neuroimage* 25, 1302–1309.
Niv, Y., Daw, N. D., and Dayan, P. (2006). Choice values. *Nat. Neurosci.* 9, 987–988.
Nunez Castellar, E., Notebaert, W., Van den Bossche, L., and Fias, W. (2011). How monitoring other's actions influences one's own performance. *Exp. Psychol.* 58, 499–508.
Oberman, L. M., Hubbard, E. M., Mccleery, J. P., Altschuler, E. L., Ramachandran, V. S., and Pineda, J. A. (2005). EEG evidence for mirror neuron dysfunction in autism spectrum disorders. *Brain Res. Cogn. Brain Res.* 24, 190–198.
Olsson, A., Nearing, K. I., and Phelps, E. A. (2007). Learning fears by observing others: the neural systems of social fear transmission. *Soc. Cogn. Affect. Neurosci.* 2, 3–11.
Orgs, G., Dombrowski, J. H., Heil, M., and Jansen-Osmann, P. (2008). Expertise in dance modulates alpha/beta event-related desynchronization during action observation. *Eur. J. Neurosci.* 27, 3380–3384.
O'Toole, A. J., Deffenbacher, K. A., Valentin, D., and Abdi, H. (1994). Structural aspects of face recognition and the other-race effect. *Mem. Cognit.* 22, 208–224.
Oztop, E., Wolpert, D., and Kawato, M. (2005). Mental state inference using visual control parameters. *Brain Res. Cogn. Brain Res.* 22, 129–151.
Pecenka, N., and Keller, P. E. (2011). The role of temporal prediction abilities in interpersonal sensorimotor synchronization. *Exp. Brain Res.* 211, 505–515.
Perrett, D. I., and Emery, N. J. (1994). Understanding the intentions of others from visual signals: neurophysiological evidence. *Curr. Psychol. Cogn.* 13, 683–694.
Perry, A., Bentin, S., Bartal, I. B., Lamm, C., and Decety, J. (2010). "Feeling" the pain of those who are different from us: modulation of EEG in the mu/alpha range. *Cogn. Affect. Behav. Neurosci.* 10, 493–504.
Prinz, W. (1997). Perception and action planning. *Eur. J. Cogn. Psychol.* 9, 129–154.
Radke, S., De Lange, F. P., Ullsperger, M., and De Bruijn, E. R. (2011). Mistakes that affect others: an fMRI study on processing of own errors in a social context. *Exp. Brain Res.* 211, 405–413.
Rao, R. P., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87.
Raymaekers, R., Wiersema, J. R., and Roeyers, H. (2009). EEG study of the mirror neuron system in children with high functioning autism. *Brain Res.* 1304, 113–121.
Rhodes, G., Tan, S., Brake, S., and Taylor, K. (1989). Expertise and configural coding in face recognition. *Br. J. Psychol.* 80(Pt 3), 313–331.
Rigoni, D., Polezzi, D., Rumiati, R., Guarino, R., and Sartori, G. (2010). When people matter more than money: an ERPs study. *Brain Res. Bull.* 81, 445–452.
Rilling, J., Gutman, D., Zeh, T., Pagnoni, G., Berns, G., and Kilts, C. (2002). A neural basis for social cooperation. *Neuron* 35, 395–405.
Rilling, J. K., Sanfey, A. G., Aronson, J. A., Nystrom, L. E., and Cohen, J. D. (2004). Opposing BOLD responses to reciprocated and unreciprocated altruism in putative reward pathways. *Neuroreport* 15, 2539–2543.
Rushworth, M. F., Mars, R. B., and Summerfield, C. (2009). General mechanisms for making decisions? *Curr. Opin. Neurobiol.* 19, 75–83.
Sallet, J., Mars, R. B., Noonan, M. P., Andersson, J. L., O'Reilly, J. X., Jbabdi, S., Croxson, P. L., Jenkinson, M., Miller, K. L., and Rushworth, M. F. (2011). Social network size affects neural circuits in macaques. *Science* 334, 697–700.
Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., and Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science* 300, 1755–1758.
Sartori, L., Becchio, C., Bara, B. G., and Castiello, U. (2009a). Does the intention to communicate affect action kinematics? *Conscious. Cogn.* 18, 766–772.
Sartori, L., Becchio, C., Bulgheroni, M., and Castiello, U. (2009b). Modulation of the action control system by social intention: unexpected social requests override preplanned action. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 1490–1500.
Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., and Frith, C. (2012). The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Soc. Cogn. Affect. Neurosci.* 7, 413–422.
Schacter, D. L., and Addis, D. R. (2007). The cognitive neuroscience of constructive memory: remembering the past and imagining the future. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 362, 773–786.
Schlicht, E. J., Shimojo, S., Camerer, C. F., Battaglia, P., and Nakayama, K. (2010). Human wagering behavior depends on opponents' faces. *PLoS ONE* 5:e11663. doi: 10.1371/journal.pone.0011663
Schober, H., and Rentschler, I. (1972). *Optische Täuschungen in Wissenschaft und Kunst*. München: Heinz Moos Verlag.
Schubotz, R. I., and von Cramon, D. Y. (2002). Predicting perceptual events activates corresponding motor schemes in lateral premotor cortex: an fMRI study. *Neuroimage* 15, 787–796.
Schultz, W., and Dickinson, A. (2000). Neuronal coding of prediction errors. *Annu. Rev. Neurosci.* 23, 473–500.
Scully, D. M., and Newell, K. M. (1985). Observational learning and the acquisition of motor skills: toward a visual perception perspective. *J. Hum. Mov. Stud.* 11, 169–186.
Sebanz, N., Bekkering, H., and Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends Cogn. Sci.* 10, 70–76.

Serino, A., Giovagnoli, G., and Ladavas, E. (2009). I feel what you feel if you are similar to me. *PLoS ONE* 4:e4930. doi: 10.1371/journal.pone.0004930

Serino, A., Pizzoferrato, F., and Ladavas, E. (2008). Viewing a face (especially one's own face) being touched enhances tactile perception on the face. *Psychol. Sci.* 19, 434–438.

Sescousse, G., Redoute, J., and Dreher, J. C. (2010). The architecture of reward value coding in the human orbitofrontal cortex. *J. Neurosci.* 30, 13095–13104.

Seymour, B., Yoshida, W., and Dolan, R. (2009). Altruistic learning. *Front. Behav. Neurosci.* 3:23. doi: 10.3389/neuro.08.023.2009

Shane, M. S., Stevens, M., Harenski, C. L., and Kiehl, K. A. (2008). Neural correlates of the processing of another's mistakes: a possible underpinning for social and observational learning. *Neuroimage* 42, 450–459.

Shane, M. S., Stevens, M. C., Harenski, C. L., and Kiehl, K. A. (2009). Double dissociation between perspective-taking and empathic-concern as predictors of hemodynamic response to another's mistakes. *Soc. Cogn. Affect. Neurosci.* 4, 111–118.

Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., and Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature* 439, 466–469.

Singh, F., Pineda, J., and Cadenhead, K. S. (2011). Association of impaired EEG mu wave suppression, negative symptoms and social functioning in biological motion processing in first episode of psychosis. *Schizophr. Res.* 130, 182–186.

Sinigaglia, C., and Rizzolatti, G. (2011). Through the looking glass: self and others. *Conscious. Cogn.* 20, 64–74.

Spengler, S., Brass, M., Kuhn, S., and Schutz-Bosbach, S. (2010). Minimizing motor mimicry by myself: self-focus enhances online action-control mechanisms during motor contagion. *Conscious. Cogn.* 19, 98–106.

Sperry, R. W. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *J. Cogn. Neurosci.* 43, 482–489.

Spoormaker, V. I., Andrade, K. C., Schroter, M. S., Sturm, A., Goya-Maldonado, R., Samann, P. G., and Czisch, M. (2011). The neural correlates of negative prediction error signaling in human fear conditioning. *Neuroimage* 54, 2250–2256.

Stahl, J., Wiese, H., and Schweinberger, S. R. (2010). Learning task affects ERP-correlates of the own-race bias, but not recognition memory performance. *Neuropsychologia* 48, 2027–2040.

Stephens, G. J., Silbert, L. J., and Hasson, U. (2010). Speaker-listener neural coupling underlies successful communication. *Proc. Natl. Acad. Sci. U.S.A.* 107, 14425–14430.

Suddendorf, T., Addis, D. R., and Corballis, M. C. (2009). Mental time travel and the shaping of the human mind. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 1317–1324.

Tabibnia, G., Satpute, A. B., and Lieberman, M. D. (2008). The sunny side of fairness: preference for fairness activates reward circuitry (and disregarding unfairness activates self-control circuitry). *Psychol. Sci.* 19, 339–347.

Takahashi, H., Kato, M., Matsuura, M., Mobbs, D., Suhara, T., and Okubo, Y. (2009). When your gain is my pain and your pain is my gain: neural correlates of envy and schadenfreude. *Science* 323, 937–939.

Thabit, M. N., Nakatsuka, M., Koganemaru, S., Fawi, G., Fukuyama, H., and Mima, T. (2011). Momentary reward induce changes in excitability of primary motor cortex. *Clin. Neurophysiol.* 122, 1764–1770.

Thibaut, J. W., and Kelley, H. H. (1959). *The Social Psychology of Groups.* London, UK: Wiley.

Tinbergen, N. (1951). *The Study of Instinct.* New York, NY: Oxford University Press.

Tipper, C. M., Handy, T. C., Giesbrecht, B., and Kingstone, A. (2008). Brain responses to biological relevance. *J. Cogn. Neurosci.* 20, 879–891.

Tsakiris, M. (2008). Looking for myself: current multisensory input alters self-face recognition. *PLoS ONE* 3:e4040. doi: 10.1371/journal.pone.0004040

Turella, L., Erb, M., Grodd, W., and Castiello, U. (2009). Visual features of an observed agent do not modulate human brain activity during action observation. *Neuroimage* 46, 844–853.

Turner, M. S., Simons, J. S., Gilbert, S. J., Frith, C. D., and Burgess, P. W. (2008). Distinct roles for lateral and medial rostral prefrontal cortex in source monitoring of perceived and imagined events. *Neuropsychologia* 46, 1442–1453.

Valdesolo, P., and Desteno, D. (2011). Synchrony and the social tuning of compassion. *Emotion* 11, 262–266.

Valdesolo, P., Ouyang, J., DeSteno, D. (2010). The rhythm of joint action: synchrony promotes cooperative behavior. *J. Exp. Soc. Psychol.* 46, 693–695.

Van den Bos, E., and Jeannerod, M. (2002). Sense of body and sense of action both contribute to self-recognition. *Cognition* 85, 177–187.

Van Meel, C. S., and Van Heijningen, C. A. (2010). The effect of interpersonal competition on monitoring internal and external error feedback. *Psychophysiology* 47, 213–222.

Van Schie, H. T., Mars, R. B., Coles, M. G., and Bekkering, H. (2004). Modulation of activity in medial frontal and motor cortices during error observation. *Nat. Neurosci.* 7, 549–554.

Van 'T Wout, M., and Sanfey, A. G. (2008). Friend or foe: the effect of implicit trustworthiness judgments in social decision-making. *Cognition* 108, 796–803.

von Holst, E., and Mittelstaedt, H. (1950). Das Reafferenzprinzip (Wechselwirkungen zwischen Zentralnervensystem und Peripherie). *Naturwissenschaften* 37, 464–476.

Walker, P. M., and Tanaka, J. W. (2003). An encoding advantage for own-race versus other-race faces. *Perception* 32, 1117–1125.

Walter, W. G., Cooper, R., Aldridge, V. J., McCallum, W. C., and Winter, A. L. (1964). Contingent negative variation: an electric sign of sensorimotor association and expectancy in the human brain. *Nature* 203, 380–384.

Webster, M. A., Kaping, D., Mizokami, Y., and Duhamel, P. (2004). Adaptation to natural facial categories. *Nature* 428, 557–561.

Wenke, D., Atmaca, S., Holländer, A., Baess, P., Liepelt, R., and Prinz, W. (2011). What is shared in joint action? Issues of co-representation, response conflict, and agent identification. *Rev. Phil. Psych.* 2, 147–172.

Williams, J. M., Ellis, N. C., Tyers, C., Healy, H., Rose, G., and Macleod, A. K. (1996). The specificity of autobiographical memory and imageability of the future. *Mem. Cognit.* 24, 116–125.

Wiltermuth, S. S., and Heath, C. (2009). Synchrony and cooperation. *Psychol. Sci.* 20, 1–5.

Wolpert, D. M., and Miall, R. C. (1996). Forward models for physiological motor control. *Neural Netw.* 9, 1265–1279.

Wolpert, D. M., Miall, R. C., and Kawato, M. (1998). Internal models in the cerebellum. *Trends Cogn. Sci.* 2, 338–347.

Wolpert, D. M., Doya, K., and Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358, 593–602.

Yomogida, Y., Sugiura, M., Sassa, Y., Wakusawa, K., Sekiguchi, A., Fukushima, A., Takeuchi, H., Horie, K., Sato, S., and Kawashima, R. (2010). The neural basis of agency: an fMRI study. *Neuroimage* 50, 198–207.

Yoshida, W., Seymour, B., Friston, K. J., and Dolan, R. J. (2010). Neural mechanisms of belief inference during cooperative games. *J. Neurosci.* 30, 10744–10751.

Yu, R., and Zhou, X. (2006). Brain responses to outcomes of one's own and other's performance in a gambling task. *Neuroreport* 17, 1747–1751.

Zaki, J., Schirmer, J., and Mitchell, J. P. (2011). Social influence modulates the neural computation of value. *Psychol. Sci.* 22, 894–900.

Zwickel, J., Grosjean, M., and Prinz, W. (2010a). What part of an action interferes with ongoing perception? *Acta Psychol. (Amst.)* 134, 403–409.

Zwickel, J., Grosjean, M., and Prinz, W. (2010b). On interference effects in concurrent perception and action. *Psychol. Res.* 74, 152–171.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 15 February 2012; accepted: 09 May 2012; published online: 24 May 2012.

Citation: Brown EC and Brüne M (2012) The role of prediction in social neuroscience. Front. Hum. Neurosci. ***6****:147. doi: 10.3389/fnhum.2012.00147*

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 11 June 2012
doi: 10.3389/fnhum.2012.00171

Personality and error monitoring: an update

Sven Hoffmann*, Edmund Wascher and Michael Falkenstein

Leibniz Research Centre for Working Environment and Human Factors, Dortmund, Germany

Edited by:
Christian Bellebaum, Ruhr University Bochum, Germany

Reviewed by:
Marco Steinhauser, University of Konstanz, Germany
Christian Bellebaum, Ruhr University Bochum, Germany

****Correspondence:***
Sven Hoffmann, Leibniz Research Centre for Working Environment and Human Factors, Ardeystr. 67, 44139 Dortmund, Germany.
e-mail: shoffmann@ifado.de

People differ considerably with respect to their ability to initiate and maintain cognitive control. A core control function is the processing and evaluation of errors from which we learn to prevent maladaptive behavior. People differ strongly in the degree of error processing, and how errors are interpreted and appraised. In the present study it was investigated whether a correlate of error monitoring, the error negativity (Ne or ERN), is related to personality factors. Therefore, the EEG was measured continuously during a task that provoked errors, and the Ne was tested with respect to its relation to personality traits. The results indicate a substantial trait-like relation of error processing and personality factors: the Ne was more pronounced for subjects scoring low on the "Openness" scale, the "Impulsiveness" scale and the "Emotionality" scale. Inversely, the Ne was less pronounced for subjects scoring low on the "Social Orientation" scale. The results implicate that personality traits related to emotional valences and rigidity are reflected in the way people monitor and adapt to erroneous actions.

Keywords: personality, response monitoring, error monitoring, error negativity, EEG

INTRODUCTION

From our everyday experience we know that the efficiency of monitoring behavior and coping with undesired outcomes of one's own actions differs considerably between subjects. Furthermore, subjects differ considerably with respect to how they interpret and appraise such outcomes. Especially the committing and processing of errors plays a crucial role with respect to the adequate adaptation of behavior. A correlate of response monitoring has caught considerable attention during the last two decades. Following erroneous responses in a typical choice reaction experiment a sharp negative deflection at fronto-central electrode position in a simultaneous EEG measurement can be observed: the error negativity (Ne, Falkenstein et al., 1990) or error-related negativity (ERN, Gehring et al., 1993). This finding supports the idea of a common response monitoring system, which is involved in monitoring and adapting behavior with respect to external or internal events. The sources of the Ne have been located repeatedly in the anterior cingulate cortex (ACC) (Dehaene et al., 1994; Debener et al., 2005; Hoffmann and Falkenstein, 2010), and it is assumed that the Ne is elicited by striatal dopamine projections (Holroyd and Coles, 2002).

Indeed, several neurophysiological studies indicate a relation of the Ne to the functioning of the dopamine system, for example studies which investigated schizophrenia (Mathalon et al., 2002), attention deficit disorder (Liotti et al., 2005), alcohol consumption (Holroyd and Yeung, 2003), or aging (Nieuwenhuis et al., 2002; Beste et al., 2009; Hoffmann and Falkenstein, 2011).

Most models about the functional role of the Ne assume that it reflects a detection mechanism of undesirable or unexpected events (Holroyd and Coles, 2002; Ridderinkhof et al., 2004; Alexander and Brown, 2011). The Ne is thought to reflect the degree by which an event (be it external or internal) is worse than expected Holroyd and Coles (2002), or unexpected (Alexander and Brown, 2011). A competing hypothesis, the conflict hypothesis suggests that the Ne is the result of a temporal overlap of competing responses. The higher the concordance between all activated possible responses, the higher the conflict and thus the larger is the Ne amplitude (Botvinick et al., 2001, 2004; van Veen et al., 2001). However, several more recent results contradict the conflict hypothesis (e.g., Carbonnell and Falkenstein, 2006; Masaki et al., 2007).

All hypotheses so far assume that the Ne is related to the detection of unexpected or undesirable or conflicting events. Such events need control to be avoided on-line or at least in the future (Hoffmann and Falkenstein, 2012). More recent studies suggest the Ne to reflect the activity of a common response monitoring system (Debener et al., 2005; Hoffmann and Falkenstein, 2010; Roger et al., 2010) evaluating state-goal discrepancies permanently and inducing cognitive control. This activity as reflected in the Ne appears to be integrated in pre-frontal cortex (PFC) functions dealing with the maintaining or execution of cognitive control. It is well known that the PFC, and more specifically the ACC, is central to inhibitory control (Braver et al., 2001). Also, ACC functioning has top down effects even on sensory processes (Sarter et al., 2006; Danielmeier et al., 2011). Thus, ACC and PFC appear to be central to the control of internal states (Devinsky et al., 1995).

Anyway, people differ considerably not only with respect to performance, but also with respect to how they are concerned with making errors. For example people suffering from obsessive compulsive disorder show an enhanced Ne, and its amplitude is related to the severity of symptoms (Gehring et al., 2000). Also, simply being more worried about committing errors is predictive for an enhanced Ne (Hajcak et al., 2003). Furthermore, the Ne appears to be related to negative affective experience (Hajcak et al., 2004). A reversed pattern can be observed in poorly

socialized individuals, here the Ne is reduced (Dikman and Allen, 2000).

Taking into account the core structures involved in response monitoring and learning, i.e., dopamine system, amygdala, insula, and PFC (de Bruijn et al., 2009; Gentsch et al., 2009; Jocham and Ullsperger, 2009; Ullsperger et al., 2010), reasonable predictions can be made about the expected morphology of the Ne with respect to the way errors are processed and interpreted by the individual (for a recent review see e.g., Hoffmann and Falkenstein, 2012). For example, in anxious subjects like obsessive subjects and worriers, negative affect arising from committing an error might activate the ACC more than in normal subjects which is reflected in a larger Ne (Gehring et al., 2000; Ernst et al., 2006; Segalowitz and Dywan, 2009). On the other hand, individuals with reduced amygdala activations might have reduced PFC activation to inhibit responses of the nucleus accumbens, which in turn might lead to a reduced Ne (Ernst et al., 2006). Typically, these subjects experience a heightened sensation-seeking tendency. Accordingly, risk-taking is correlated with the Ne amplitude (Santesso and Segalowitz, 2009). Also, subjects scoring high on neuroticism show a larger alteration of the Ne if motivation is being manipulated (Pailing and Segalowitz, 2004). In sum, it appears that the Ne is strongly affected by emotional aspects of personality. Boksem and his coworkers (Boksem et al., 2006) related the Ne to the Behavioral Inhibition System (BIS/BAS) scales, which are based on a biopsychological theory of personality (Gray, 1987, 1989). They found that subjects scoring high on the BIS scale, i.e., who are particularly sensitive to punishment, displayed larger Ne amplitudes. Also, it was already shown that the Ne was increased for subjects scoring high in negative affect and negative emotionality scales, at least at an early time point of the conducted experiment (Luu et al., 2000). Finally, Ruchsow et al. (2005) found the Ne to be related to impulsivity: highly impulsive subjects display a reduced Ne. However, Santesso and Segalowitz (2009) found no relation of the Ne to impulsivity and punishment, but rather to empathy and risk raking.

In sum there does not exist very much work on the relation of error processing and personality, and the results are partly contradictory. Thus, the purpose of the study was to test whether the Ne is related to a wider range of personality factors. Therefore, a classic personality inventory, the Freiburg Personality Inventory (FPI-R, Fahrenberg et al., 1989), was utilized in order to test relations of the Ne and personality factors which conceptually integrate the "normal" range of personality dimensions which are, if showing extreme expressions, indicators of pathologies typically mediating Ne effects (anxiety, violation of social norms, social orientation). Also, these personality dimensions should incorporate emotional reactivity (e.g., empathy, impulsivity). Therefore, the study focused on the personality traits "Openness", "Social Orientation", "Emotionality", and "Impulsiveness" as measured by the FPI-R. **Table 1** provides an overview of how these dimensions are conceptualized.

Table 1 | Measured personality dimension of the FPI-R in the present study.

Dimension	Description
Openness	Oriented to social norms, concerned of making a good impression, unable to be self-critical, closed vs. admitting minor weaknesses and common violations of social norms, unembarrassed
Social orientation	Self-concerned, uncooperative, little solidarity vs. taking social responsibility, helpful, considerable
Emotionality	Emotionally stable, self-confident, having life content vs. emotionally labile, hypersensitive
Impulsiveness	Controlled, calm, composed vs. easy aroused, hypersensitive, uncontrolled

METHODS

PARTICIPANTS

A student sample of 28 healthy young subjects participated (11 females). Subjects were aged between 19 and 30 years (mean = 22.7, SD = 2.5) and gave written informed consent prior to participation. Subjects received 10, –€/h payment for participation. The study was conducted according to the code of ethics of the World Medical Association (Declaration of Helsinki) and was proofed by the local ethics committee. The data of three participants had to be rejected due to too few errors ($N = 1$) and bad data channels in the EEG ($N = 2$).

GENERAL PROCEDURE AND EXPERIMENTAL DESIGN

Participants were seated in an ergonomic chair in front of a backprojection beamer display. The distance from display to participants was about 120 cm. The viewing angle was for all target stimuli 1.5°. Stimuli were presented in mild gray on a dark gray background. Subjects were asked to response by a button press of the left or right index finger. The experiment consisted of 720 trials, from which the first 80 trials (training trials) were discarded from further analysis. A short break was provided at 1/3 and 2/3 of the trials. Prior to participation of the EEG experiment all participants filled out the FPI-R and the STAI-T in a separate room. Instructions for the questionnaires were given according to the instruction of the manuals.

For inducing errors a modified global-local task was conducted. During the experiment a combination of two letters (F, H) of different size (small, large) was presented during each trial. The large letters had a viewing angle of 1.5°, and the small ones a viewing angle of 0.5°. The instruction of the participants was to respond to the smaller (local) or larger (global) letter according to a predefined instruction: "attend to the small letter" (local), "attend to the large letter"(global), "ignore the small letter"(global), and "ignore the large letter"(local). This instruction was presented to the subjects as a pre-cue 1000 ms prior to presentation of the target stimulus. Subject had to respond to the to-be-attended letter, or to the not-to-be ignored letter. The attended letter H required a button press of the right index finger, and the letter F a response with the left index finger. This stimulus-response mapping was fully balanced, i.e., half of the participants should respond the other way round.

Following their button press, participants received a feedback according to their performance ("good", "error", "too slow"). An adaptive deadline was utilized (Rinkenauer et al., 2004; Hoffmann

and Falkenstein, 2010, 2011) in order to minimize speed-accuracy tradeoff. This means that the maximum time during which participants had to respond without receiving the feedback "too slow" was adapted online by estimating the percentage of errors during the last 40 trials. If the error rate was lower than 8% the deadline was decreased by 50 ms; if the error rate was above 12% the deadline was increased by 50 ms. This way the deadline was adapted to the error rate, and not to the mean RT, which prevents reliably from a strong speed-accuracy trade-off.

EEG-RECORDINGS AND ANALYSIS

The EEG was recorded from 60 standard electrode sites using an active electrode system (actiCap©, BrainProducts). EOG was recorded vertically from above and below the right eye (vEOG) and from the outer canthi of both eyes (hEOG). EEG and EOG were digitized at 1000 Hz and the recording was conducted using an average reference amplifier system (QuickAmp©, BrainProducts).

The EEG was analyzed using EEGLAB (Delorme and Makeig, 2004) and Matlab©. Initially, data were filtered (0.1–30 Hz) using a phase-shift free IIR Butterworth filter. Subsequently, the data were segmented relative to the button press (–0.5 to 1.0 s) and pruned automatically from artifacts by a statistical thresholding procedure (Delorme et al., 2007). Then the EEG was re-referenced to linked mastoids. For removal of ocular artifacts independent component analysis was conducted by applying extended infomax (Lee et al., 1999) with default settings as implemented in EEGLAB. Ocular artifacts were removed by applying an objective automatic artifact removal algorithm which combines spatial and temporal features of independent components for artifact detection (Mognon et al., 2010).

Following this the automated thresholding protocol was conducted a second time in order to remove residual artifacts. Finally, event-related potentials were calculated for correct and erroneous responses, respectively. The Ne was measured as a difference wave between correct and erroneous response-related activity. From this difference wave the Ne was extracted as a peak-to-peak measure at FCz, i.e., as the difference between the local maximum in the time range from –50 to 0 ms and the local minimum in the time range from 0 to 100 ms. The rationale for this was to yield an estimate of the error-specific variation in the post-response ERP. In the following this difference is termed Ne. This measure was correlated with the personality questionnaire scores.

Additionally, sLORETA estimations of the source of variation between the response-related ERPs of correct and erroneous responses corresponding to the Ne peak were conducted. sLORETA can be used to estimate the sources generating the variance between two experimental conditions. sLORETA is an improved version of LORETA. The main difference is that sources are estimated based on standardized current density allowing a more precise source localization than the previous LORETA-method (Pascual-Marqui, 2002; Pascual-Marqui et al., 2002). sLORETA was performed with the scalp maps of the Ne and corresponding time window of the correct response to find the generators of these maps. This was done by comparison of the voxel-based sLORETA-images (6239 voxels at a spatial resolution of 5 mm; MNI template) of both response types. Statistical quantification was conducted by using the sLORETA-built-in voxelwise randomization tests (5000 permutations) based on statistical non-parametric mapping (SnPM), corrected for multiple comparisons (Nichols and Holmes, 2002). The voxels with significant differences ($p < 0.05$) were located in specific brain regions.

BEHAVIORAL DATA

Reaction time data were analyzed by means of repeated measures ANOVAs with the factors accuracy (correct vs. erroneous responses) and cue (global vs. local). Only RTs in the time range from 100 to 1000 ms were included in the analysis. Error rates were analyzed by a T-test. Effect sizes are provided by means of partial eta squared (η_p^2) and Cohen's d.

QUESTIONAIRE DATA AND Ne

For assessing personality dimensions the Freiburg Personality Inventory was utilized in its revised version (FPI-R, Fahrenberg et al., 1989). The FPI-R is a structured objective measure of personality traits. It consists of 128 items which are grouped on 10 primary scales and two secondary scales. The former consist of the dimensions "Life satisfaction", "Social orientation", "Achiement orientation", "Inhibitedness", "Impulsiveness", "Aggressiveness", "Strain", "Somatic complaints", "Health concern", and "Frankness/Openness". The two secondary scales consist of "Extraversion" and "Emotionality," which shall measure Eysencks's Extraversion/Introversion and Neuroticism constructs. The FPI-R is one of the most used and best assessed personality inventories in the German language area and yields sufficient reliability (Cronbachs' alpha between 0.73 and 0.83). All analyses were restricted to personality factors from which it can be assumed that they are conceptually related to cognitive control, which are "Openness", "Social Orientation", "Emotionality", and Impulsiveness". Due to the sample size, which was not sufficient for multiple regression analysis, the sample was divided into high and low scoring subjects by a median split of the questionnaire data. Therefore, "high" and "low" groups did not consist exactly of the same number of subjects (which was due to a slight skewness in the data and the fact that the final sample consisted of 25 subjects). Subsequently, the relation of Ne and personality factors was assessed by means of Welch T-Tests that account for different N and unequal variances. Reported are always T-Values and corrected degrees of freedom. Additionally, effect sizes are provided by means of Cohen's d in order to ease power estimations.

RESULTS

BEHAVIORAL DATA

Repeated measures ANOVAs revealed a significant effect for the factors cue [$F_{(1,\ 24)} = 24.76$, $p = 4.42 - 05$, $\eta_p^2 = 0.51$] and accuracy [$F_{(1,\ 24)} = 27.48$, $p = 2.25e - 05$, $\eta_p^2 = 0.53$]. However, a significant interaction of cue and accuracy [$F_{(1,\ 24)} = 36.23$, $p = 3.25e - 06$, $\eta_p^2 = 0.6$] revealed that RTs did not differ between the global and local conditions if the responses were erroneous, but they were considerably longer for the local condition of correct responses compared to correct responses in the global conditions (compare **Figure 1**). Error rates were lower

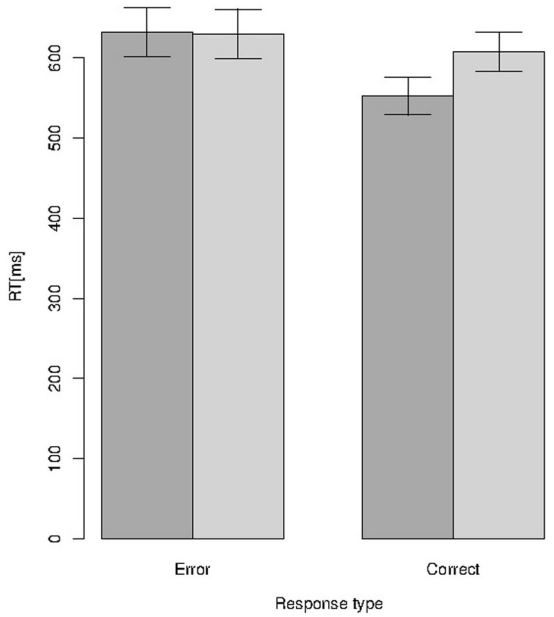


FIGURE 1 | Reaction times (RT) as a function of Cue (global = dark gray, local = gray) and response type (errors vs. correct), bars = standard error of the mean.

for the global compared to the local condition [4.6 vs. 6%; $t_{(24)} = -2.36, p = 0.013, d = 0.47$].

Ne AND QUESTIONNAIRE DATA

There was a clear error negativity observable in the data [**Figure 2**; $t_{(24)} = 8.81$, $p = 2.76e-09$, $d = 1.32$]. sLORETA localized the Ne complex in the ACC [$\text{Tal}_{(x,y,z)} = Tal_{(x,y,z)} = -10, 35, 25$, $t < -4.3$, $p = 0.01$]. In this task, the Ne amplitude was negatively correlated with the error rate [$r = -0.53$, $t_{(23)} = -2.99$, $p = 0.006$].

The Ne was larger for subjects scoring high vs. low on the "Social Orientation" scale [$t_{(21.71)} = 2.29$, $p = 0.01$, $d = 0.91$] and smaller for subjects scoring high vs. low on the scales "Openness" [$t_{(18.6)} = 3.01$, $p = 0.003$, $d = 1.22$], "Impulsiveness" [$t_{(17.889)} = 2.19, p = 0.02, d = 0.88$] and "Emotionality" [$t_{(17.75)} = 1.92$, $p = 0.03$, $d = 0.78$]. **Figure 3** summarizes these results.

However, since the error rate was correlated significantly with the Ne, and the error rate itself might well be correlated with personality traits further analyses were necessary. Therefore, partial correlations between Ne and personality dimensions, controlled for error rates, were calculated. This was done by calculating a linear regression between Ne and error rate (Ne = b*error rate) and correlating the residuals of this linear fit with the personality scores. Again, the Ne was less pronounced for subjects scoring high on the scales "Openness" [$t_{(18.1)} = 1.833, p = 0.04, d = 0.74$], "Impulsiveness" [$t_{(19.028)} = 2.304, p = 0.01, d = 0.93$], and "Emotionality" [$t_{(17.97)} = 1.89, p = 0.037, d = 0.77$]. However, the relation of Ned and "Social Orientation" did only show a trend toward significance [$t_{(21.57)} = 1.34$, $p = 0.096$, $d = 0.53$].

SUMMARY AND DISCUSSION

The core result of the present study is that the Ne was clearly related to personality traits closely associated with social and emotional dimensions as well as behavioral flexibility. As such the Ne was correlated with the factors "Openness", "Impulsiveness", "Emotionality", and "Social Orientation". After cancelling out the variance due to error rate, the first three dimensions still correlated with the Ne. The large effect sizes (Cohen's $d > 0.7$ on average) indicate a strong relation, which cannot be due to spurious correlations.

This is, at least to our knowledge, the first time that such strong relations are shown. Although a moderate relation of the Ne and personality factors has already been shown (e.g., Pailing and Segalowitz, 2004; Santesso and Segalowitz, 2009), this pattern is a bit heterogeneous. This might be due to different tasks, measurement systems and more importantly, different personality constructs. For example the study of Pailing and Segalowitz (2004) utilized the NEO-PI-R that measures five more global personality dimensions (with comparable reliability like the FPI-R), whereas the FPI-R measures 12 dimensions. Also, the number of participants ($N = 18$) was relatively low for regression analyses. Though the present study tested at least 25 subjects this number is still not sufficient for an adequate multivariate testing, i.e., multiple regression analyses. Therefore, we focused on "dimensions of interest" in one specific test that has been tested extensively in the German area. For more general analyses with higher statistical power more participants are necessary (Tabachnick and Fidell, 2001). Also, it has to be kept in mind that personality tests are sensitive to cultural influences. Furthermore, personality questionnaires do not have the same reliability as performance measures. This difference in the measurement error might lead to false insignificant correlations: if one of two samples is less reliable than the other one, the correlation of both might be underestimated. From the Ne it is known that it is quite reliable across measurements and within subjects and thus it can be interpreted as a trait-like correlate of ACC integrity (Segalowitz et al., 2010; Weinberg and Hajcak, 2011). Moreover, the reported Ne reliability was between 0.56 and 0.67, which is compared to modern psychological tests relatively low. For example the reliability of the test utilized in the study reported here lies between 0.73 and 0.81 (according to the test manual). These coefficients are typical for good scales in this research area. Also, it has to be kept in mind that in contrast to psychometric tests the measurement of the Ne is not standardized, nor exist norm values. Furthermore, the acquisition and analysis protocol differs from laboratory to laboratory. Thus, in the case of the Ne it is not clear where the measurement error comes from: is the Ne itself less reliable or is this reliability due to a measurement error? This question cannot be answered from concepts of classical test theory (like reliability). In sum this might be one reason why there exist only few studies reporting relations of the Ne and personality factors. Nevertheless, it can be expected that the Ne, as a correlate of response monitoring *per se*, should be correlated with personality factors which incorporate the evaluation of one's own behavior.

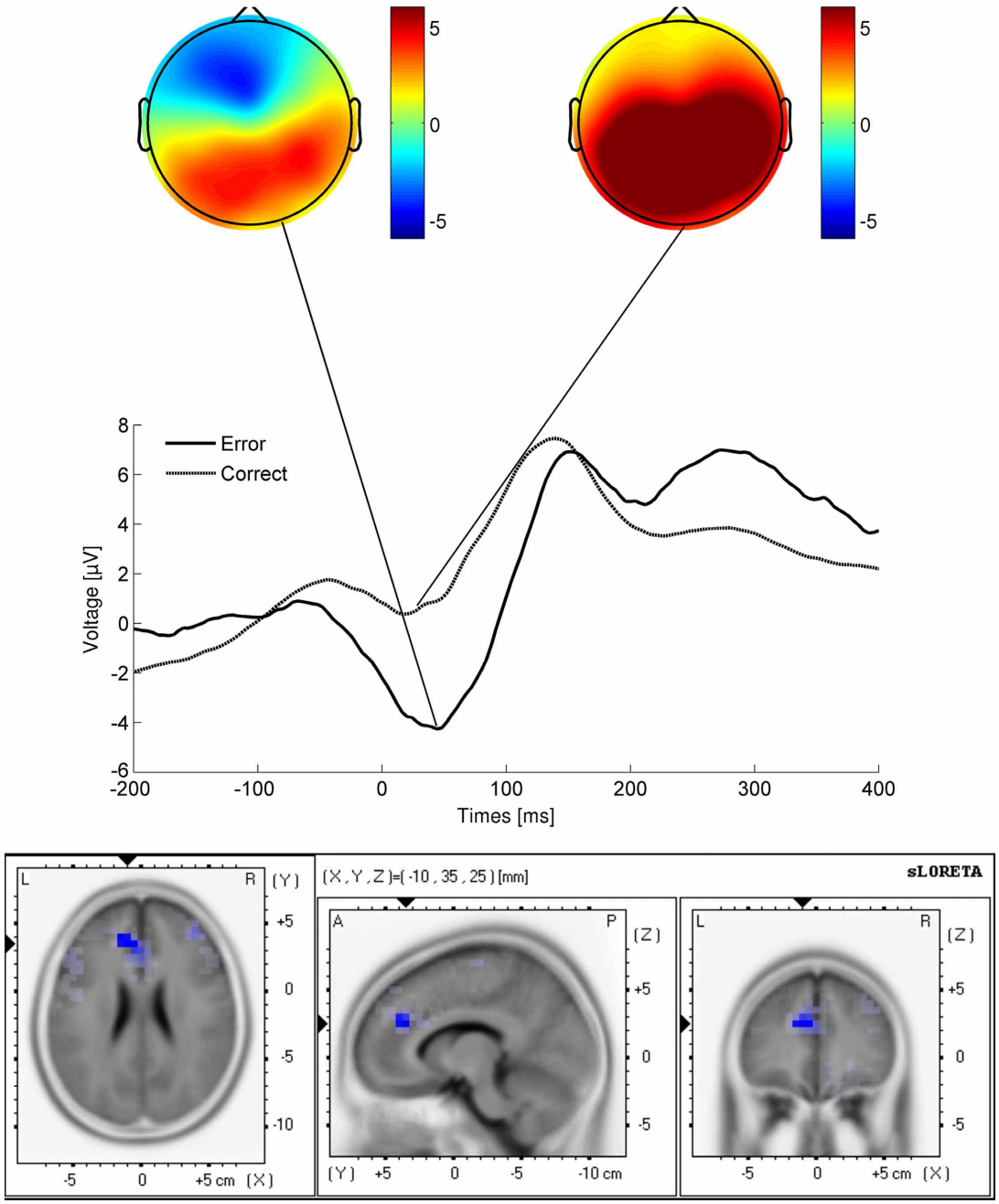


FIGURE 2 | Response-related potentials for correct and erroneous responses (response: time point zero) as well as corresponding topographies at the time point of maximum deflection in the Ne. Lower panel: sLORETA estimate of the Ne ($Tal_{(x,y,z)} = -10, 35, 25$). The sLORETA estimations of all participants' topographies were projected on an averaged normalized brain (MNI template).

What is surprising in the present study is the relation between the Ne and several personality dimensions. However, this becomes clearer if having a look at the inter-correlation of the personality factors: of course they are, at least to some degree, conceptually related to each other. Thus, with respect to the relation of the measured personality dimensions these results are not so unexpected. Subjects scoring low in the "Openness", "Impulsivity", and "Emotionality" scale showed a larger Ne, which is in line with results from clinical samples and previous studies dealing with the relation of personality and error monitoring.

For recapitulation: "Openness/frankness" refers to a personality ranging from oriented toward social norms (following norms, introverted, making a good impression) to admitting violations of social norms or being unconventional. In the present study, subjects that tended to admit social norms or to be unconventional

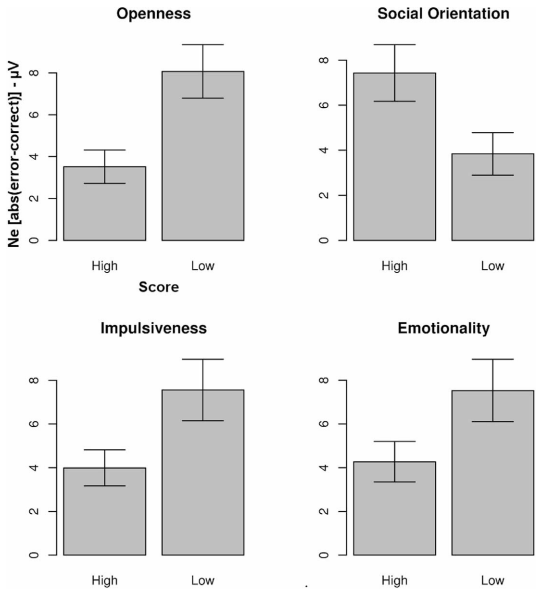


FIGURE 3 | Average absolute values of the error-correct difference for four personality dimensions according to the FPI-R (bars = standard errors of the mean). The Ne was more pronounced for subjection scoring low on the "Openness" scale, the "Impulsiveness" scale and the "Emotionality" scale. Inversely, the Ne was less pronounced for subjects scoring low on the "Social Orientation" scale.

showed decreased error negativities compared to subjects who reported of being oriented toward social norms. According to the authors of the FPI-R, high scoring subjects just don't care very much about conventions (Fahrenberg et al., 1989, p. 77). The analogy to the results from studies dealing with psychopathy lies at hand: in subjects with a psychopathy diagnosis, the Ne is reduced (Munro et al., 2007; von Borries et al., 2010). Typically, psychopaths show a lack of interest in social norms.

With respect to the trait "Impulsiveness," which ranges from calm, self-controlled to aggressive and spontaneous the results indicate that subjects that reported being self-controlled showed enhanced error negativities, which appears also to be predictable: maybe these subjects are able to gain more "self-control," which is reflected in their capability of monitoring errors.

With respect to "Emotionality," which ranges from emotionally stable, controlled to being a bit anxiously and being more pessimistic the results indicate that subjects which reported of being more emotionally stable showed increased Ne amplitudes. This appears to be not in line with studies showing enhanced Ne amplitudes for subjects suffering from anxiety disorders or depression (Olvet and Hajcak, 2008; Weinberg et al., 2010). However, in sum it may be that low scoring subjects (in all three scales) do not engage to a comparable degree like high scoring subjects in the task. Here further research is necessary, for example with studies experimentally varying task engagement. Also, it has to be kept in mind, that the personality traits have to be distinguished from pathologies and thus only cautious conclusion might be made.

The final scale, "social orientation" ranges from being self-concerned, being uncooperative to feeling social responsible and being helpful. Here the data pattern is inversed to the one that could be observed in the other scales: subjects that scored low, i.e., who reported being more helpful than being self-concerned showed enhanced error negativities. If this scale is interpreted as a scale that might capture a concept like empathy, this data pattern reminds of the results of Santesso and Segalowitz (2009). They showed that the Ne was enhanced in subjects scoring high on an empathy scale.

In a nutshell, it could be concluded that since the personality variables measured herein are trait measures, the found effects might reflect the degree to which individuals are concerned with the outcome of their behaviors. Thus, the Ne might also reflect trait-like emotional reactivity.

REFERENCES

Alexander, W. H., and Brown, J. W. (2011). Medial prefrontal cortex as an action-outcome predictor. *Nat. Neurosci.* 14, 1338–1344.

Beste, C., Willemssen, R., Saft, C., and Falkenstein, M. (2009). Error processing in normal aging and in basal ganglia disorders. *Neuroscience* 159, 143–149.

Boksem, M. A., Tops, M., Wester, A. E., Meijman, T. F., and Lorist, M. M. (2006). Error-related ERP components and individual differences in punishment and reward sensitivity. *Brain Res.* 1101, 92–101.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., and Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychol. Rev.* 108, 624–652.

Botvinick, M. M., Cohen, J. D., and Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: an update. *Trends Cogn. Sci.* 8, 539–546.

Braver, T. S., Barch, D. M., Gray, J. R., Molfese, D. L., and Snyder, A. (2001). Anterior cingulate cortex and response conflict: effects of frequency, inhibition and errors. *Cereb. Cortex* 11, 825–836.

Carbonnell, L., and Falkenstein, M. (2006). Does the error negativity reflect the degree of response conflict? *Brain Res.* 1095, 124–130.

Danielmeier, C., Eichele, T., Forstmann, B. U., Tittgemeyer, M., and Ullsperger, M. (2011). Posterior medial frontal cortex activity predicts post-error adaptations in task-related visual and motor areas. *J. Neurosci.* 31, 1780–1789.

de Bruijn, E. R. A., de Lange, F. P., Von Cramon, D. Y., and Ullsperger, M. (2009). When errors are rewarding. *J. Neurosci.* 29, 12183–12186.

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

Dehaene, S., Posner, M., and Tucker, D. M. (1994). Localization of an neural system for error detection and compensation. *Psychol. Sci.* 5, 303–305.

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21.

Delorme, A., Sejnowski, T., and Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *Neuroimage* 34, 1443–1449.

Devinsky, O., Morrell, M. J., and Vogt, B. A. (1995). Contributions of anterior cingulate cortex to behaviour. *Brain* 118(Pt 1), 279–306.

Dikman, Z. V., and Allen, J. J. (2000). Error monitoring during reward and avoidance learning in high- and low-socialized individuals. *Psychophysiology* 37, 43–54.

Ernst, M., Pine, D. S., and Hardin, M. (2006). Triadic model of the neurobiology of motivated behavior in adolescence. *Psychol. Med.* 36, 299–312.

Fahrenberg, J., Hampel, R., and Selg, H. (1989). *Das Freiburger Persönlichkeitsinventar FPI. Revidierte Fassung FPI-R.* Göttingen, Toronto, Zürich: Verlag für Psychologie, Hogrefe.

Falkenstein, M., Hohnsbein, J., Hoomann, J., and Blanke, L. (1990). "Effects of errors in choice reaction tasks on the {ERP} under

focused and divided attention," in *Psychophysiological Brain Research*, eds C. H. M. Brunia, A. W. K. Gaillard, and A. Kok (Tilburg: Tilburg University Press), 192–195.

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., and Donchin, E. (1993). A neural system for error-dedection and compensation. *Psychol. Sci.* 4, 385–390.

Gehring, W. J., Himle, J., and Nisenson, L. G. (2000). Action-monitoring dysfunction in obsessive-compulsive disorder. *Psychol. Sci.* 11, 1–6.

Gentsch, A., Ullsperger, P., and Ullsperger, M. (2009). Dissociable medial frontal negativities from a common monitoring system for self- and externally caused failure of goal achievement. *Neuroimage* 47, 2023–2030.

Gray, J. A. (1987). "The neuropsychology of emotion and personality," in *Cognitive Neurochemistry*, eds S. M. Stahl, S. D. Iverson, and E. C. Goodman (Oxford: Oxford University Press), 171–190.

Gray, J. A. (1989). "Fundamental systems of emotion in the mammalian brain," in *Coping with Uncertainty: Behavioral and Developmental Perspectives,* ed D. S. Palermo (Hillsdale, NJ: Lawrence Erlbaum), 173–195.

Hajcak, G., McDonald, N., and Simons, R. F. (2003). Anxiety and error-related brain activity. *Biol. Psychol.* 64, 77–90.

Hajcak, G., McDonald, N., and Simons, R. F. (2004). Error-related psychophysiology and negative affect. *Brain Cogn.* 56, 189–197.

Hoffmann, S., and Falkenstein, M. (2010). Independent component analysis of erroneous and correct responses suggests online response control. *Hum. Brain Mapp.* 31, 1305–1315.

Hoffmann, S., and Falkenstein, M. (2011). Aging and error processing: age related increase in the variability of the error-negativity is not accompanied by increase in response variability. *PLoS ONE* 6:e17482. doi: 10.1371/journal.pone.0017482

Hoffmann, S., and Falkenstein, M. (2012). Predictive information processing in the brain: errors and response monitoring. *Int. J. Psychophysiol.* 83, 208–212.

Holroyd, C. B., and Coles, M. G. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychol. Rev.* 109, 679–709.

Holroyd, C. B., and Yeung, N. (2003). Alcohol and error processing. *Trends Neurosci.* 26, 402–404.

Jocham, G., and Ullsperger, M. (2009). Neuropharmacology of performance monitoring. *Neurosci. Biobehav. Rev.* 33, 48–60.

Lee, T. W., Girolami, M., and Sejnowski, T. J. (1999). Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural Comput.* 11, 417–441.

Liotti, M., Pliszka, S. R., Perez, R., Kothmann, D., and Woldorff, M. G. (2005). Abnormal brain activity related to performance monitoring and error detection in children with ADHD. *Cortex* 41, 377–388.

Luu, P., Collins, P., and Tucker, D. M. (2000). Mood, personality, and self-monitoring: negative affect and emotionality in relation to frontal lobe mechanisms of error monitoring. *J. Exp. Psychol. Gen.* 129, 43–60.

Masaki, H., Falkenstein, M., Sturmer, B., Pinkpank, T., and Sommer, W. (2007). Does the error negativity reflect response conflict strength? Evidence from a simon task. *Psychophysiology* 44, 579–585.

Mathalon, D. H., Fedor, M., Faustman, W. O., Gray, M., Askari, N., and Ford, J. M. (2002). Response-monitoring dysfunction in schizophrenia: an event-related brain potential study. *J. Abnorm. Psychol.* 111, 22–41.

Mognon, A., Jovicich, J., Bruzzone, L., and Buiatti, M. (2010). ADJUST: an automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology* 48, 229–240.

Munro, G. E., Dywan, J., Harris, G. T., McKee, S., Unsal, A., and Segalowitz, S. J. (2007). ERN varies with degree of psychopathy in an emotion discrimination task. *Biol. Psychol.* 76, 31–42.

Nichols, T. E., and Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum. Brain Mapp.* 15, 1–25.

Nieuwenhuis, S., Ridderinkhof, K. R., Talsma, D., Coles, M. G., Holroyd, C. B., Kok, A., and van der Molen, M. W. (2002). A computational account of altered error processing in older age: dopamine and the error-related negativity. *Cogn. Affect. Behav. Neurosci.* 2, 19–36.

Olvet, D. M., and Hajcak, G. (2008). The error-related negativity (ERN) and psychopathology: toward an endophenotype. *Clin. Psychol. Rev.* 28, 1343–1354.

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

Pascual-Marqui, R. D. (2002). Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. *Methods Find. Exp. Clin. Pharmacol.* 24(Suppl. D), 5–12.

Pascual-Marqui, R. D., Esslen, M., Kochi, K., and Lehmann, D. (2002). Functional imaging with low-resolution brain electromagnetic tomography (LORETA): a review. *Methods Find. Exp. Clin. Pharmacol.* 24(Suppl. C), 91–95.

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

Rinkenauer, G., Osman, A., Ulrich, R., Muller-Gethmann, H., and Mattes, S. (2004). On the locus of speed-accuracy trade-off in reaction time: inferences from the lateralized readiness potential. *J. Exp. Psychol. Gen.* 133, 261–282.

Roger, C., Benar, C. G., Vidal, F., Hasbroucq, T., and Burle, B. (2010). Rostral cingulate zone and correct response monitoring: ICA and source localization evidences for the unicity of correct- and error-negativities. *Neuroimage* 51, 391–403.

Ruchsow, M., Spitzer, M., Gron, G., Grothe, J., and Kiefer, M. (2005). Error processing and impulsiveness in normals: evidence from event-related potentials. *Brain Res. Cogn. Brain Res.* 24, 317–325.

Santesso, D. L., and Segalowitz, S. J. (2009). The error-related negativity is related to risk taking and empathy in young men. *Psychophysiology* 46, 143–152.

Sarter, M., Gehring, W. J., and Kozak, R. (2006). More attention must be paid: the neurobiology of attentional effort. *Brain Res. Rev.* 51, 145–160.

Segalowitz, S. J., and Dywan, J. (2009). Individual differences and developmental change in the ERN response: implications for models of ACC function. *Psychol. Res.* 73, 857–870.

Segalowitz, S. J., Santesso, D. L., Murphy, T. I., Homan, D., Chantziantoniou, D. K., and Khan, S. (2010). Retest reliability of medial frontal negativities during performance monitoring. *Psychophysiology* 47, 260–270.

Tabachnick, B. G., and Fidell, L. S. (2001). *Using Multivariate Statistics.* Boston, MA: Allyn and Bacon.

Ullsperger, M., Harsay, H. A., Wessel, J. R., and Ridderinkhof, K. R. (2010). Conscious perception of errors and its relation to the anterior insula. *Brain Struct. Funct.* 214, 629–643.

van Veen, V., Cohen, J. D., Botvinick, M. M., Stenger, V. A., and Carter, C. S. (2001). Anterior cingulate cortex, conflict monitoring, and levels of processing. *Neuroimage* 14, 1302–1308.

von Borries, A. K., Brazil, I. A., Bulten, B. H., Buitelaar, J. K., Verkes, R. J., and de Bruijn, E. R. (2010). Neural correlates of error-related learning deficits in individuals with psychopathy. *Psychol. Med.* 40, 1559–1568.

Weinberg, A., and Hajcak, G. (2011). Longer term test-retest reliability of error-related brain activity. *Psychophysiology* 48, 1420–1425.

Weinberg, A., Olvet, D. M., and Hajcak, G. (2010). Increased error-related brain activity in generalized anxiety disorder. *Biol. Psychol.* 85, 472–480.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 28 February 2012; accepted: 25 May 2012; published online: 11 June 2012.

Citation: Hoffmann S, Wascher E and Falkenstein M (2012) Personality and error monitoring: an update. Front. Hum. Neurosci. ***6****:171. doi: 10.3389/fnhum.2012.00171*

frontiers in
HUMAN NEUROSCIENCE

GENERAL COMMENTARY
published: 17 October 2012
doi: 10.3389/fnhum.2012.00283

An updated update to personality and error monitoring

Mattie Tops[1,2,3]* and Sander L. Koole[1]

[1] *Department of Clinical Psychology, VU University Amsterdam, Amsterdam, Netherlands*
[2] *Centre for Child and Family Studies, University of Leiden, Leiden, Netherlands*
[3] *Leiden Institute for Brain and Cognition, Leiden University Medical Center, Leiden, Netherlands*
**Correspondence: m.tops@vu.nl*

Edited by:
Christian Bellebaum, Ruhr University Bochum, Germany

Reviewed by:
Christian Bellebaum, Ruhr University Bochum, Germany

A commentary on

Personality and error monitoring: an update
by Hoffmann, S., Wascher, E., and Falkenstein, M. (2012). Front. Hum. Neurosci. 6:171. doi: 10.3389/fnhum.2012.00171

People differ strongly in the degree of error processing, and how errors are interpreted and appraised. In a recent study in Frontiers in Human Neuroscience, Hoffmann et al. (2012) investigated whether a correlate of error monitoring, the error negativity (Ne or ERN), is related to personality factors. They measured the EEG continuously during a task that provoked errors, and the Ne was tested with respect to its relation to personality traits. The amplitude of the Ne was smaller in individuals who scored higher on the "Openness" scale, the "Impulsiveness" scale, and the "Emotionality" scale. By contrast, the Ne was larger in individuals who scored higher on the "Social Orientation" scale. These results are partly consistent with previous studies of associations between Ne and personality, and extent those associations to traits that had not been investigated before in this context. However, Hoffmann and colleagues missed some recent findings that may be important in the interpretation of their results.

Previous studies associated Ne amplitude with various traits that often seemed very different from each other. We recently convincingly demonstrated that one thing those traits have in common is that they predict task engagement, suggesting that task engagement is a common underlying factor that predicts the amplitude of the Ne (Tops and Boksem, 2010). In a two-study paper, we first showed that the traits that have been related to Ne amplitude in previous research are interrelated and have in common that they are correlated with the motivational trait of *Persistence*. This by itself supports the hypothesis that engagement is a common underlying factor that predicts the amplitude of the Ne. An alternative factor, such as *concern over social evaluation*, may relate to Persistence and may perhaps explain the association of traits such as BIS and neuroticism with persistence, but does not seem involved in obvious ways in some of the other traits, such as Drive for reward, Impulsivity, and Absorption. Moreover, the second study provided additional support for the engagement hypothesis by showing that the traits interact with context to predict the Ne, such that trait–context combinations that are likely to be associated with increased engagement predict larger Ne amplitudes. For instance, a trait measure of intrinsic motivation (Absorption) predicted both larger Ne amplitudes during the first part of performance when boredom had not yet set in, and a larger decrease in amplitudes during later performance. By contrast, Constraint, a trait related to the resistance of temptation and distraction, predicted larger Ne amplitudes only during later performance when boredom and fatigue increased temptations to disengage. We also review evidence that externalizing psychopathological syndromes in which reduced Ne amplitudes have been found are characterized by reduced Persistence, while internalizing syndromes such as obsessive compulsive disorder in which increased Ne amplitudes have been found are characterized by increased Persistence (Tops and Boksem, 2010).

The traits investigated by Hoffmann and colleagues are related to previously investigated traits. For instance, "Impulsiveness" is associated with externalizing and the opposite pole of Constraint. "Openness" (vs. following social norms and making good impression) and "Social orientation" (helpful vs. uncooperative) are both very likely correlates of the Agreeableness trait that we showed to relate positively to task engagement, trait Persistence, and Ne amplitude (Tops et al., 2006; Tops and Boksem, 2010). Indeed, in their Discussion, Hoffmann and colleagues argued themselves that their traits were interrelated and may be associated with the level of engagement during task performance. Although this suggests that task engagement may provide a parsimonious account for individual differences in Ne amplitudes at the trait and state levels, it cannot be ruled out that mechanisms behind the Ne are functionally implicated in negative affect or behavioral inhibition (e.g., Tops and Boksem, 2011). More studies are needed to address this unresolved issue.

One part of the results of Hoffmann and colleagues that appears to deviate from previous findings is their finding of an association between "Emotionality" and smaller Ne amplitudes. Many studies related Ne amplitude to individual differences that reflect anxiety, punishment sensitivity, or negative emotion traits and it has been suggested that the Ne reflects concern with the outcome of events, which may increase engagement (Hajcak et al., 2005; Boksem et al., 2006, 2008; Tops et al., 2006; Santesso and Segalowitz, 2009). However, it was recently found that Ne amplitude does not relate to such individual differences if trial-by-trial performance feedback is provided (Olvet and Hajcak, 2009). This could mean that also in the Hoffmann et al. study, the provision of performance feedback may have altered performance monitoring processes

and error-related potentials because it was possible to rely on the performance feedback to monitor performance accuracy. However, significant Ne effects showed that there was error processing before feedback. Possibly, in the local-global task, some baseline level of performance monitoring is performed regardless of the presence of feedback, and is sufficient to detect error responses; the feedback may decrease excessive error-monitoring or error-related orienting responses that are related to anxious and emotional traits, or may increase anticipatory processes related to potentially distressing feedback.

To conclude, future studies of individual trait and state differences in Ne amplitude should measure and/or manipulate task engagement to help interpret results and to investigate potential additional determinants of Ne amplitude besides engagement. Among the factors that can be used for this purpose are measurements of traits such as Persistence, variations in task demands, task length and motivation, and interaction between relevant traits and conditions (Tops and Boksem, 2010).

ACKNOWLEDGMENT

This research was supported by a Veni grant of the Netherlands Organization for Scientific Research (NWO; 451-07-013).

REFERENCES

Boksem, M. A. S., Tops, M., Kostermans, E., and De Cremer, D. (2008). Sensitivity to punishment and reward omission: evidence from error-related ERP components. *Biol. Psychol.* 79, 185–192.

Boksem, M. A. S., Tops, M., Wester, A. E., Meijman, T. F., and Lorist, M. M. (2006). Error related ERP components and individual differences in punishment and reward sensitivity. *Brain Res.* 1101, 92–101.

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

Hoffmann, S., Wascher, E., and Falkenstein, M. (2012). Personality and error monitoring: an update. *Front. Hum. Neurosci.* 6:171. doi: 10.3389/fnhum.2012.00171

Olvet, D. M., and Hajcak, G. (2009). The effect of trial-to-trial feedback on the error-related negativity and its relationship with anxiety. *Cogn. Affect. Behav. Neurosci.* 9, 427–433.

Santesso, D. L., and Segalowitz, S. J. (2009). The error-related negativity is related to risk taking and empathy in young men. *Psychophysiology* 46, 143–152.

Tops, M., and Boksem, M. A. S. (2010). Absorbed in the task: Personality measures predict engagement during task performance as tracked by error negativity and asymmetrical frontal activity. *Cogn. Affect. Behav. Neurosci.* 10, 441–453.

Tops, M., and Boksem, M. A. S. (2011). Cortisol involvement in mechanisms of behavioral inhibition. *Psychophysiology* 48, 723–732.

Tops, M., Boksem, M. A. S., Wester, A. E., Lorist, M. M., and Meijman, T. F. (2006). Task engagement and the relationships between the error-related negativity, agreeableness, behavioral shame proneness and cortisol. *Psychoneuroendocrinology* 31, 847–858.

Received: 23 August 2012; accepted: 27 September 2012; published online: 17 October 2012

*Citation: Tops M and Koole SL (2012) An updated update to personality and error monitoring. Front. Hum. Neurosci. **6**:283. doi: 10.3389/fnhum.2012.00283*

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 27 June 2012
doi: 10.3389/fnhum.2012.00186

Punishment sensitivity modulates the processing of negative feedback but not error-induced learning

Kerstin Unger*, Sonja Heintz and Jutta Kray

Department of Psychology, Development of Language, Learning, and Action, Saarland University, Saarbruecken, Germany

Edited by:
Patrizia Thoma, Ruhr-University Bochum, Germany

Reviewed by:
Josep Marco-Pallares, University of Barcelona, Spain
Michael Falkenstein, Leibniz Research Centre für Working Environment and Human Factors, Germany

****Correspondence:***
Kerstin Unger, Department of Psychology, Development of Language, Learning, and Action, Saarland University, P. O. Box 15 11 50, D-66041 Saarbruecken, Germany.
e-mail: k.unger@mx.uni-saarland.de

Accumulating evidence suggests that individual differences in punishment and reward sensitivity are associated with functional alterations in neural systems underlying error and feedback processing. In particular, individuals highly sensitive to punishment have been found to be characterized by larger mediofrontal error signals as reflected in the error negativity/error-related negativity (Ne/ERN) and the feedback-related negativity (FRN). By contrast, reward sensitivity has been shown to relate to the error positivity (Pe). Given that Ne/ERN, FRN, and Pe have been functionally linked to flexible behavioral adaptation, the aim of the present research was to examine how these electrophysiological reflections of error and feedback processing vary as a function of punishment and reward sensitivity during reinforcement learning. We applied a probabilistic learning task that involved three different conditions of feedback validity (100%, 80%, and 50%). In contrast to prior studies using response competition tasks, we did not find reliable correlations between punishment sensitivity and the Ne/ERN. Instead, higher punishment sensitivity predicted larger FRN amplitudes, irrespective of feedback validity. Moreover, higher reward sensitivity was associated with a larger Pe. However, only reward sensitivity was related to better overall learning performance and higher post-error accuracy, whereas highly punishment sensitive participants showed impaired learning performance, suggesting that larger negative feedback-related error signals were not beneficial for learning or even reflected maladaptive information processing in these individuals. Thus, although our findings indicate that individual differences in reward and punishment sensitivity are related to electrophysiological correlates of error and feedback processing, we found less evidence for influences of these personality characteristics on the relation between performance monitoring and feedback-based learning.

Keywords: reinforcement learning, BIS, BAS, punishment sensitivity, reward sensitivity, error-related negativity (ERN), feedback-related negativity (FRN), error positivity (Pe)

INTRODUCTION

Learning from reward and punishment is a prerequisite for flexible behavioral adaptation to changing environmental conditions. There is, however, considerable evidence to suggest that individuals vary in their responsiveness to rewarding and punishing stimuli (Depue and Collins, 1999; Pickering and Gray, 2001; Corr, 2004). According to a prominent neurophysiologically oriented theory of personality, three systems underlie interindividual differences in reward and punishment processing (Gray, 1982; Gray and McNaughton, 2000; McNaughton and Corr, 2004). The behavioral activation system (BAS) is thought to be activated by appetitive stimuli and to promote reward-directed approach behavior. In contrast, the fight-flight-freeze system (FFFS) is presumed to be activated by aversive cues and to mediate defensive avoidance. Activation of the behavioral inhibition system (BIS) has been linked to the detection of conflict between competing goals (e.g., approach-avoidance conflict), resulting in increased arousal, focused attention, and enhanced information processing. The BIS is assumed to inhibit prepotent response tendencies and to arbitrate between conflicting BAS- and FFFS-controlled behaviors by promoting risk-assessment along with a negative processing bias. While reward sensitivity has primarily been related to BAS-functioning, punishment sensitivity has been-related to combined FFFS/BIS-functioning (Corr, 2004).

Recent findings indicate that BAS-reactivity is associated with dopamine-dependent activity cortex (e.g., Beaver et al., 2006; Hahn et al., 2009; Simon et al., 2010). BIS/FFFS-reactivity has been linked to functional variations in a distributed network of neural structures including septo-hippocampal system and amygdala, possibly mediated by serotonergic and noradrenergic mechanisms (Gray and McNaughton, 2000; Smillie, 2008). Moreover, a number of event-related potential (ERP) studies point to a link between self-reported punishment sensitivity and functioning of the medial prefrontal cortex (mPFC), specifically the anterior cingulate cortex (ACC) (e.g., Boksem et al., 2006; Amodio et al., 2008; Balconi and Crivelli, 2010). The ACC has been shown to be involved in the processing of motivationally salient events such as errors, conflict, and punishment cues, and more generally, in integrating action selection with motivational and affective processes (Devinsky et al., 1995; Shackman et al., 2011).

The error negativity (Ne; Falkenstein et al., 1990), or error-related negativity (ERN; Gehring et al., 1993) and the feedback-related negativity (FRN; Miltner et al., 1997) are ERP correlates of error or conflict monitoring and feedback processing that are thought to reflect the evaluative functions subserved by the mPFC/ACC (Ridderinkhof et al., 2004; Taylor et al., 2007). The Ne/ERN is a fronto-centrally distributed negative deflection that peaks within 100 ms after an individual's erroneous response. A morphologically similar component, the FRN, is elicited ∼250–300 ms following the presentation of performance feedback. The FRN is more pronounced after negative compared to positive feedback, indicating that it is sensitive to the valence of an outcome (e.g., Gehring and Willoughby, 2002; Yeung and Sanfey, 2004). Subjects scoring high on measures of negative affectivity and punishment sensitivity appear to be characterized by a larger Ne/ERN (Hajcak et al., 2003, 2004; Boksem et al., 2006, 2008; Amodio et al., 2008; Dennis and Chen, 2009) and FRN (Sato et al., 2005; Balconi and Crivelli, 2010; De Pascalis et al., 2010; Santesso et al., 2011a,b), presumably reflecting enhanced reactivity of the medial prefrontal action monitoring system to outcomes signaling potential threat. In line with this notion, Boksem and colleagues (2008) found that high punishment sensitivity was associated with larger Ne/ERN amplitudes when participants tried to prevent monetary loss but not when they aimed to maximize monetary gain.

Interestingly, Boksem and colleagues (2006, 2008) also reported a positive correlation between reward sensitivity and the error positivity (Pe; Falkenstein et al., 1990), a slow positive-going deflection with a maximum amplitude between 200 and 400 ms after an erroneous response. The Pe shows a centro-parietal scalp distribution and has been mapped to distinct neural generators in the (rostral) ACC and the parietal cortex (Van Veen and Carter, 2002; O'Connell et al., 2007). There is some evidence that the Pe reflects salience or motivational significance of an error and thus may be functionally related to the P300 (Overbeek et al., 2005; Ridderinkhof et al., 2009). In addition, the Pe has been linked to the conscious recognition of an error (Falkenstein et al., 1990; Leuthold and Sommer, 1999; Nieuwenhuis et al., 2001; Endrass et al., 2007). According to Boksem and colleagues (2006, 2008), higher Pe amplitudes in subjects highly sensitive to reward might indicate proactive engagement in the service of maximizing future rewards.

Although the error-related ERP components have been proposed to reflect processes that support flexible behavioral adaptation (Falkenstein et al., 1990; Gehring et al., 1993; Holroyd and Coles, 2002; Yeung et al., 2004; Frank, 2005; Frank et al., 2007a), it remains largely unclear whether variations in Ne/ERN, FRN, and Pe amplitude as a function of punishment and reward sensitivity are accompanied by behavioral alterations. On the one hand, a central implication following from the conceptualization of BIS/FFFS and BAS is that highly punishment sensitive individuals should learn more efficiently from negative action outcomes than less punishment sensitive individuals, whereas high reward sensitivity should be associated with better learning under positive reinforcement (Pickering and Gray, 2001; Corr, 2004). On the other hand, previous studies using reinforcement learning paradigms indicate that Ne/ERN and FRN are neural manifestations of negative reward prediction errors, possibly coded by phasic activity of the midbrain dopamine system (Holroyd and Coles, 2002; Frank et al., 2005). These error signals are assumed to be used by the mPFC to guide adaptive action selection. In support of this view, it has been demonstrated that larger Ne/ERN and FRN amplitudes are associated with a stronger tendency to subsequently avoid the same maladaptive response (Frank et al., 2005; van der Helden et al., 2010; Unger et al., 2012).

So far, most studies reporting a relationship between punishment/reward sensitivity and ERP correlates of error and feedback processing have used response conflict and gambling tasks (Boksem et al., 2006, 2008; Amodio et al., 2008; Santesso et al., 2011b). To our knowledge, only one study has investigated the influence of individual differences in punishment and reward sensitivity on feedback processing in a Go-NoGo learning task (De Pascalis et al., 2010). Although this study failed to obtain a significant correlation between punishment sensitivity and the FRN, individuals with higher trait sensitivity to punishment showed larger FRN amplitudes on NoGo trials than less punishment sensitive individuals when the groups were defined by median split. The main goal of the present research was to further investigate the influence of individual differences in punishment and reward sensitivity on error and feedback processing as reflected in the Ne/ERN, FRN, and Pe. Specifically, we aimed to determine whether the effects of punishment sensitivity on the Ne/ERN and FRN are associated with changes in error-induced behavioral adjustments during reinforcement learning.

To address these issues, we applied a reinforcement learning task that has been used by a number of previous studies to examine learning-related changes in the Ne/ERN and FRN (e.g., Holroyd and Coles, 2002; Eppinger et al., 2008). Since the neural mechanisms of error processing have been shown to be sensitive to the uncertainty of stimulus-response (S-R) mappings inherent in a probabilistic learning task (e.g., Eppinger et al., 2008; Gründler et al., 2009), we manipulated the validity of feedback information by including a deterministic learning condition (100% valid), a probabilistic learning condition (80% valid), and a chance condition (50%). In addition, we administered the Carver and White (1994) BIS/BAS Scales to measure punishment and reward sensitivity. It should be noted, that Ne/ERN, FRN, and Pe have not consistently been found to vary as a function of punishment and reward sensitivity (e.g., Cavanagh and Allen, 2008; Van den Berg et al., 2011). These inconsistencies might partly result from the fact that some of the relevant studies used relatively small samples (<30), limiting the generalizability of the corresponding findings. The present study therefore included a comparatively large sample of 105 participants.

At the behavioral level, we expected higher punishment sensitivity to be associated more efficient error-related behavioral adjustments, i.e., higher post-error accuracy. We also expected to find a positive, albeit weaker, relationship between overall accuracy and both punishment and reward sensitivity, as punishment *and* reward learning can contribute to better overall performance on the task employed in this study. However, given that BIS and BAS are thought to interact in that the activation of one system inhibits the other, effects of punishment and reward sensitivity on overall learning performance might be hard to detect in tasks

involving both reward and punishment cues (Pickering et al., 1997; Corr, 2002).

Regarding the relationship between reward and punishment sensitivity and the ERP correlates of error and feedback processing we expected to replicate previous findings that (i) punishment sensitivity correlates positively with the magnitude of Ne/ERN and FRN and (ii) reward sensitivity correlates positively with the Pe.

Moreover, we examined whether the relations between punishment/reward sensitivity and the error-related ERP components vary over the course of learning. Specifically, we expected the effects of punishment and reward sensitivity on the Ne/ERN and Pe to be larger toward the end compared to the beginning of learning as well as in the deterministic compared to the probabilistic condition, reflecting the participants' ability to represent the correctness of their responses (Holroyd and Coles, 2002). Importantly, previous findings indicate that larger Ne/ERN and FRN amplitudes are associated with more efficient error-related behavioral adjustments (van der Helden et al., 2010; Unger et al., 2012). Moreover, it has recently been shown that hyperresponsivity to punishment cues might be reflected in a strengthening of the *coupling* between error-related neural responses and behavioral adaptation rather than in performance differences *per se* (Cavanagh et al., 2011a,b). On the basis of these findings, we hypothesized that the relationship between the error-related ERP components and accuracy measures might vary as a function of punishment (and reward) sensitivity.

MATERIALS AND METHODS

PARTICIPANTS

One-hundred and five participants (71 female, 34 male) were recruited from the student population of Saarland University. All had normal or corrected-to-normal vision, no known neurological or psychiatric diseases and were free from psychoactive medication or drug use. Mean age was 22 years (range = 18–33 years). Participants gave informed written consent in accordance with the protocols approved by the local ethics committee and received course credit or payment at a rate of € 8/h as well as a small monetary bonus. A further 20 students participated in the study but had to be excluded from analyses because of (i) fewer than 15 electroencephalogram (EEG) epochs in one or more conditions (2), (ii) poor performance on the learning task (less than 55% correct in the deterministic learning condition) (9), (iii) excessive noise in the EEG data (3), and (iv) technical problems during EEG recordings (6).

LEARNING TASK

The stimuli used were colored images of objects (Snodgrass and Vanderwart, 1980). The learning task required participants to make a forced-choice decision (left vs. right button press) upon presentation of the target stimulus. Following the response, either the word "RICHTIG" ("correct"), or "FALSCH" ("incorrect") was shown. If a response exceeded the adaptive deadline (see below), the feedback "ZU LANGSAM" ("too slow") was presented. Participants had to learn the response mappings by trial and error. The task involved three conditions of feedback validity (100%, 80%, and 50%). In the *deterministic* learning condition, feedback was always valid (100%). In the *probabilistic* learning condition, feedback was valid on 80% of the trials only. That is, if a stimulus was assigned to the right response key, participants received "Correct" feedback in 80% and "Incorrect" feedback in 20% of right button presses and vice versa for left button presses. In the *chance* condition, "Correct" and "Incorrect" feedback was delivered at random (see **Figure 1** for an illustration of the experimental paradigm). Four stimuli were associated with each condition, resulting in a total of 12 stimuli presented throughout the task. Within each learning condition, two stimuli were mapped to the left and right response key, respectively. Each stimulus was presented 50 times in pseudo-randomized order, resulting in a total of 600 trials. Participants first completed 60 practice trials. During the experiment, they took self-paced breaks every 30 trials.

Trials began with a variable fixation period of 250–500 ms, which was followed by the presentation of the imperative stimulus for 500 ms. Stimuli were presented on a light gray background. The screen was then blanked for a variable interval of 600–1500 ms, depending on the variable response time window. In order to obtain a sufficient number of error trials, we applied an adaptive response deadline. Based on the proportion of timeout trials, the response window was individually adjusted in steps

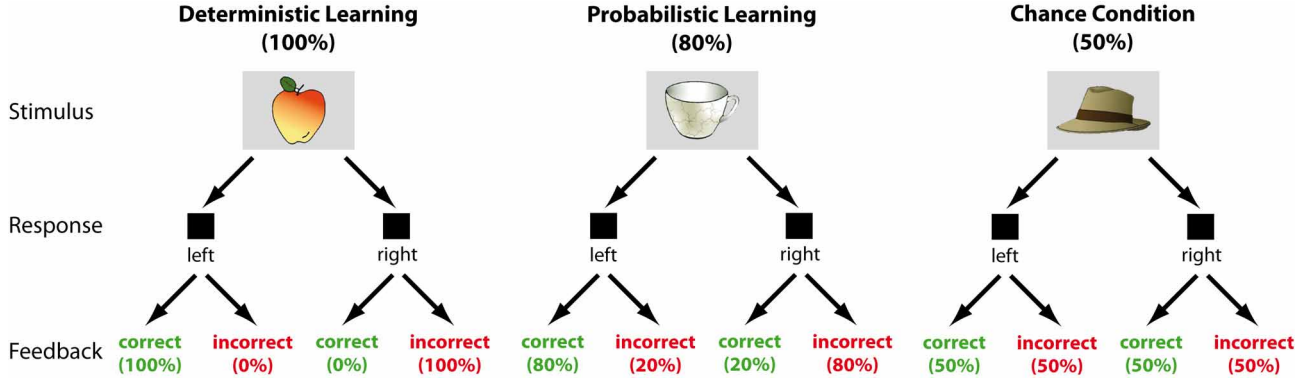


FIGURE 1 | Schematic illustration of the learning paradigm.

of 100 ms within an overall range of 400–1000 ms (for a similar procedure see (Eppinger et al., 2008)). The time window between the response and the feedback was fixed to 500 ms and feedback was displayed for 500 ms. The next trial started after a randomly jittered 1250–2000 ms interval. Participants were informed that they would gain a point for each correct response and lose a point for each incorrect or too slow response and that they could earn a monetary bonus up to 10 Euro depending on the total sum of points obtained. During the breaks, a feedback screen indicated the sum of points they had collected.

PSYCHOMETRIC TESTS AND QUESTIONNAIRES

A German version of the Carver and White (1994) behavioral inhibition scale/behavioral activation scale (BIS/BAS) Scale was used to measure trait-level punishment (seven items; Cronbachs $\alpha = 0.73$; e.g., "I worry about making mistakes.") and reward sensitivity (13 items; Cronbachs $\alpha = 0.59$; e.g., "When good things happen to me, it affects me strongly."). Note that the "BIS" scale contains both FFFS and BIS items. The BIS/BAS scores were in normal ranges for healthy young adults (Strobel et al., 2001). **Table 1** shows means and standard deviations, for the total sample and separately for male and female participants. Consistent with previous reports (Leone et al., 2001), females were characterized by higher punishment sensitivity. We additionally administered the questionnaires assessing positive and negative affectivity (PANAS, Watson et al., 1988) as well as state vs. action orientation (ACS-90, Kuhl, 1994). However, none of these variables was related to learning performance or ERP measures.

PROCEDURE

After a brief description of the experiment, participants filled out a consent form and a short demographic questionnaire. Prior to the electrophysiological recordings, they completed the two psychometric tests and the BIS/BAS scales. The learning task was run in an electrically shielded, dimly lit, sound-attenuated chamber. Stimuli were presented on a CTX 17" monitor and participants responded by pressing the keys C or M on a standard computer keyboard with the left and right index finger, respectively. The entire experiment took approximately 1.5 h.

ELECTROPHYSIOLOGICAL RECORDING

The continuous EEG was recorded from 58 Ag/AgCl electrodes arranged according to the extended 10–20 system and referenced to the left mastoid, using Brain Amp DC Recorder (BrainVision recorder acquisition software). EEG signals were sampled in DC mode with a low-pass filter at 70 Hz and digitized at a sampling rate of 500 Hz. Impedances were kept below 5 kΩ. Electrooculographic activity (EOG) was recorded from electrodes placed on the outer canthi of the two eyes (horizontal EOG) and on the infra- and supra-orbital ridges of the left eye (vertical EOG). The data were re-referenced offline to the linked mastoids and band-pass filtered from 0.1 to 30 Hz. The impact of blinks and eye movements was corrected using an independent component analysis algorithm embedded in the BrainVision Analyzer Software Package (Brain products, Gilching, Germany). Trials containing EEG activity exceeding ±100 μV, changing more than 50 μV between samples or containing DC drifts were removed by means of a semiautomatic artifact inspection procedure.

Table 1 | BIS/BAS scores (Means, Standard deviations, and *t*-test results).

	Total ($N = 105$)	Female ($n = 71$)	Male ($n = 34$)	t	p
BIS	2.87 (0.46)	2.98 (0.44)	2.60 (0.39)	3.32	0.001
BAS	3.17 (0.30)	3.17 (0.29)	3.16 (0.32)	0.16	0.87

DATA ANALYSIS

Behavioral data analyses

Responses faster than 240 ms (<2 SD) or exceeding the adaptive response deadline were excluded from further analyses (on average, the mean number of responses exceeding the individually adjusted deadline was 4%). To examine the course of learning, the behavioral data were averaged into six bins of 100 trials each, i.e., Bin 1 contained Trials 1–100, Bin 2 contained Trials 101–200, and so on. Within each bin, mean reaction times (RTs) and mean accuracy rates were computed separately for the three learning conditions. Only valid trials were included for the probabilistic learning condition. To analyze trial-to-trial behavioral adjustments, we additionally determined post-error accuracy by calculating mean accuracy rates for the next presentation of a given stimulus after an erroneous response, separately for each learning condition.

ERP analyses

Artifact-free EEG data were segmented relative to response and feedback onset to extract response-related and feedback-related ERPs. The response-locked and feedback-locked epochs were baseline corrected with respect to the average voltage during a −200 to −50 ms pre-response interval and a 100 ms pre-stimulus interval, respectively.

As in previous studies employing reinforcement learning tasks (e.g., Frank et al., 2005; Eppinger et al., 2008), we defined the Ne/ERN after 15 Hz low-pass filtering at electrode site FCz as peak-to-peak difference in voltage between the most negative peak in a time window between −50 and 100 ms around the response and the most positive peak within the preceding 100 ms time window. Peak-to-peak voltage was measured to determine baseline-independent amplitudes and to minimize distortions due to the positivity on which the Ne/ERN is superimposed. However, to keep the present results compatible with those from previous studies, we additionally created difference waveforms by subtracting the activity correct trials from the activity on error trials (ΔNe). The ΔNe was defined as the mean amplitude in a 0–100 ms post-response time window covering the peak of the difference wave in each learning condition. Similarly, the FRN was quantified as peak-to-peak voltage difference between the most negative peak in a 200–400 ms time window after feedback onset and the preceding positive peak in a 150–300 ms post-feedback interval at electrode FCz (see Yeung and Sanfey, 2004; Frank et al., 2005). In a further step, ΔFRN amplitude was determined by subtracting the activity after correct feedback from the activity after negative feedback. The ΔFRN was defined as mean amplitude in a 50 ms time window centered on the individual

peaks of the difference waves in a 200 and 400 ms post-feedback interval. Note that larger *negative* values correspond to larger Ne/ERN and FRN amplitudes. We selected FCz for analyses based on visual inspection of the waveforms and the corresponding scalp topographies, which showed a fronto-central maximum of Ne and FRN (see **Figure 3**). Following previous studies (Hajcak et al., 2004; Wiswede et al., 2009), the Pe was measured as the mean amplitude between 200 and 400 ms after the response at electrode Pz.

To examine learning-related changes in Ne/ERN, FRN, and Pe (Holroyd and Coles, 2002), EEG epochs were averaged separately for each incentive condition across the first (Bin 1) and the second half of trials within each block (Bin 2). This analysis, however, included only a subsample of 68 participants (46 female, 22 male), which committed at least 15 errors in each learning condition in both halves of the learning task.

Statistical analyses

Pearson's correlations were calculated to examine the relationships between personality measures, accuracy, and ERP components. Differences between correlation coefficients were tested using the Hotelling–William test (Steiger, 1980). Learning-related effects in the behavioral and ERP data were analyzed using repeated measures analyses of covariance (ANCOVAs) with BIS/BAS as continuous between-subjects factors. The covariates were mean-centered before entering the analysis (Delaney and Maxwell, 1981). To test for differential effects of personality measures in males and females, we performed the same set of analyses with gender as a further between-subjects factor. These analyses did not yield evidence for moderator effects of gender and thus are not reported in the following. In order to test for moderator effects of punishment and reward sensitivity on the relationship between the ERP components and behavioral adjustments, we used multiple regression analyses. The regression models included BIS/BAS, ERP amplitude (Ne/ERN vs. FRN vs. Pe), and the corresponding cross-product terms as predictors and overall accuracy vs. post-error accuracy as criterion. Separate models were tested for the deterministic and probabilistic learning condition. Similar to the ANCOVAs, the independent variables were mean-centered.

Whenever necessary, the Geisser-Greenhouse correction was applied (Geisser and Greenhouse, 1958) and corrected p-values are reported together with the uncorrected degrees of freedom and the epsilon-values (ε). Planned comparisons were performed to decompose significant high-level interactions.

RESULTS

The results section is structures into three parts. In the first part, we will report the analyses of the behavioral data (RT, accuracy, post-error accuracy). The second part involves the correlation analyses of behavioral, personality, and ERP measures. In the third part, we will present the findings on the effects of personality on learning-related modulations of the ERP components.

REACTION TIMES

Mean RTs for correct and incorrect trials were 443 ms (SD = 33 ms) and 434 ms (SD = 52 ms) in the deterministic learning condition, 451 ms (SD = 36 ms) and 450 ms (SD = 45 ms) in the probabilistic learning condition, and 454 ms (SD = 41 ms) and 454 ms (SD = 42 ms) in the chance condition. Response latencies were subjected to an ANCOVA with the within-subjects factors *learning condition* (deterministic, probabilistic, and chance condition, *bin* (Bins 1–6), and *correctness* (correct vs. incorrect) and the continuous between-subjects factors *BIS* and *BAS*. The analysis yielded significant main effects of learning condition [$F_{(2, 204)} = 74.80$, $p < 0.001$, $\eta^2 = 0.43$] and bin [$F_{(5, 510)} = 211.34$, $p < 0.001$, $\eta^2 = 0.68$, $\varepsilon = 0.38$]. Contrasts revealed shorter RTs in the deterministic and probabilistic learning condition compared to the chance condition as well as in the deterministic compared to the probabilistic learning condition (ps < 0.001). Response latencies decreased from Bin 1 to Bin 6, reflected in a significant linear trend across the bins ($p < 0.001$). There were also a main effect of correctness [$F_{(1, 102)} = 72.95$, $p < 0.001$, $\eta^2 = 0.42$], an interaction between bin and correctness [$F_{(5, 510)} = 15.99$, $p < 0.001$, $\eta^2 = 0.14$, $\varepsilon = 0.76$], and an interaction between learning condition, bin, and correctness [$F_{(10, 1020)} = 11.52$, $p < 0.001$, $\eta^2 = 0.10$, $\varepsilon = 0.71$]. Decomposing the Three-Way interaction yielded significant bin × correctness interactions for the deterministic and probabilistic learning condition (ps < 0.001, η^2s > 0.06), indicating that RTs decreased more strongly for erroneous compared to correct responses with increasing time on task in these two conditions. Thus, toward the end of the learning task, RTs were faster on erroneous compared to correct trials, particularly in the deterministic learning condition. No main effect or interaction involving BIS/BAS approached significance.

ACCURACY

Overall accuracy

An ANCOVA with the within-subjects factors *learning condition* and *bin* and the continuous between-subjects factors *BIS* and *BAS* yielded a reliable main effect of learning condition [$F_{(2, 204)} = 533.17$, $p < 0.001$, $\eta^2 = 0.84$], indicating that accuracy was higher in the deterministic and probabilistic learning condition compared to the chance condition as well as in the deterministic compared to the probabilistic learning condition (ps < 0.001, η^2s > 0.53) (see **Figure 2**). Moreover, we found a reliable main effect of bin [$F_{(5, 510)} = 56.92$, $p < 0.001$, $\eta^2 = 0.36$, $\varepsilon = 0.79$] that was qualified by an interaction between learning condition and bin [$F_{(5, 510)} = 14.28$, $p < 0.001$, $\eta^2 = 0.12$, $\varepsilon = 0.81$]. Contrasts revealed a significant interaction when comparing the linear increase of accuracy across bins for deterministic and probabilistic learning condition to the linear increase in the chance condition [$F_{(1, 102)} = 47.72$, $p < 0.001$, $\eta^2 = 0.32$], but not for the deterministic compared to the probabilistic learning condition ($p = 0.64$). As can be seen from **Figure 2**, these findings indicate that accuracy increased across bins in the deterministic and probabilistic learning condition but not in the chance condition. Furthermore, we obtained significant quadratic and cubic interactions between learning condition and bin ($p < 0.001$ and 0.01, $\eta^2 = 0.34$ and 0.09, respectively), reflecting that accuracy increased only from Bin 1 to Bin 3 and reached asymptote thereafter.

As was indicated by a significant main effect of BIS [$F_{(1, 102)} = 5.24$, $p < 0.05$, $\eta^2 = 0.05$], higher punishment sensitivity

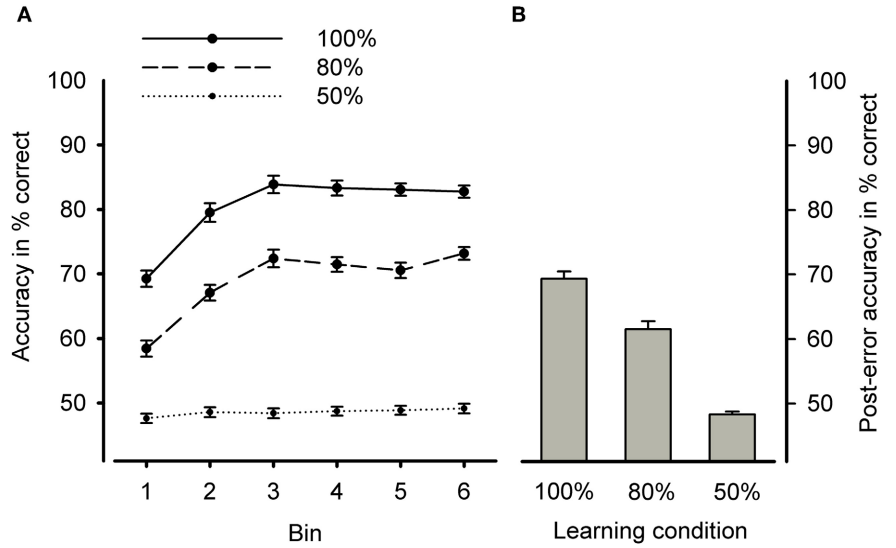


FIGURE 2 | (A) Learning curves (mean accuracy) for the three learning conditions. **(B)** Mean post-error accuracy rates (collapsed across bins) for the three learning conditions. Error bars indicate standard error.

predicted lower overall accuracy (partial $r = -0.20$, $p < 0.05$). By contrast, a reliable main effect of BAS [$F_{(1, 102)} = 4.88$, $p < 0.05$, $\eta^2 = 0.05$] and an interaction between BAS and learning condition [$F_{(2, 210)} = 3.52, p < 0.05, \eta^2 = 0.03$], showed that higher reward sensitivity was associated with higher overall accuracy in the deterministic learning condition (partial $r = 0.25$, $p < 0.05$) but not in probabilistic learning or chance condition (ps > 0.10).

Post-error accuracy

Mean post-error accuracy rates (see **Figure 2**) were subjected to an ANCOVA with the within-subject factor *learning condition* and the continuous between-subjects factors *BIS* and *BAS*. The analysis revealed significant main effects of learning condition [$F_{(2, 210)} = 132.09$, $p < 0.001$, $\eta^2 = 0.57$]. Contrasts revealed post-error accuracy to be higher in the deterministic compared to the probabilistic learning condition as well as for the two learning conditions compared to the chance condition (ps < 0.001, η^2s > 0.22).

Moreover, we found a main effect of BAS [$F_{(1, 102)} = 5.44, p < 0.05$, $\eta^2 = 0.05$] and an interaction between BAS and learning condition [$F_{(2, 210)} = 3.79$, $p < 0.05$, $\eta^2 = 0.04$]. Similar to the findings for overall accuracy, higher reward sensitivity was associated with higher post-error accuracy in the deterministic learning condition only (partial $r = 0.26$, $p < 0.01$). In contrast to overall accuracy, post-error accuracy did not relate to punishment sensitivity.

CORRELATIONS BETWEEN PERSONALITY, BEHAVIOR, AND ERP COMPONENTS

Figure 3 displays the response- and feedback-locked ERPs on correct and incorrect trials, separately for the three learning conditions. The Ne/ERN and the FRN were evident as negative going deflections over fronto-central scalp regions, whereas the Pe was evident as a centro-parietally distributed positive slow wave. Bivariate correlations between Ne/ERN, FRN, Pe, personality measures, and behavior are shown in **Table 2** and **Table 3**, separately for the deterministic and probabilistic learning condition, respectively.

NE/ERN

Contrary to our predictions, we did not observe a significant relationship between BIS score and Ne/ERN measures in either learning condition ($|r$s$| < 0.08$, ps > 0.42). Instead, higher BAS scores were related to larger (i.e., more negative) ΔNe/ERN amplitude in the deterministic learning condition ($r = -0.25$, $p < 0.05$). However, this latter correlation failed to reach significance after partialling out the influence of overall accuracy and post-error accuracy ($p = 0.37$). As illustrated in **Figure 4**, larger Ne/ERN amplitudes were also associated with higher overall accuracy and post-error accuracy in both the deterministic and probabilistic learning condition (rs < -0.33, ps < 0.001). Considering that the negative correlation between BIS and overall performance may have disguised a relationship between BIS and Ne/ERN, we conducted partial correlations controlling for overall accuracy. Nonetheless, the correlation between punishment sensitivity and Ne/ERN remained non-significant (rs < 0.09, ps > 0.39).

FRN

As expected, higher BIS scores were related to larger FRN amplitudes in the deterministic ($r = -0.29$, $p < 0.01$), probabilistic ($r = -0.26$, $p < 0.05$), and chance condition ($r = -0.27$, $p < 0.01$) (see **Figure 4**). Similarly, self-reported punishment sensitivity correlated with the ΔFRN in the probabilistic learning and chance condition (rs < -0.19, ps < 0.05). In contrast to the Ne/ERN, however, the FRN was largely unrelated to learning

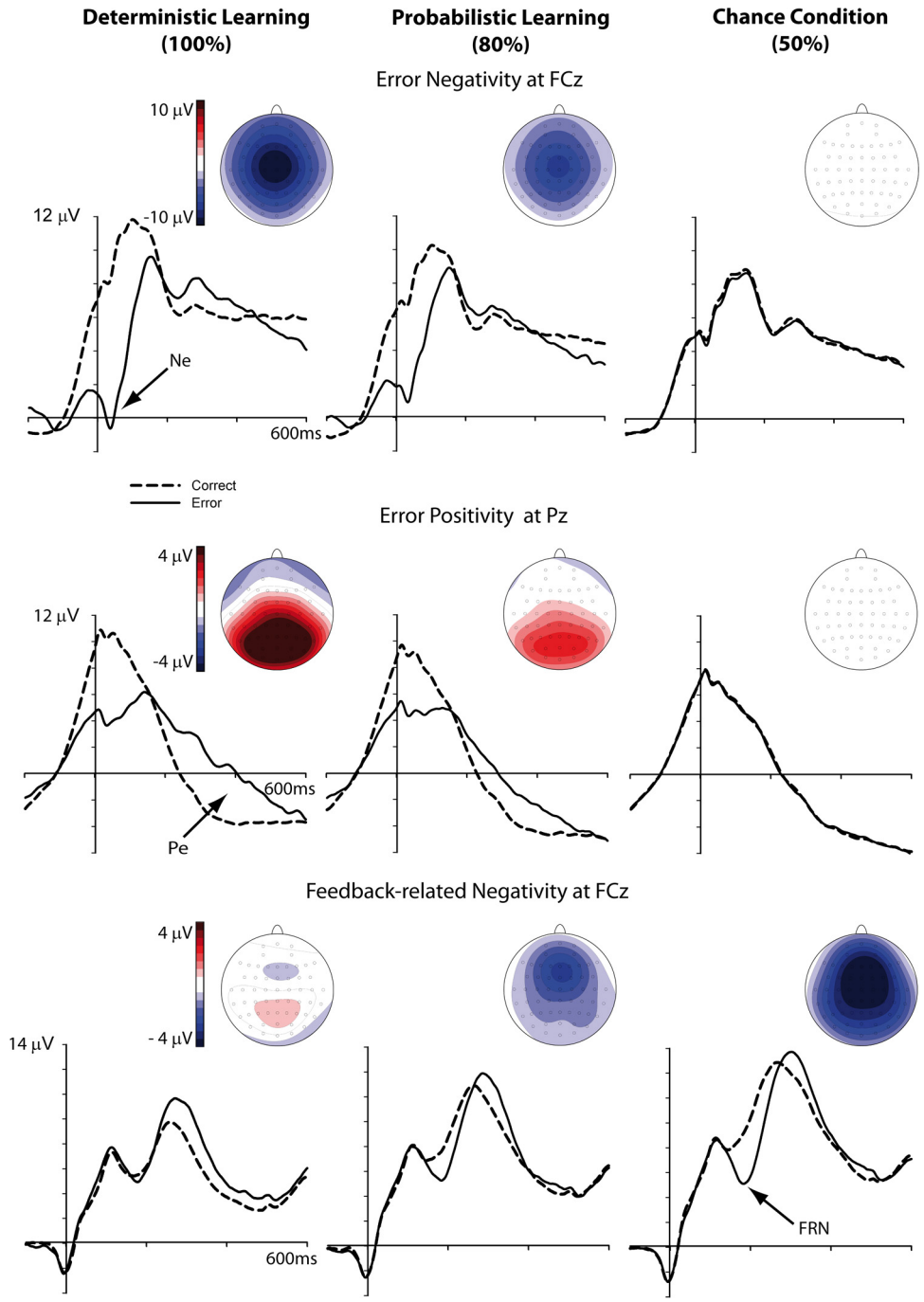


FIGURE 3 | Response- and feedback-locked ERPs on correct (dashed lines) and incorrect (solid lines) trials and corresponding topographical maps, displayed separately for the three learning conditions.

performance and post-error accuracy. Only in the deterministic learning condition, ΔFRN correlated with overall accuracy ($r = 0.31$, $p < 0.01$). Since previous studies reported an association between punishment sensitivity and larger FRN amplitudes to positive feedback (Balconi and Crivelli, 2010; Santesso et al., 2011b), we additionally tested the correlation between BIS/BAS scores and the FRN on correct trials. The analyses only revealed a marginally significant correlation between punishment sensitivity and FRN amplitude in the chance condition ($r = -0.18$, $p = 0.07$; deterministic and probabilistic learning condition: $ps > 0.15$). Furthermore, we probed the relationship between punishment sensitivity and the FRN to *invalid* negative feedback in the probabilistic learning condition. The correlation coefficient was highly similar to that observed for valid negative feedback

Table 2 | Pearson's correlations between personality measures, behavioral measures, and ERP components in the deterministic learning condition.

	BAS	Acc	AccPost	RTcorr	RTerr	Ne[a]	ΔNe[a]	FRN[a]	ΔFRN[a]	Pe	ΔPe
BIS	0.13	**−0.21**	0.04	−0.05	−0.09	−0.04	0.03	**−0.29**	0.02	−0.02	−0.08
BAS		**0.21**	**0.27**	0.03	0.05	−0.16	**−0.25**	−0.06	0.13	**0.25**	**0.26**
Acc			**0.61**	**0.34**	**0.37**	**−0.39**	**−0.44**	0.03	**0.31**	0.18	**0.32**
AccPost				0.02	0.08	**−0.45**	**−0.45**	0.05	0.16	**0.23**	**0.39**
RTcorr					**0.73**	−0.01	0.18	−0.02	0.02	−0.11	−0.09
RTerr						−0.09	0.07	−0.09	−0.09	−0.17	**−0.21**
Ne[a]							**0.41**	−0.05	−0.15	−0.04	−0.11
ΔNe[a]								−0.11	**−0.27**	**−0.24**	−0.18
FRN[a]									**0.23**	−0.06	−0.12
ΔFRN[a]										−0.12	−0.08
Pe											**0.59**

Correlation coefficients printed in bold are significant at least at $\alpha = 0.05$.

Note: BIS = punishment sensitivity, BAS = reward sensitivity, Acc = overall accuracy, AccPost = post-error accuracy, RTcorr = reaction time correct responses, RTerr = reaction time erroneous responses, Ne = error negativity (peak-to-peak measure), ΔNe = error negativity (difference wave), FRN = feedback-related negativity (peak-to-peak measure), ΔFRN = feedback-related negativity (difference wave), Pe = error positivity, ΔPe = error positivity (difference wave).

[a]Note that larger Ne/ERN and FRN amplitudes are reflected in larger negative values.

Table 3 | Pearson's correlations between personality measures, behavioral measures, and ERP components in the probabilistic learning condition.

	BAS	Acc	AccPost	RTcorr	RTerr	Ne[a]	ΔNe[a]	FRN[a, b]	ΔFRN[a, b]	Pe	ΔPe
BIS	0.13	**−0.20**	−0.12	−0.08	−0.07	−0.01	0.08	**−0.26**	**−0.20**	0.04	−0.05
BAS		0.13	0.12	0.02	0.07	−0.07	−0.14	−0.04	0.15	0.15	0.11
Acc			**0.66**	−0.02	0.03	**−0.34**	**−0.38**	−0.06	0.13	0.13	**0.23**
AccPost				0.05	−0.08	**−0.48**	**−0.46**	−0.03	0.09	0.12	0.10
RTcorr					**0.92**	−0.12	0.12	0.05	−0.14	**−0.20**	0.19
RTerr						−0.16	0.12	0.05	−0.16	−0.16	**0.23**
Ne[a]							**0.26**	−0.04	−0.03	0.06	−0.09
ΔNe[a]								0.01	−0.07	−0.09	−0.08
FRN[a,b]									−0.05	−0.06	−0.12
ΔFRN[a,b]										−0.12	−0.11
Pe											**0.50**

Correlation coefficients printed in bold are significant at least at $\alpha = 0.05$.

Note: BIS = punishment sensitivity, BAS = reward sensitivity, Acc = overall accuracy, AccPost = post-error accuracy, RTcorr = reaction time correct responses, RTerr = reaction time erroneous responses, Ne = error negativity (peak-to-peak measure), ΔNe = error negativity (difference wave), FRN = feedback-related negativity (peak-to-peak measure), ΔFRN = feedback-related negativity (difference wave), Pe = error positivity, ΔPe = error positivity (difference wave).

[a]Note that larger Ne/ERN and FRN amplitudes are reflected in larger negative values.

[b]Valid trials (a highly similar pattern of correlations was obtained for invalid trials).

($r = -0.25$, $p < 0.05$), suggesting that the relationship between BIS and FRN was not modulated by the degree of expectancy violation.

Pe

Subjects scoring higher on BAS showed greater (i.e., more positive) Pe/ΔPe amplitudes in the deterministic learning condition (rs > 0.24, ps < 0.05) (see **Figure 4**) but not in the probabilistic learning condition (rs < 0.16, ps > 0.12). However, only for the ΔPe, there was a marginally significant difference between the two correlation coefficients [$t_{(102)} = 1.34$, $p < 0.10$]. In addition, as displayed in **Figure 4**, larger ΔPe amplitudes were associated with higher overall accuracy in both learning conditions (rs > 0.22, ps < 0.05), whereas only in the deterministic learning condition, ΔPe was significantly related to post-error accuracy ($r = 0.39$, $p < 0.001$). To examine whether BAS and ΔPe contributed independently to learning performance in the deterministic learning condition, we included them as predictors in multiple regression analyses with overall and post-error accuracy as criterion. Higher overall accuracy was related to larger ΔPe amplitudes ($\beta = 0.28$, $t = 2.87$, $p < 0.01$), whereas the relationship with BAS was only marginally significant ($\beta = 0.16$, $t = 1.69$, $p = 0.09$). Similarly, higher post-error accuracy was significantly associated with larger ΔPe amplitudes ($\beta = 0.34$, $t = 3.61$, $p < 0.001$), but not with higher BAS scores ($\beta = 0.17$, $t = 1.81$, $p = 0.06$). These findings suggest that the positive relationship between reward sensitivity

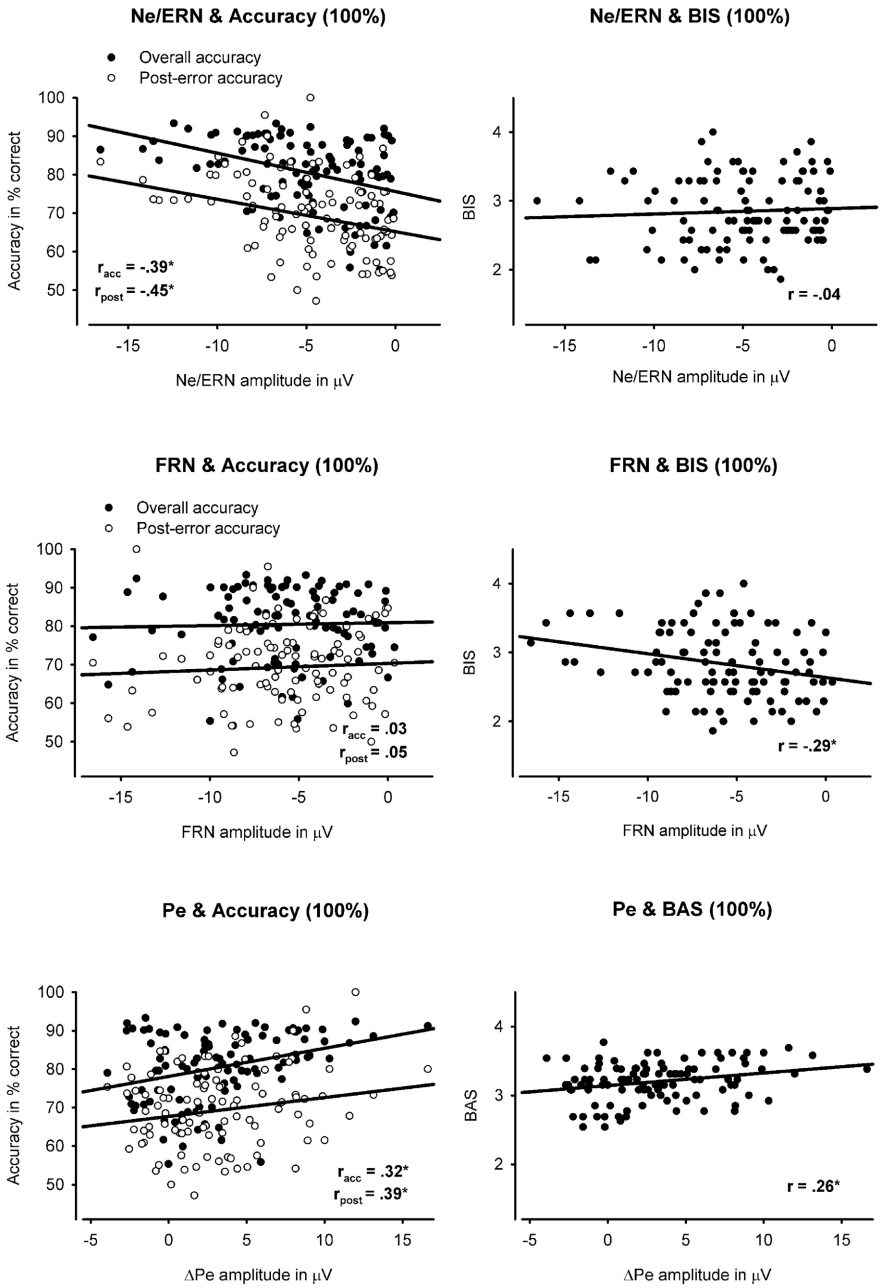


FIGURE 4 | Scatterplots showing the relationships between the ERP components (Ne/ERN, FRN, Pe) and learning performance (overall accuracy, post-error accuracy), and personality (BIS, BAS). The first row shows the correlation between the Ne/ERN (measured peak-to-peak) and learning performance (left) and BIS (right). The second row displays the correlation between the FRN (measured peak-to-peak) and learning performance (left) and BIS (right). The first row shows the correlation between the ΔPe and learning performance (left) and BAS (right).

and learning performance was partly mediated by shared variance with the Pe. We also regressed the two accuracy measures as a function of ΔPe and Ne/ERN. These analyses revealed that higher overall accuracy in the deterministic learning condition was associated with both greater ΔPe ($\beta = 0.28$, $t = 3.18$, $p < 0.01$) and Ne/ERN amplitudes ($\beta = -0.36$, $t = 4.10$, $p < 0.001$). Likewise, the two components made independent contributions to post-error accuracy ($|\beta s| > 0.33$, $|ts| > 4.08$, $ps < 0.001$). Finally, in contrast to the Ne/ERN, ΔPe correlated negatively with error RT ($rs < -0.20$, $ps < 0.05$), reflecting that faster responses on error trials were associated with smaller ΔPe amplitudes.

MODERATOR EFFECTS OF BIS AND BAS ON THE RELATIONSHIP BETWEEN ERP COMPONENTS AND BEHAVIOR

Previous research suggested that affect-related modulations in neuroelectric responses to errors may be associated with a

stronger impact of these error signals on learning-related behavioral adaptation (Cavanagh et al., 2011a,b). Therefore, in a further step, we tested whether the relation between the ERP components and behavioral adjustments varies as a function of punishment/reward sensitivity. Separate moderated multiple regression models for the deterministic and probabilistic learning condition included *BIS*, *BAS*, *ERP amplitude* (Ne/ERN vs. FRN vs. Pe), and the corresponding *interaction terms* (i.e., Ne/ERN × BIS, Ne/ERN × BAS vs. FRN × BIS, FRN × BAS vs. Pe × BIS, Pe × BAS) as predictors and *overall accuracy* vs. *post-error accuracy* as criterion.

The interaction terms were non-significant in all analyses ($|\beta s| < 0.18$, $|ts| < 1.60$, $ps > 0.10$). Thus, we did not find evidence for a moderating effect of punishment sensitivity or reward sensitivity on the relationship between the ERP components (Ne/ERN, FRN, Pe) and learning performance in terms of overall accuracy or post-error accuracy.

INFLUENCE OF PERSONALITY ON LEARNING-RELATED MODULATIONS IN THE ERP COMPONENTS

For a subsample of 68 participants who committed enough errors in both halves of the learning task to obtain reliable measures of the ERP components, Ne/ERN, FRN, and Pe amplitudes were subjected to separate ANCOVAs with the within-subject factors *learning condition* (deterministic, probabilistic, and chance condition) and *bin* (Bin 1 vs. 2) and the continuous between-subjects factors *BIS* and *BAS*. For reasons of parsimony, we will only report analyses of the peak-to-peak measures of Ne/ERN and FRN as well as analyses of ΔPe amplitudes.

Ne/ERN

The ANCOVA yielded a significant main effect of learning condition [$F_{(2, 130)} = 47.28$, $p < 0.001$, $\eta^2 = 0.42$, $\varepsilon = 0.82$]. Contrasts revealed the Ne/ERN to be larger in the deterministic compared to the probabilistic learning condition and in the two learning conditions compared to the chance condition ($ps < 0.01$, $\eta^2 s > 0.12$) (see **Figures 3** and **5**). As was indicated by an interaction of learning condition and bin [$F_{(2, 130)} = 6.97$, $p < 0.01$, $\eta^2 = 0.10$], the Ne/ERN was differentially modulated over the course of learning in the three conditions. Follow-up comparisons showed a significant increase in Ne/ERN amplitude for the deterministic learning condition only [$t_{(67)} = 1.90$, $p < 0.05$, one-tailed]. While the Ne/ERN did not reliably change from Bin 1 to Bin 2 in the probabilistic learning condition ($p = 0.33$), it decreased in the chance condition [$t_{(67)} = -3.31$, $p < 0.01$, two-tailed].

Furthermore, the analysis revealed a significant interaction between BIS, learning condition, and bin [$F_{(2, 130)} = 3.80$, $p < 0.05$, $\eta^2 = 0.06$]. As can be seen from **Figure 5**, this interaction reflects that only highly punishment sensitive individuals showed a learning-related increase of the Ne/ERN in the deterministic learning condition, whereas the Ne/ERN did not change from Bin 1 to Bin 2 for less punishment sensitive individuals (defined by median split). Follow-up correlation analyses yielded a marginally significant relation between BIS and learning-related changes in Ne/ERN amplitude (Ne_2—Ne_1) in the deterministic learning condition (partial $r = -0.24$, $p = 0.06$). The correlation between punishment sensitivity and Ne/ERN, however, was non-significant both in Bin 1 and Bin 2 (partial $rs < 0.17$, $ps > 0.18$).

FRN

The analysis yielded a significant main effect of BIS only [$F_{(1, 65)} = 11.88$, $p < 0.01$, $\eta^2 = 0.15$], reflecting that higher punishment sensitivity predicted larger FRN amplitudes (partial $r = -0.38$, $p < 0.01$; FRN collapsed across bins and learning conditions). **Figure 6** illustrates that the FRN did not reliably change over the course of learning in either the deterministic or probabilistic learning condition.

Pe

The ANCOVA revealed a reliable main effect of learning condition [$F_{(2, 130)} = 42.84$, $p < 0.001$, $\eta^2 = 0.40$]. As can be seen

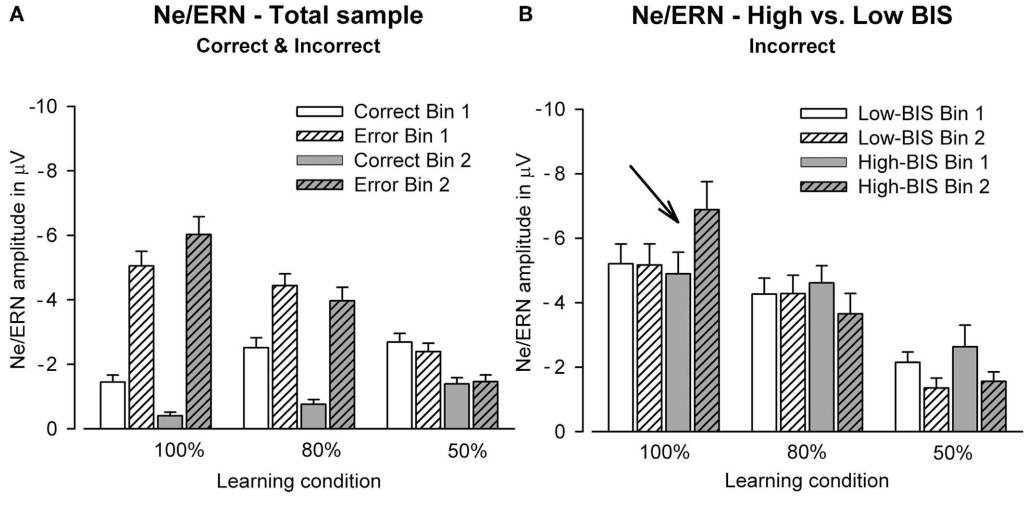


FIGURE 5 | Bar graphs show the amplitude of the Ne/ERN at FCz in Bin 1 and Bin 2 for (A) the total sample (error and correct trials) and (B) high vs. low BIS subjects (only error trials).

from **Figure 7** (see also **Figure 3**), the ΔPe was larger for the deterministic compared to the probabilistic learning conditions as well as for the two learning conditions compared to the chance condition (ps < 0.01, η^2s > 0.12). Furthermore, we found a significant main effect of bin [$F_{(1, 65)} = 43.11$, $p < 0.001$, $\eta^2 = 0.40$] and an interaction between learning condition and bin [$F_{(2, 130)} = 11.13$, $p < 0.001$, $\eta^2 = 0.15$]. Contrasts revealed that the learning-related changes in the ΔPe were larger for the deterministic and probabilistic learning condition compared to the chance condition ($p < 0.001$, $\eta^2 = 0.25$), but did not differ between the two learning conditions ($p = 0.79$). Follow-up comparisons confirmed that the ΔPe increased with learning in the deterministic and probabilistic learning condition only (ps < 0.001).

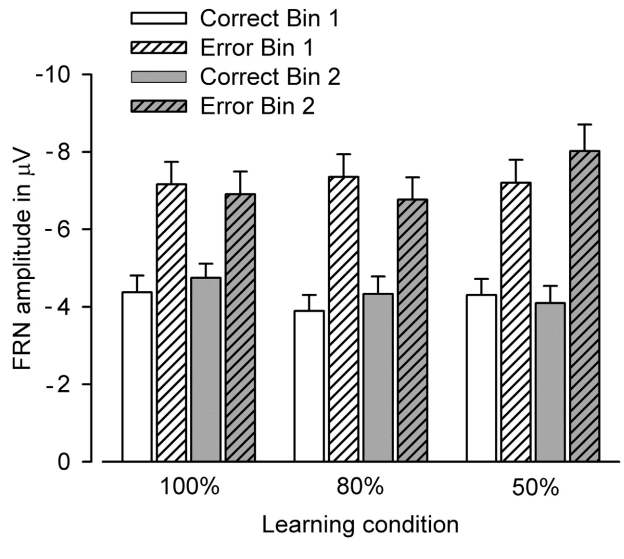


FIGURE 6 | Bar graphs show the amplitude of the FRN at FCz in Bin 1 and Bin 2 for "correct" and "incorrect" feedback.

In addition, the analysis revealed a significant interaction between BAS, learning condition, and bin [$F_{(2, 130)} = 3.61$, $p < 0.05$, $\eta^2 = 0.05$]. Contrasts showed that the interaction between bin and BAS differed for the deterministic compared to the probabilistic learning condition ($p < 0.05$, $\eta^2 = 0.07$) but not for the two learning conditions compared to the chance condition ($p = 0.10$). **Figure 7** illustrates that highly reward sensitive individuals showed a more pronounced learning-related increase in ΔPe amplitude in the deterministic learning condition than subjects with lower BAS scores. In line with this, correlation analyses yielded a significant relationship between reward sensitivity and the increase of ΔPe from Bin 1 to Bin 2 (ΔPe_2—ΔPe_1) for the deterministic learning condition (partial $r = 0.29$, $p < 0.05$). Notably, we found a significant correlation between BAS and ΔPe in Bin 2 (partial $r = 0.29$, $p < 0.05$) but not in Bin 1 ($p = 0.81$).

SUMMARY OF MAIN FINDINGS

Analyses of accuracy data showed that higher reward sensitivity was associated with better overall learning performance and higher post-error accuracy in the deterministic learning condition. Conversely, and contrary to our predictions, higher punishment sensitivity was associated with impaired performance both in the deterministic and probabilistic learning condition, but was not related to post-error accuracy in either of the two conditions.

Critically, correlation analyses did not reveal a significant relationship between punishment sensitivity and Ne/ERN. However, as expected, larger Ne/ERN amplitudes were associated with better learning performance and higher post-error accuracy. Moreover, punishment sensitivity modulated learning-related changes of the Ne/ERN. Only for highly punishment sensitive individuals, we found an increase of the Ne/ERN over the course of learning in the deterministic learning condition.

In line with prior studies, higher punishment sensitivity was associated with enhanced FRN amplitudes. Interestingly, this relationship appeared to be insensitive to feedback validity. In

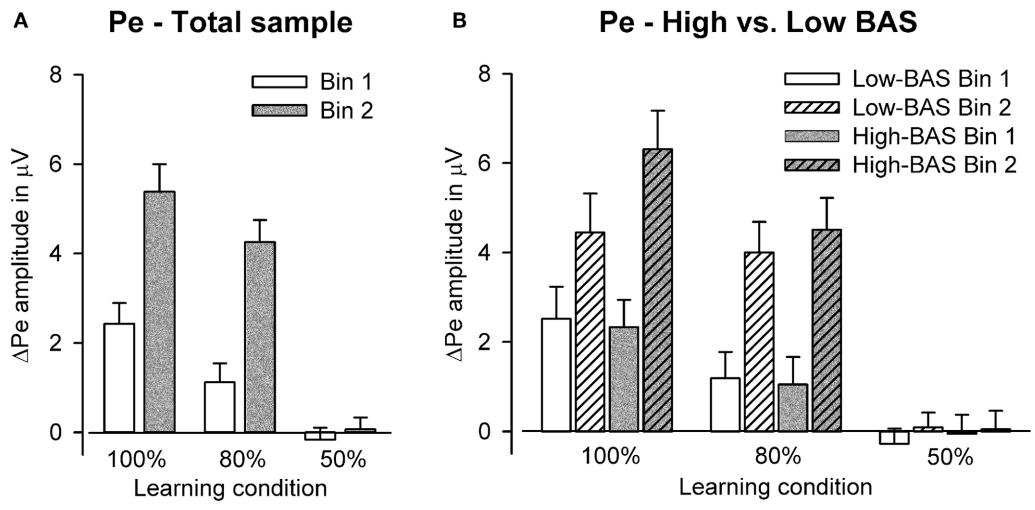


FIGURE 7 | Bar graphs show the amplitude of the ΔPe at Pz in Bin 1 and Bin 2 for (A) the total sample and (B) high vs. low BAS subjects.

contrast to the Ne/ERN, the FRN was not clearly related to learning performance.

Furthermore, the present results replicate prior findings that higher reward sensitivity relates to larger Pe amplitudes, but this was only the case toward the end of learning in the deterministic learning condition. Moreover, participants highly sensitive to reward showed a more pronounced learning-related increase of the Pe in the deterministic learning condition. Similar to the Ne/ERN, greater Pe amplitudes were associated higher overall and post-error accuracy.

Finally, we found no evidence that individual differences in punishment or reward sensitivity modulate the relationship between error- and feedback-processing—as reflected in the Ne/ERN, FRN, and Pe—and learning-related behavioral adjustments.

DISCUSSION

Numerous reports have suggested that individual differences in punishment (BIS/FFFS) and reward sensitivity (BAS) are reflected in neurocognitive mechanisms of error and feedback processing. The main goal of the present investigation was to further examine the impact of these interactions between affect-related traits and action monitoring on the ability to use error signals for behavioral adaptation during reinforcement learning. In contrast to previous studies employing simple motor tasks, such as the Flankers and Go/No-Go task (Boksem et al., 2006, 2008; Amodio et al., 2008), we found no relation between punishment sensitivity and the Ne/ERN. However, consistent with past research, higher punishment sensitivity was related to larger FRN amplitudes (Balconi and Crivelli, 2010; De Pascalis et al., 2010; Santesso et al., 2011b). These results indicate that highly punishment sensitive individuals were characterized by an enhanced responsivity to external rather than internal error cues. Furthermore, higher reward sensitivity was associated with increased neural responses during later stages of error processing as reflected in the Pe, replicating prior findings (Boksem et al., 2006, 2008). Although both FRN and Pe are thought to play a functional role in post-error adaptation, only reward sensitivity was related to better overall learning performance and higher post-error accuracy. By contrast, participants with higher trait sensitivity to punishment showed impaired learning performance.

The negative correlation between punishment sensitivity and overall accuracy was somewhat surprising, as higher BIS-reactivity has been claimed to trigger enhanced attention and information processing (Gray and McNaughton, 2000; Smillie, 2008). Still, BIS-activation has also been linked to anxious rumination and worry, which might interfere with task-related processing such as updating of S-R mappings. Moreover, as was pointed out by Pickering and colleagues (1997), learning tasks involving both rewards and punishments can cause mutually inhibitory interactions between BIS/FFFS and BAS. One should note that learning was accompanied by an increasing proportion of positive feedback, perhaps shifting the balance between the two systems toward a relative dominance of the BAS. Thus, relatively stronger reward reactivity may have contributed to better overall performance in less punishment sensitive individuals by facilitating appetitive learning or proactive engagement (Corr, 2004; Braver et al., 2007).

Given the comparatively large sample size, the lack of BIS/FFFS-related variations in Ne/ERN amplitude was unlikely to reflect insufficient statistical power, at least if the effect size is assumed to be small to moderate. One might argue that the negative correlation between punishment sensitivity and overall accuracy on the one hand, and the positive correlation between overall accuracy and Ne/ERN magnitude on the other hand, have neutralized the relationship between punishment sensitivity and Ne/ERN. Partial correlation analysis controlling for overall learning performance suggested that this was not the case. There is also no indication that the correlation coefficient was deflated due to restricted variability of BIS scores. However, the Ne/ERN was relatively small as is typically the case when using probabilistic learning tasks, in which participants are less certain about the correctness of their responses. It is thus possible that reduced variability of the Ne/ERN has decreased the probability of obtaining a significant correlation with punishment sensitivity.

Otherwise, it has been suggested that the delivery of trial-to-trial performance feedback leads participants to rely more strongly on external than internal error cues (Nieuwenhuis et al., 2005). This might be especially true for individuals highly sensitive to punishment as they appear to be characterized by low-level personal agency, which means that their actions are controlled by environmental cues rather than internal standards (Balconi and Crivelli, 2010). The unique association between punishment sensitivity and FRN found in the present study is consistent with this view. Interestingly, the relationship did not vary as a function of feedback validity or learning, suggesting that highly punishment sensitive individuals were generally more vigilant to negative feedback cues, irrespective of whether they were unexpected or not. Moreover, we found no clear evidence for a relation between punishment sensitivity and the FRN to positive feedback, consistent with what has been reported for individuals high in trait negative affect as well as moderately depressed subjects (Tucker et al., 2003; Sato et al., 2005; Santesso et al., 2011a). Thus, while punishment sensitivity has also been shown to be associated with an increased FRN elicited by unexpected (large) rewards (Santesso et al., 2011b), our findings indicate that highly punishment sensitive individuals are particularly characterized by enhanced mPFC responses to environmental cues signaling punishment. However, future studies should determine under what circumstances positive feedback elicits increased FRN amplitudes in highly punishment sensitive and whether these modulations reflect blunted responses to reward or higher vigilance to both positive and negative performance feedback.

Although high trait-level sensitivity to punishment was not associated with an overall enhancement of Ne/ERN amplitudes, self-reported BIS/FFFS-reactivity modulated learning-related changes of this component. The Ne/ERN increased with learning of the S-R mappings only for highly punishment sensitive individuals in deterministic learning condition, whereas no learning-related changes in Ne/ERN amplitude were observed for less punishment sensitive individuals or in the probabilistic learning condition. An explanation of this finding could be that highly punishment sensitive individuals were less prone to motivational

disengagement. Punishment sensitivity has been linked to higher persistence, reflected in a relatively smaller decrease in behavioral performance and Ne/ERN amplitude with increasing time on task (Boksem et al., 2006; Tops and Boksem, 2010). Thus, disengagement could have attenuated a learning-related enhancement of the Ne/ERN more clearly for individuals with low compared to high BIS scores. This explanation, however, leaves open the question of why higher punishment sensitivity was related to worse overall performance. Further studies are necessary to clarify whether this might reflect differences ability to use positive feedback for behavioral adaptation.

Previous ERP studies have shown that BIS/FFFS-related differences in mPFC functioning are more pronounced in aversive compared to appetitive motivational contexts and in response to intense negative events (Boksem et al., 2008; Santesso et al., 2011a). The motivational context could be an important determinant of whether or not punishment sensitivity is also reflected in higher responsivity to internal indicators of response errors, even if continuous external performance feedback is provided. Indeed, we recently found that highly punishment sensitive participants showed a larger Ne/ERN to errors resulting in loss or gain omission during a learning task involving trial-to-trial manipulation of incentive value (Unger and Kray, in preparation). By contrast, consistent with the present results, punishment sensitivity did not relate to Ne/ERN amplitude on neutral trials. Interestingly, the association between punishment sensitivity and Ne/ERN was stronger at the beginning than at the end of learning, arguing against the view that undetermined S-R mappings *per se* account for the present null-finding. Under threatening conditions, activity of the medial prefrontal performance monitoring system appears to be more sensitive to individual differences in self-reported BIS/FFFS-reactivity when the optimal course of action is uncertain and cognitive control demands are high.

According to a recent proposal, the ACC integrates punishment-related information from multiple sources in order to support instrumental behaviors, particularly in unstable and threatening environments (Shackman et al., 2011). From this perspective, the relation between punishment sensitivity and FRN might reflect that affect-related traits bias cognitive processing and regulate action selection in accordance with an individual's overarching goals and beliefs (Huys and Dayan, 2009; Cavanagh et al., 2011a). Even so despite the proposed link between FRN and future behavioral adaptation (Holroyd and Coles, 2002; Frank et al., 2005), accuracy data suggest that larger error signals to negative feedback in highly punishment sensitive individuals were not beneficial for learning or may even reflect dysfunctional processing. One interpretation of this finding could be that the FRN enhancement is primarily related to the regulation of negative emotions (Pizzagalli, 2011; Santesso et al., 2011a,b). The ACC has been assigned an important role in controlling amygdala responsivity to fear-related stimuli. Dysregulated interactions between ACC and amygdala may be associated with a negative processing bias that is reflected in enhanced attentional capture by potential threat cues, anxious rumination, and inability to disengage from negative events and have been linked to anxiety and depression (Bishop, 2007; Pizzagalli, 2011). Moreover, it may be important to consider that rapid trial-to-trial adjustments as assessed in the current investigation are thought to primarily reflect explicit/declarative learning (Frank et al., 2007b). Previous research, however, suggests that individual differences in punishment sensitivity rather affect implicit/habitual learning. In particular, Cavanagh and colleagues (2011a,b) showed that increased mPFC responses to negative feedback in punishment hypersensitive participants were specifically associated with alterations in slow integrative avoidance learning, presumably mediated by phylogenetically old non-declarative learning systems.

The second set of findings from our study concerns the relationship between reward sensitivity and Pe. In line with previous reports (Boksem et al., 2006, 2008), self-reported reward sensitivity correlated positively with the magnitude of the Pe. However, this relationship was only significant during later stages of learning in the deterministic learning condition, indicating that it depended on the participants' ability to internally represent the correct response. Further corroborating this notion, higher BAS scores were related to a more pronounced learning-related increase in Pe amplitude in the deterministic learning condition.

Drawing on the proposal that there is a link between approach motivation and a bias toward proactive control (Braver et al., 2007), Boksem and colleagues (2006, 2008) suggested that larger Pe amplitudes in highly reward sensitive individuals are functionally related to subsequent engagement in proactive behaviors. Our finding that greater Pe amplitudes were associated with higher overall accuracy and post-error accuracy seems consistent with the proposed link. Although strictly speaking, for action control to be implemented proactively, predictive contextual cues have to be present prior to the imperative stimulus (Braver et al., 2007). This is typically not the case during reinforcement learning, presumably limiting the utility of proactive strategies in a narrow sense. Nonetheless, it is conceivable that highly reward sensitive individuals tend to respond to errors with positive approach behaviors such as reactivation of the potentially disrupted representation of the correct S-R mappings. The idea that BAS-related modulations of the Pe reflect active updating of task-set representations in working memory corresponds to previous reports stressing the morphological and functional similarity between the Pe and the stimulus-evoked P300 (Leuthold and Sommer, 1999; Davies et al., 2001; Overbeek et al., 2005). In this regard, it seems noteworthy that high reward sensitivity has also been found to be associated with enhanced P300 amplitudes to negative feedback (Balconi and Crivelli, 2010).

Although the neurobiological basis of the BAS has been described in terms of dopaminergic mechanisms (Gray and McNaughton, 2000; Smillie, 2008), the Pe and the P300 have primarily been linked to noradrenergic neurotransmission (Nieuwenhuis et al., 2005; Frank et al., 2007a). Moreover, the Pe has previously been found to be affected by functional polymorphisms of the serotonin transporter gene, possibly mediated by its regulatory influence on the amygdala (Althaus et al., 2009; but see Beste et al., 2010). Despite the pivotal role that dopamine is assumed to play in the generation of the FRN (Holroyd and Coles, 2002), serotonergic functioning is also likely to be involved in the

observed relationship between punishment sensitivity and FRN amplitude. Several reports showed that genetic and pharmacological variations in serotonergic neurotransmission are accompanied by changes in mPFC responses to errors and conflict as well as amygdala/hippocampus reactivity to aversive and threatening stimuli (Canli et al., 2005; Cools et al., 2005; Chamberlain et al., 2006; Harmer et al., 2006; Finger et al., 2007). In addition, variations in serotonin transmission have been associated with individual differences in anxiety and depression-related traits (Sen et al., 2004). It has been proposed that the modulatory influence of serotonin on the prefrontal dopamine system may constitute the neurophysiological basis of altered action monitoring functions in individuals high in negative affectivity, including anxiety and depression (Beste et al., 2010). Clearly, more research is needed to determine whether opponency between the serotonergic and dopaminergic system underlies cognitive-affective interactions in learning and decision making (Cools et al., 2008; Jocham and Ullsperger, 2009).

Some limitations of the present study should be noted. First, the observed effects of personality measures on ERP correlates of error and feedback processing were rather small-sized ($r \leq 0.30$), particularly when compared to the relationship between accuracy measures and ERP components. Although larger correlation coefficients have been reported in the literature, these were typically derived from small samples and hence likely to be inflated (Ioannidis, 2008). Note that the strength of the relations is already constrained by the internal reliability of the BIS/BAS measures (Cronbachs $\alpha = 0.73/0.59$). Second, the current investigation included a very homogeneous sample of under-graduate university students. It is possible that higher correlations will be found in more heterogeneous samples such as clinical populations or different age groups. Finally, the present study reported only correlational data, leaving unspecified the direction of the observed effects.

To summarize and conclude, the present study shows that individual differences in punishment sensitivity are associated with larger FRN amplitudes, indicating an increased mPFC responsivity to negative performance feedback. However, the negative correlation between punishment sensitivity and overall accuracy suggests that the alterations in mPFC functioning are not beneficial for learning-related behavioral adaptation and may reflect non-adaptive forms of emotion regulation. Future research is needed to determine whether the negative processing bias specifically affects incremental habitual learning mechanisms rather than rapid trial-to-trial adjustments as assessed in the current task. Furthermore, higher reward sensitivity was related to larger Pe amplitudes and better learning performance, suggesting that self-reported BAS-reactivity is associated with an enhanced use of deliberate proactive strategies to support future performance. Importantly, the Pe and the Ne/ERN appeared to make independent contributions to overall learning performance and error-related behavioral adjustments, consistent with the notion that the two components reflect activity of separable action monitoring systems, which may mediate automatic vs. more controlled forms of post-error adaptation (cf. Ridderinkhof et al., 2009). In line with previous studies, the present findings indicate that individual differences in reward and punishment sensitivity are associated with unique functional alterations of these systems.

ACKNOWLEDGMENTS

This work was funded by the German Research Foundation (Deutsche Forschungsgesellschaft; grant IRTG 1457). We gratefully thank Michael Herbert, Anna Orth, Svenja Schieren, Verena Schnitzler, and Jenny Sinzig for help during data acquisition.

REFERENCES

Althaus, M., Groen, Y., Wijers, A. A., Mulder, L. J., Minderaa, R. B., Kema, I. P., Dijck, J. D., Hartman, C. A., and Hoekstra, P. J. (2009). Differential effects of 5-HTTLPR and DRD2/ANKK1 polymorphisms on electrocortical measures of error and feedback processing in children. *Clin. Neurophysiol.* 120, 93–107.

Amodio, D. M., Master, S. L., Yee, C. M., and Taylor, S. E. (2008). Neurocognitive components of the behavioral psychophysiology inhibition and activation systems: implications for theories of self-regulation. *Psychophysiology* 45, 11–19.

Balconi, M., and Crivelli, D. (2010). FRN and P300 ERP effect modulation in response to feedback sensitivity: the contribution of punishment-reward system (BIS/BAS) and behaviour identification of action. *Neurosci. Res.* 66, 162–172.

Beaver, J. D., Lawrence, A. D., van Ditzhuijzen, J., Davis, M. H., Woods, A., and Calder, A. J. (2006). Individual differences in reward drive predict neural responses to images of food. *J. Neurosci.* 26, 5160–5166.

Beste, C., Domschke, K., Kolev, V., Yordanova, J., Baffa, A., Falkenstein, M., and Konrad, C. (2010). Functional 5-HT1a receptor polymorphism selectively modulates error-specific subprocesses of performance monitoring. *Hum. Brain Mapp.* 31, 621–630.

Bishop, S. J. (2007). Neurocognitive mechanisms of anxiety: an integrative account. *Trends Cogn. Sci.* 11, 307–316.

Boksem, M. A. S., Meijman, T. F., and Lorist, M. M. (2006). Mental fatigue, motivation and action monitoring. *Biol. Psychol.* 72, 123–132.

Boksem, M. A. S., Tops, M., Kostermans, E., and De Cremer, D. (2008). Sensitivity to punishment and reward omission: evidence from error-related ERP components. *Biol. Psychol.* 79, 185–192.

Braver, T. S., Gray, J. R., and Burgess, G. C. (2007). "Explaining the many varieties of working memory variation: dual mechanisms of cognitive control," in *Variation in Working Memory,* eds A. R. A. Conway, C. Jarrold, M. Kane, A. Miyake, and J. N. Towse (Oxford: Oxford University Press), 76–106.

Canli, T., Omura, K., Haas, B. W., Fallgatter, A., Constable, R. T., and Lesch, K. P. (2005). Beyond affect: a role for genetic variation of the serotonin transporter in neural activation during a cognitive attention task. *Proc. Natl. Acad. Sci. U.S.A.* 102, 12224–12229.

Carver, C. S., and White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the BIS/BAS scales. *J. Pers. Soc. Psychol.* 67, 319–333.

Cavanagh, J. F., and Allen, J. J. B. (2008). Multiple aspects of the stress response under social evaluative threat: an electrophysiological investigation. *Psychoneuroendocrinology* 33, 41–53.

Cavanagh, J. F., Frank, M. J., and Allen, J. J. B. (2011a). Social stress reactivity alters reward and punishment learning. *Soc. Cogn. Affect. Neurosci.* 6, 311–320.

Cavanagh, J. F., Bismark, A. J., Frank, M. J., and Allen, J. B. (2011b). Larger error signals in major depression are associated with better avoidance learning. *Front. Psychol.* 2:331. doi: 10.3389/fpsyg.2011.00331

Chamberlain, S. R., Muller, U., Blackwell, A. D., Clark, L., Robbins, T. W., and Sahakian, B. J. (2006). Neurochemical modulation of response inhibition and probabilistic learning in humans. *Science* 311, 861–863.

Cools, R., Calder, A. J., Lawrence, A. D., Clark, L., Bullmore, E., and Robbins, T. W. (2005). Individual differences in threat sensitivity predict serotonergic modulation of

amygdala response to fearful faces. *Psychopharmacology* 180, 670–679.
Cools, R., Roberts, A. C., and Robbins, T. W. (2008). Serotoninergic regulation of emotional and behavioural control processes. *Trends Cogn. Sci.* 12, 31–40.
Corr, P. J. (2002). Gray's reinforcement sensitivity theory: tests of the joint subsystems hypothesis of anxiety and impulsivity. *Pers. Individ. Dif.* 33, 511–532.
Corr, P. J. (2004). Reinforcement sensitivity theory and personality. *Neurosci. Biobehav. Rev.* 28, 317–332.
Davies, P. L., Segalowitz, S. J., Dywan, J., and Pailing, P. E. (2001). Error-negativity and positivity as they relate to other ERP indices of attentional control and stimulus processing. *Biol. Psychol.* 56, 191–206.
Delaney, H. D., and Maxwell, S. E. (1981). On using analysis of covariance in repeated measures designs. *Multivariate Behav. Res.* 16, 105–123.
Dennis, T. A., and Chen, C. C. (2009). Trait anxiety and conflict monitoring following threat: an ERP study. *Psychophysiology* 46, 122–131.
De Pascalis, V., Varriale, V., and D'Antuono, L. (2010). Event-related components of the punishment and reward sensitivity. *Clin. Neurophysiol.* 121, 60–76.
Depue, R. A., and Collins, P. F. (1999). Neurobiology of the structure of personality: dopamine, facilitation of incentive motivation, and extraversion. *Behav. Brain Sci.* 22, 491–517.
Devinsky, O., Morrell, M. J., and Vogt, B. A. (1995). Contributions of anteror cingulate cortex to behaviour. *Brain* 118, 279–306.
Endrass, T., Reuter, B., and Kathmann, N. (2007). ERP correlates of conscious error recognition: aware and unaware errors in an antisaccade task. *Eur. J. Neurosci.* 26, 1714–1720.
Eppinger, B., Kray, J., Mock, B., and Mecklinger, A. (2008). Better or worse than expected? Aging, learning, and the Ne/ERN. *Neuropsychologia* 46, 521–539.
Falkenstein, M., Hohnsbein, J., Hoormann, J., and Blanke, L. (1990). "Effects of errors in choice reaction tasks on the ERP under focused and divided attention," in *Psychophysiological Brain Research,* eds C. H. M. Brunia, A. W. K. Gaillard, and A. Kok (Tilburg: TilburgUniversity Press), 192–195.
Finger, E. C., Marsh, A. A., Buzas, B., Kamel, N., Rhodes, R., Vythilingham, M., Pine, D. S., Goldman, D., and Blair, J. R. (2007). The impact of tryptophan depletion and 5-HTTLPR genotype on passive avoidance and response reversal instrumental learning tasks. *Neuropsychopharmacology* 32, 206–215.
Frank, M. J. (2005). Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. *J. Cogn. Neurosci.* 17, 51–72.
Frank, M. J., D'Lauro, C., and Curran, T. (2007a). Cross-task individual differences in error processing: neural, electrophysiological, and genetic components. *Cogn. Affect. Behav. Neurosci.* 7, 297–308.
Frank, M. J., Moustafa, A. A., Haughey, H. M., Curran, T., and Hutchison, K. E. (2007b). Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proc. Natl. Acad. Sci. U.S.A.* 104, 16311–16316.
Frank, M. J., Woroch, B. S., and Curran, T. (2005). Error-related negativity predicts reinforcement learning and conflict biases. *Neuron* 47, 495–501.
Gehring, J. W., Goss, B., Coles, M. G., Meyer, D. E., and Donchin, E. (1993). A neural system for error detection and compensation. *Psychol. Sci.* 4, 385–390.
Gehring, W. J., and Willoughby, A. R. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. *Science* 295, 2279–2282.
Geisser, S., and Greenhouse, S. W. (1958). An extension of box's results on the use of the F-distribution in multivariate analysis. *Ann. Math. Stat.* 29, 885–891.
Gray, J. A. (1982). *The Neuropsychology of Anxiety: An Enquiry into the Functions of the Septo-Hippocampal System.* Oxford: Oxford University Press.
Gray, J. A., and McNaughton, N. (2000). *The Neuropsychology of Anxiety: An Enquiry into the Functions of the Septo-Hippocampal System.* Oxford: Oxford University Press.
Gründler, T. O. J., Cavanagh, J. F., Figueroa, C. M., Frank, M. J., and Allen, J. J. B. (2009). Task-related dissociation in ERN amplitude as a function of obsessive-compulsive symptoms. *Neuropsychologia* 47, 1978–1987.
Hahn, T., Dresler, T., Ehlis, A. C., Plichta, M. M., Heinzel, S., Polak, T., Lesch, K. P., Breuer, F., Jakob, P. M., and Fallgatter, A. J. (2009). Neural response to reward anticipation is modulated by Gray's impulsivity. *Neuroimage* 46, 1148–1153.
Hajcak, G., McDonald, N., and Simons, R. F. (2003). Anxiety and error-related brain activity. *Biol. Psychol.* 64, 77–90.
Hajcak, G., McDonald, N., and Simons, R. F. (2004). Error-related psychophysiology and negative affect. *Brain Cogn.* 56, 189–197.
Harmer, C. J., Mackay, C. E., Reid, C. B., Cowen, P. J., and Goodwin, G. M. (2006). Antidepressant drug treatment modifies the neural processing of nonconscious threat cues. *Biol. Psychiatry* 59, 816–820.
Holroyd, C. B., and Coles, M. G. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychol. Rev.* 109, 679–709.
Huys, Q. J., and Dayan, P. (2009). A Bayesian formulation of behavioral control. *Cognition* 113, 314–328.
Ioannidis, J. P. (2008). Why most discovered true associations are inflated. *Epidemiology* 19, 640–648.
Jocham, G., and Ullsperger, M. (2009). Neuropharmacology of performance monitoring. *Neurosci. Biobehav. Rev.* 33, 48–60.
Kuhl, J. (1994). "Action and state orientation: psychometric properties of the action control scales (ACS-90)," in *Volition and Personality: Action versus State Orientation,* eds J. Kuhl and J. Beckmann (Göttingen: Hogrefe), 47–59.
Leone, L., Perugini, M., Bagozzi, R. P., Pierro, A., and Mannetti. L. (2001). Construct validity and generalizability of the carver-white behavioural inhibition system/behavioural activation system scales. *Eur. J. Pers.* 15, 373–390.
Leuthold, H., and Sommer, W. (1999). ERP correlates of error processing in spatial S-R compatibility tasks. *Clin. Neurophysiol.* 110, 342–357.
McNaughton, N., and Corr, P. J. (2004). A two-dimensional neuropsychology of defense: fear/anxiety and defensive distance. *Neurosci. Biobehav. Rev.* 28, 285–305.
Miltner, W. H. R., Braun, C. H., and Coles, M. G. H. (1997). Event-related brain potentials following incorrect feedback in a time estimation task: evidence for a "generic" neural system for error detection. *J. Cogn. Neurosci.* 9, 788–798.
Nieuwenhuis, S., Nielen, M. M., Mol, N., Hajcak, G., and Veltman, D. J. (2005). Performance monitoring in obsessive–compulsive disorder. *Psychiatry Res.* 134, 111–122.
Nieuwenhuis, S., Ridderinkhof, K. R., Blom, J., Band, G. P. H., and Kok, A. (2001). Error-related brain potentials are differentially related to awareness of response errors: evidence from an antisaccade task. *Psychophysiology* 38, 752–760.
O'Connell, R. G., Dockree, P. M., Bellgrove, M. A., Kelly, S. P., Hester, R., Garavan, H., and Foxe, J. J. (2007). The role of cingulate cortex in the detection of errors with and without awareness: a high-density electrical mapping study. *Eur. J. Neurosci.* 25, 2571–2579.
Overbeek, T. J. M., Nieuwenhuis, S., and Ridderinkhof, K. R. (2005). Dissociable components of error processing-on the functional significance of the Pe vis-à-vis the ERN/Ne. *J. Psychophysiol.* 19, 319–329.
Pickering, A. D., Corr, P. J., Powell, J. H., Kumari, V., Thornton, J. C., and Gray, J. A. (1997). "Individual differences in reactions to reinforcing stimuli are neither black nor white: to what extent are they Gray?" in *The Scientific Study of Human Nature: Tribute to Hans J. Eysenck At Eighty,* ed H. Nyborg (London: Elsevier Sciences), 36–67.
Pickering, A. D., and Gray, J. A. (2001). "Dopamine, appetitive reinforcement, and the neuropsychology of human learning: an individual differences approach," in *Advances in Individual Differences Research,* eds A. Eliaszand and A. Angleitner (Lengerich, Germany: PABST Science Publishers), 113–149.
Pizzagalli, D. A. (2011). Frontocingulate dysfunction in depression: towards biomarkers of treatment response. *Neuropsychopharmacol. Rev.* 36, 183–206.
Ridderinkhof, K. R., Ramautar, J. R., and Wijnen, J. G. (2009). To Pe or not to Pe: a P3-like ERP component reflecting the processing of response errors. *Psychophysiology* 46, 531–538.
Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., and Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science* 306, 443–446.
Santesso, D. L., Bogdan, R., Birk, J. L., Goetz, E. L., Holmes, A. J., and Pizzagalli, D. A. (2011a). Neural responses to negative feedback are related to negative emotionality in healthy adults. *Soc. Cogn. Affect. Neurosci.* doi: 10.1093/scan/nsr054. [Epub ahead of print].
Santesso, D. L., Dzyundzyak, A., and Segalowitz, S. J. (2011b). Age,

sex and individual differences in punishment sensitivity: factors influencing the feedback-related negativity. *Psychophysiology* 48, 1481–1489.

Sato, A., Yasuda, A., Ohira, H., Miyawaki, K., Nishikawa, M., Kumano, H., and Kuboki, T. (2005). Effects of value and reward magnitude on feedback negativity and P300. *Neuroreport* 16, 407–411.

Sen, S., Burmeister, M., and Ghosh, D. (2004). Meta-analysis of the association between a serotonin transporter promoter polymorphism (5-HTTLPR) and anxiety-related personality traits. *Am. J. Med. Genet.* 127B, 85–89.

Shackman, A. J., Salomons, T. V., Slagter, H. A., Fox, A. S., Winter, J. J., and Davidson, R. J. (2011). The integration of negative affect, pain and cognitive control in the cingulated cortex. *Nat. Rev. Neurosci.* 12, 154–167.

Simon, J. J., Walther, S., Fiebach, C. J., Friederich, H. C., Stippich, C., Weisbrod, M., and Kaiser, S. (2010). Neural reward processing is modulated by approach- and avoidance-related personality traits. *Neuroimage* 49, 1868–1874.

Smillie, L. D. (2008). What is reinforcement sensitivity? Neuroscience paradigms for approach-avoidance process theories of personality. *Eur. J. Pers.* 22, 359–384.

Snodgrass, J. G., and Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *J. Exp. Psychol. Hum. Learn. Mem.* 6, 174–215.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychol. Bull.* 87, 245–251.

Strobel, A., Beauducel, A., Debener, S., and Brocke, B. (2001). Psychometrische und strukturelle Merkmale einer deutschsprachigen Version des BIS/BAS Fragebogens von Carver und White [Psychometric and structural features of a German version of the BIS/BAS scales of Carver and White]. *Zeitschrift für Differentielle und Diagnostische Psychologie* 22, 216–227.

Taylor, S. F., Stern, E. R., and Gehring, W. J. (2007). Neural systems for error monitoring: recent findings and theoretical perspectives. *Neuroscientist* 13, 160–172.

Tops, M., and Boksem, M. A. S. (2010). Absorbed in the task. Personality measures predict engagement during task performance as tracked by error negativity and asymmetrical frontal activity. *Cogn. Affect. Behav. Neurosci.* 20, 441–453.

Tucker, D. M., Luu, P., Frishkoff, G., Quiring, J., and Poulsen, C. (2003). Frontolimbic response to negative feedback in clinical depression. *J. Abnorm. Psychol.* 112, 667–678.

Unger, K., Kray, J., and Mecklinger, A. (2012). Worse than feared? Failure induction modulates the electrophysiolgical signature of error monitoring during subsequent learning. *Cogn. Affect. Behav. Neurosci.* 12, 34–51.

Van Veen, V., and Carter, C. S. (2002). The timing of action-monitoring processes in the anterior cingulate cortex. *J. Cogn. Neurosci.* 14, 593–602.

Van den Berg, I., Franken, I. H., and Muris, P. (2011). Individual differences in sensitivity to reward: association with electrophysiological responses to monetary gains and losses. *J. Psychophysiol.* 25, 81–86.

van der Helden, J., Boksem, M. A. S., and Blom, J. H. G. (2010). The importance of failure: feedback related negativity predicts motor learning efficiency. *Cereb. Cortex* 20, 1596–1603.

Watson, D., Clark, L. A., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *J. Pers. Soc. Psychol.* 54, 1063–1070.

Wiswede, D., Münte, T. F., and Rüsseler, J. (2009). Negative affect induced by derogatory verbal feedback modulates the neural signature of error detection. *Soc. Cogn. Affect. Neurosci.* 4, 227–237.

Yeung, N., Botvinick, M. M., and Cohen, J. D. (2004). The neural basis of error detection: conflict monitoring and the error-related negativity. *Psychol. Rev.* 111, 931–959.

Yeung, N., and Sanfey, A. G. (2004). Independent coding of reward magnitude and valence in the human brain. *J. Neurosci.* 24, 6258–6264.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 16 March 2012; paper pending published: 28 March 2012; accepted: 04 June 2012; published online: 27 June 2012.

*Citation: Unger K, Heintz S and Kray J (2012) Punishment sensitivity modulates the processing of negative feedback but not error-induced learning. Front. Hum. Neurosci. **6**:186. doi: 10.3389/fnhum.2012.00186*

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 09 November 2011
doi: 10.3389/fnhum.2011.00131

Social comparison affects brain responses to fairness in asset division: an ERP study with the ultimatum game

Yin Wu*[1,2,3], *Yuqin Zhou*[2,3], *Eric van Dijk*[4], *Marijke C. Leliveld*[5] and *Xiaolin Zhou*[2,3,6]

[1] Key Laboratory of Child Development and Learning Science (Ministry of Education), Southeast University, Nanjing, China
[2] Department of Psychology, Peking University, Beijing, China
[3] Center for Brain and Cognitive Sciences, Peking University, Beijing, China
[4] Department of Social and Organizational Psychology, Leiden University, Leiden, Netherlands
[5] Department of Marketing, University of Groningen, Groningen, Netherlands
[6] Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing, China

Edited by:
Christian Bellebaum, Ruhr University Bochum, Germany

Reviewed by:
Bernd Weber, Rheinische-Friedrich-Wilhelms Universität, Germany
Gregor Kohls, The Children's Hospital of Philadelphia, USA

***Correspondence:**
Xiaolin Zhou, Department of Psychology, Peking University, Beijing 100871, China.
e-mail: xz104@pku.edu.cn

Previous studies have shown that social comparison influences individual's fairness consideration and other-regarding behavior. However, it is not clear how social comparison affects the brain activity in evaluating fairness during asset distribution. In this study, participants, acting as recipients in the ultimatum game, were informed not only of offers to themselves but also of the average amount of offers in other allocator–recipient dyads. Behavioral results showed that the participants were more likely to reject division schemes when they were offered less than the other recipients, especially when the offers were highly unequal. Event-related brain potentials recorded from the participants showed that highly unequal offers elicited more negative-going medial frontal negativity than moderately unequal offers in an early time window (270–360 ms) and this effect was not significantly modulated by social comparison. In a later time window (450–650 ms), however, the late positive potential (LPP) was more positive for moderately unequal offers than for highly unequal offers when the other recipients were offered less than the participants, whereas this distinction disappeared when the other recipients were offered the same as or more than the participants. These findings suggest that the brain activity in evaluating fairness in asset division entails both an earlier (semi-) automatic process in which the brain responds to fairness at an abstract level and a later appraisal process in which factors related to social comparison and fairness norms come into play.

Keywords: social comparison, fairness consideration, ultimatum game, ERP, MFN, LPP

INTRODUCTION

Fairness is important in interpersonal interaction and for social stability. A large number of studies, employing different paradigms, show that people demand fairness in wealth allocation and are willing to sacrifice economic interests to punish unfair behavior (Fehr and Gächter, 2002; Camerer, 2003). One way to investigate fairness consideration in asset division is to let individuals play economic exchange games, like the ultimatum game (UG; Güth et al., 1982), and to examine these individuals' behavioral responses and/or brain activities. In the standard UG, two players have to divide a certain amount of money between them. One player is the allocator and proposes a division of the money; the other is the recipient and can either accept or reject the division scheme. If the recipient accepts, the asset is divided as proposed. If the recipient rejects, both players end up empty-handed. Ample evidence shows that allocators often offer an equal split, and that recipients are unwilling to accept offers that leave them with approximately 20% of the pie or less (Camerer and Thaler, 1995). Studies manipulating the size of the bargaining property and the population of players obtain essentially the same pattern of effects (Hoffman et al., 1996; Henrich et al., 2006).

As the UG is a typical dyadic bargaining situation, the recipient automatically compares the amount offered to him with the amount the allocator would have, and this comparison helps the recipient to judge whether the division scheme is fair (Handgraaf et al., 2003). If the amount offered to the recipient compares unfavorably to the amount left to the allocator, negative feelings are elicited, and drive the recipient to reject the offer (Sanfey et al., 2003). In daily life, however, individuals may focus not only on the outcomes of those who are involved in the current negotiation, but also on the outcomes of people who are in similar roles but in different negotiations (Loewenstein et al., 1989). For example, in salary negotiations, prospective employees typically do not compare their wages with those of their employers, but rather with wages of similarly situated employees (Babcock et al., 1996). Bohnet and Zeckhauser (2004) termed the comparison in UG between recipients in different allocator–recipient dyads as *social comparison*, and found it to affect both the allocator's and the recipient's bargaining behavior. In particular, when recipients were informed of the average offer of allocators in other allocator–recipient dyads, the allocators were more likely to propose higher offers and the recipients were more likely to reject unequal offers. They suggested that social comparison could highlight and facilitate attention to fairness norms and affect fairness consideration and other-regarding behavior in strategic situations.

Note that social comparison can be in different directions (Festinger, 1954): an upward comparison in which individuals compare themselves with peers in a better standing; a downward comparison in which individuals compare themselves with peers in an inferior situation; and a lateral comparison in which individuals compare themselves with peers in similar standings. Downward and upward comparisons may lead to different emotional responses, including schadenfreude and envy (Takahashi et al., 2009). They may also affect, in different ways, how individuals respond to unequal division of asset.

The main purpose of this study was therefore to investigate how upward and downward social comparison modulates the recipient's fairness consideration in asset division and how the brain responds to such modulations. Previous studies have demonstrated that brain regions such as ventral striatum and/or ventromedial prefrontal cortex are involved in the social comparison process (Fliessbach et al., 2007; Takahashi et al., 2009; Dvash et al., 2010; Tricomi et al., 2010), and these brain activities are modulated by individual differences, such as social value orientation (Haruno and Frith, 2010). What is lacking is the detailed knowledge about the temporal characteristics of neural processes involved in social comparison and its modulation on fairness consideration (see Qiu et al., 2010; Boksem et al., 2011). Here we developed a variant of UG in which the participant, acting as a recipient in asset division, was informed not only about the amount of money (out of 10 Chinese yuan) offered to him by the allocator in his own allocator–recipient dyad but also the average amount offered to recipients in other allocator–recipient dyads. While the offer to the participants could be moderately unequal (4 out of 10 yuan) or highly unequal (2 out of 10 yuan), a downward comparison was made possible by the presented average amount of offers in other dyads being 3 or 1 yuan, respectively; similarly, an upward comparison was made possible by the average amount being 5 or 3 yuan, respectively. From a rational perspective, the potential social comparison should not affect the participant's decision to accept or reject the offer as interests of the participant and of other recipients were independent from each other. However, it has been demonstrated that fairness consideration is strongly context-dependent and is constrained by various social or situational factors (Handgraaf et al., 2003; Güroglu et al., 2010). It was likely that the upward (and perhaps the downward) comparison would affect the participant's decision to accept or reject the offers, especially when the offer was highly unequal.

Electrophysiologically, we focused on the medial frontal negativity (MFN), an Event-related brain potential (ERP) component that has been implicated in the evaluation of fairness in asset distribution. The MFN is a negative deflection peaking between 200 and 350 ms at frontocentral recording sites. It has been found to be sensitive to violation of social expectancy or social norms (Polezzi et al., 2008; Boksem and De Cremer, 2010; Hewig et al., 2011; Wu et al., 2011; Van der Veen and Sahibdin, 2011). Unequal offers, i.e., offers deviating from the equal division of asset, elicit more negative-going MFN than equal offers in economic exchange games. The MFN is also more pronounced for lower offers than for higher offers and this effect is especially true for participants with high concerns for fairness (Boksem and De Cremer, 2010). For the present study, we would predict that highly unequal offers could elicit more negative MFN responses than moderately unequal offers, reflecting a general violation of social expectancy. Moreover, we predicted that social comparison could modulate the MFN effect for different offers. Boksem et al. (2011) found that the MFN effect for monetary gains and losses associated with outcomes in a time-estimation task is more pronounced when an individual's own reward is worse than that for others. We therefore hypothesized that highly unequal offers (2 out of 10 yuan) would elicit stronger MFN effect when the participants were offered less than the average amount of offers (3 yuan) to the recipients in other allocator–recipient dyads, as upward comparison might strengthen the negative motivational/affective significance of the highly unequal offers.

Another ERP component, the P300, which is the most positive peak in the period of 200–600 ms, has also been found to be related to various aspects of outcome evaluation or reward processing. Some studies found that the P300 is sensitive to the magnitude of reward, with a more positive response to a larger than to a smaller reward (Yeung and Sanfey, 2004; Sato et al., 2005). Other studies suggested that the P300 is also sensitive to reward valence, with a more positive amplitude for positive than for negative reward (Hajcak et al., 2005, 2007; Yeung et al., 2005; Wu and Zhou, 2009; Leng and Zhou, 2010). In a study on asset division, Wu et al. (2011) found that the P300 is more positive to equal offers than to unequal offers. Thus one might predict a similar pattern for the P300 in this study, although it was not clear how social comparison might modulate the pattern of the P300 effect. On the other hand, Qiu et al. (2010) asked participants to perform a number estimation task and to receive feedback on their own as well as others' monetary reward associated with performance. They obtained an effect on sustained late positivity potential (LPP) rather than the P300, for lateral, upward, and downward comparisons. It has been suggested the LPP may have functional significance similar to that of the P300 (Ito et al., 1998). It was not clear whether we would observe an LPP or P300 effect for offer type and/or for social comparison.

MATERIALS AND METHODS

PARTICIPANTS

Twenty-six undergraduate and graduate students (19 females; mean age 21.92 years, SD = 2.00) participated in the experiment. Seven students, who were strangers to the participants, were recruited as confederates. The purpose of using seven confederates was to reduce the reputation building effect in the repeated-trial game and to make the experimental setup more realistic since the participant would play against different allocators in rounds of the game.

All the participants were right-handed and had normal or corrected-to-normal vision. They had no history of neurological or psychiatric disorders. Informed consent was obtained from each participant before the test. The experiment was performed in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of the Department of Psychology, Peking University. Each participant was paid 45 Chinese yuan (about $ 6.9) as basic payment and was informed that additional monetary reward would be paid according to their performance in the task,

although in the end all the participants were paid extra 5 yuan on top of the basic payment.

DESIGN AND PROCEDURES

The experiment had a 2×3 within-participant factorial design, with the first factor referring to offer type (moderately unequal vs. highly unequal) and the second factor referring to social comparison (downward vs. lateral vs. upward). A highly unequal offer was 2 out of 10 yuan and a moderately unequal offer was 4 out of 10 yuan. For upward or downward comparison condition, the average amounts of offers to recipients in other allocator–recipient dyads were 1 yuan more or less than the offers to the participants, whereas for lateral comparison, the average amounts were equal to the offers to the participants.

When a participant came to the laboratory, he/she and the seven confederates were told that they would sit in separate rooms to finish a task together through the computer network. By assigning the participant and the confederates pre-determined cards, they were ostensively led to separate cubicles to play different roles in the game. The participant was then told that he/she as well as another three randomly selected participants would play as recipients in UG and the other four would play as allocators. He/she was also informed about the rules of the experiment. That is, at the beginning of each round, the computer would randomly pair each recipient with one allocator, and the allocator would then make an offer on how to divide 10 yuan. Offers in the four different dyads were independently and simultaneously made by the allocators. The participant was presented with not only the amount his/her paired allocator proposed to offer but also the average amount of offers in the other three allocator–recipient dyads. The participant was asked to press a button with the index finger of his/her left or right hand, without elaborative thinking, to indicate whether he/she would accept or reject the offer. Note that the participant was reminded that his/her response to each offer would not be sent back to the allocator immediately and therefore would not affect the allocators' offers in the following rounds.

Each trial began with the presentation of a photo of the 10 yuan bill ($2.6^\circ \times 1.3^\circ$) for 500 ms against a black background (see **Figure 1**). The sentence "The computer is randomly pairing" in Chinese (white and Song font, size 32) was presented for another 500 ms, indicating to the participant that four different dyads were being formed randomly. Then the sentence "Please wait for the offer" in Chinese (white and Song font, size 32) was presented for either 500, 750, 1000, 1250 ms, implying that the allocators were considering how to distribute the assets. After the presentation of a blank screen for a period of either 400, 500, 600, or 700 ms, the amount offered to the EEG participant as well as the average amount offered in other dyads (i.e., the division scheme) were revealed in two lines of words (e.g., "you 2, average 2," white and Song font, size 32) at the center of screen for 1000 ms. The screen turned blank again for 400 ms, followed by the presentation of two options, "accept" and "reject" (in words), on the left and right side of the screen respectively, with the positions of the two options counterbalanced over participants. The participant was asked to make the "accept" or "reject" decision as quickly as possible and his/her choice was highlighted by thickening the white outlines of the option. The next trial began 1000 ms after the button press.

The participant was seated comfortably about 1.5 m in front of a computer screen in a dimly lit room. The experiment was administered on a computer with a Del 22-in. CRT display, using Presentation software (Neurobehavioral System Inc.) to control the presentation and timing of the stimuli. Without the participant's knowledge, all the offers were pre-determined by a computer program. Each of the six experimental conditions had 40 trials. In addition, another 120 trials, with the average offers (up to one decimal place) computed according to different possible offers, were used as fillers. The 360 trials were randomly mixed and were divided in equal numbers into 3 blocks. A practice block of 9 trials was administered before the formal test to familiarize the participants with the task. Participants were debriefed, paid, and thanked at the end of the experiment.

EEG RECORDING AND ANALYSIS

EEGs were recorded from 64 scalp sites using tin electrodes mounted in an elastic cap (Brain Products, Munich, Germany) according to the international 10–20 system. The vertical electrooculogram (VEOGs) was recorded supra-orbitally from the right eye. The horizontal EOG (HEOG) was recorded from electrodes placed at the outer canthus of the left eye. All EEGs and EOGs were referenced online to an external electrode which was placed on the tip of nose and were re-referenced offline to the mean of the left and right mastoids. Electrode impedance was kept below 10 kΩ for EOG channels and below 5 kΩ for all other electrodes. The bio-signals were amplified with a bandpass from 0.016 to 100 Hz and digitized online with a sampling frequency of 500 Hz.

Separate EEG epochs of 1000 ms (with a 200-ms pre-stimulus baseline) were extracted offline, time-locked to the onset of each division scheme. Ocular artifacts were corrected with an eye-movement correction algorithm that employs a regression analysis in combination with artifact averaging (Semlitsch et al., 1986). Epochs were baseline-corrected by subtracting from each sample the average activity of that channel during the baseline period. All the trials in which EEG voltages exceeded a threshold of $\pm 80\,\mu V$ during recording were excluded from further analysis. For highly unequal offers, on average 36.36 (SD = 4.51), 35.81 (SD = 5.21), and 35.91 (SD = 4.96) trials after artifact rejection were entered into statistical analysis for the downward, lateral, and upward social comparison conditions, respectively. For moderately unequal offers, on average 36.91 (SD = 4.57), 36.09 (SD = 5.07), and 35.82 (SD = 4.55) trials were left for the three conditions, respectively. The number of trials did not differ between conditions after artifact rejection. The EEG data were low-pass filtered below 30 Hz.

For the MFN, we focused on 10 frontal electrodes, F3, F1, Fz, F2, F4, FC3, FC1, FCz, FC2, and FC4. For the LPP, we focused on these same frontal electrodes as well as 10 posterior electrodes, CP3, CP1, CPz, CP2, CP4, P3, P1, Pz, P2, and P4. We concentrated on these electrodes because the MFN effect and the late positive potentials observed were the largest on these electrodes. The mean amplitudes in the time window of 270–360 ms (for the MFN) in the frontal region and the mean amplitudes in the time window of 450–650 ms (for the LPP) in both the frontal and posterior regions were analyzed. These time windows were selected according to the classical definitions for the MFN and the LPP and

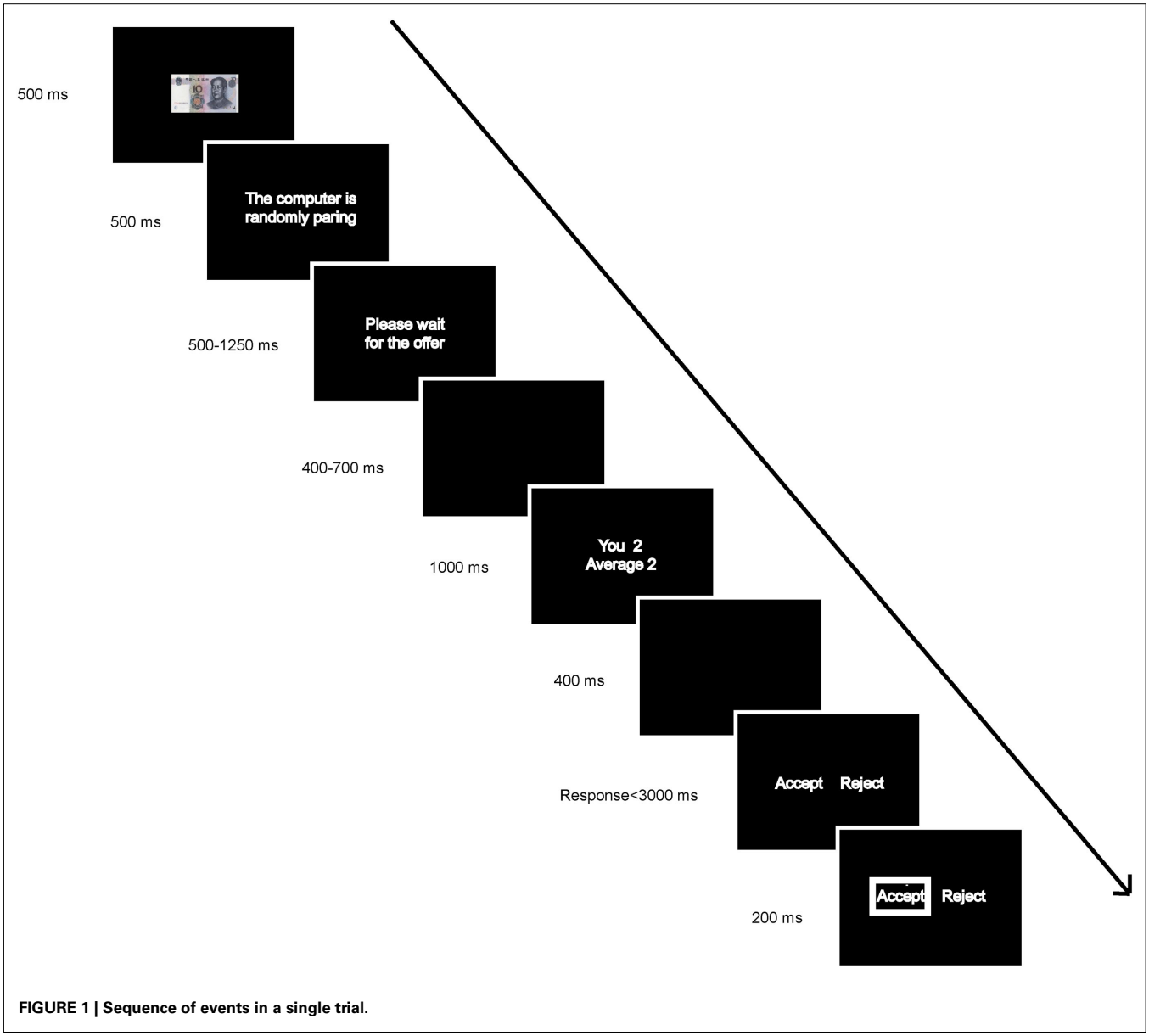


FIGURE 1 | Sequence of events in a single trial.

according to visual inspection of waveforms. Analyses of variance (ANOVAs) were conducted with three within-participant factors: offer type (highly unequal vs. moderately unequal), social comparison (upward vs. lateral vs. downward), and electrode. If the interaction between offer type and social comparison reached statistical significance, further F tests were conducted to test the simple effects, with electrode as a factor of no-interest. The Greenhouse–Geisser correction for violation of the assumption of sphericity was applied where appropriate. The Bonferroni correction was used for multiple comparisons.

RESULTS

Among the 26 EEG participants, one participant accepted all the offers and three participants displayed excessive artifacts in EEG recording. These participants were excluded from data analysis, leaving 22 participants for the following analysis.

BEHAVIORAL RESULTS

Rejection rates for different division schemes are presented in **Figure 2**. A 2 (offer type: highly unequal vs. moderately unequal) × 3 (social comparison: downward vs. lateral vs. upward) repeated-measures ANOVA revealed a significant main effect of offer type, $F(1,21) = 50.24$, $p < 0.001$, $\eta^2 = 0.71$, indicating that rejection rate for highly unequal offers (mean ± SD, 0.54 ± 0.30) was higher than that for moderately unequal offers (0.07 ± 0.11). The main effect of social comparison was significant, $F(2,42) = 18.07$, $\varepsilon = 0.69$, $p < 0.001$, $\eta^2 = 0.46$, suggesting that the rejection rate was higher for upward comparison (0.48 ± 0.25) than for either lateral (0.25 ± 0.20) or downward (0.20 ± 0.19) comparison, as confirmed by *post hoc* tests, $ps < 0.001$. Rejection rates for lateral and downward comparison did not differ from each other, $p > 0.1$. Importantly, these main effects were qualified by a significant interaction between offer

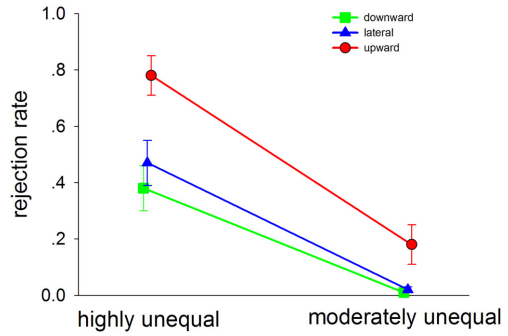


FIGURE 2 | The rejection rate depicted as a function of offer type and social comparison. Error bars represent standard errors of the means.

type and social comparison, $F(2,42) = 5.57$, $\varepsilon = 0.90$, $p = 0.01$, $\eta^2 = 0.21$. Further tests showed that for highly unequal offers, there was a main effect of social comparison, $F(2,42) = 18.20$, $p < 0.001$, $\eta^2 = 0.46$, with the rejection rate being higher for upward comparison (0.78 ± 0.31) than for either lateral (0.47 ± 0.39) or downward comparison (0.38 ± 0.37), $ps < 0.001$; for moderately unequal offers, the main effect of social comparison was also significant, $F(2,42) = 6.71$, $p < 0.05$, $\eta^2 = 0.24$, with the rejection rate being higher for upward (0.18 ± 0.31) than for downward comparison (0.01 ± 0.04), $p < 0.01$. Thus, the rejection rate was enhanced for upward comparison, and this was especially the case when the offers to the participants were highly unequal (**Figure 2**).

THE MFN

For the mean amplitudes in the 270- to 360-ms time window (**Figures 3A,B**), ANOVA revealed a significant main effect of offer type, $F(1,21) = 25.28$, $p < 0.001$, $\eta^2 = 0.55$, indicating that ERP responses were more negative going for highly unequal offers (mean ± SD, $0.35 \pm 3.06\,\mu V$) than for moderately unequal offers ($1.25 \pm 3.16\,\mu V$). However, there was no significant main effect of social comparison, $F(2,42) = 1.74$, $p > 0.1$, nor interaction between offer type and social comparison, $F(2,42) = 1.48$, $p > 0.1$, indicating that social comparison did not affect the manifestation of the MFN.

THE LATE POSITIVE POTENTIAL

At the frontal region, ANOVA revealed no significant main effect of offer type, $F(1,21) = 1.99$, $p > 0.1$, but a significant main effect of social comparison for the mean amplitudes in the 450- to 650-ms time window, $F(2,44) = 6.72$, $\varepsilon = 0.98$, $p < 0.01$, $\eta^2 = 0.24$, suggesting that the LPP was less positive for downward ($1.53 \pm 3.24\,\mu V$) and upward comparison ($1.46 \pm 3.44\,\mu V$) than for lateral comparison ($2.56 \pm 3.19\,\mu V$), $p < 0.05$ and $p < 0.01$, respectively. Moreover, this main effect was qualified by a significant interaction between offer type and social comparison, $F(2,42) = 3.36$, $\varepsilon = 0.95$, $p < 0.05$, $\eta^2 = 0.14$. Further tests showed that moderately unequal offers ($2.17 \pm 3.51\,\mu V$) elicited more positive-going responses than highly unequal offers ($0.90 \pm 3.18\,\mu V$) in downward comparison (**Figure 3C**), $F(1,21) = 11.93$, $p < 0.01$, $\eta^2 = 0.36$, whereas this contrast did not produce significant effects for either lateral or upward comparison, both $F(1,21) < 1$. On the other hand, for moderately unequal offers, social comparison did not affect LPP responses, $F(2,42) = 2.52$, $\varepsilon = 0.99$, $p = 0.09$; for highly unequal offers, social comparison did have a significant effect, $F(2,42) = 9.63$, $\varepsilon = 0.96$, $p < 0.001$, $\eta^2 = 0.31$, with LPP for downward ($0.90 \pm 3.18\,\mu V$) and upward comparison ($1.53 \pm 3.53\,\mu V$) being less positive-going than for lateral comparison ($2.61 \pm 3.12\,\mu V$), $p < 0.01$ and $p < 0.05$, respectively.

At the posterior region, ANOVA revealed only a significant main effect of social comparison, $F(2,42) = 5.33$, $\varepsilon = 0.92$, $p = 0.01$, $\eta^2 = 0.20$, indicating that the mean amplitudes were less positive for upward comparison ($5.63 \pm 3.14\,\mu V$) than for lateral comparison ($6.76 \pm 2.96\,\mu V$), $p < 0.01$. The mean amplitudes for lateral comparison were intermediate ($6.12 \pm 2.86\,\mu V$) and did not differ significantly from other conditions, $ps > 0.1$. Neither the main effect of offer type nor the interaction between offer type and social comparison was significant, $F(1,21) = 1.49$, $p > 0.1$, and $F(2,42) = 1.39$, $p > 0.1$, respectively.

DISCUSSION

This study demonstrated that social comparison influences recipients' behavioral reactions as well as their brain responses to unequal asset allocation schemes in the UG. Participants were more likely to reject division schemes when the recipients in other dyads were offered more than themselves (i.e., upward comparison), particularly when the offers were highly unequal. Electrophysiologically, highly unequal offers elicited more negative-going ERP responses than moderately unequal offers in an earlier MFN time window (270–360 ms), and this effect was not modulated by social comparison. In a later time window (450–650 ms), the late positive potential (LPP) was more positive for moderately unequal offers than for highly unequal offers when the other recipients were offered less than the participants were (i.e., downward comparison). These findings revealed the temporal characteristics of neural activity in social comparison and fairness consideration, complimenting previous fMRI studies that localized brain regions involved in social comparison (Dvash et al., 2010; Haruno and Frith, 2010; Tricomi et al., 2010) and fairness consideration (Sanfey et al., 2003; Tabibnia et al., 2008; Güroglu et al., 2010).

Previous studies suggest that upward comparison, i.e., being worse off than others, is motivationally salient and threatens self-esteem, causing individuals to feel inferior to the others (Wood, 1996). Such negative emotions elicited by upward comparison might drive the recipients to reject more often the division schemes, whether the offers are highly or moderately unequal and even though such costly punishment of the allocators might lead them and the allocators both empty-handed. Nevertheless, this effect of social comparison was more pronounced when the offers were highly unequal. Upward comparison deepens the experienced negative feeling caused by the unequal offers deviating from the equity rule in asset distribution, leading to more rejections to the division schemes. This finding is consistent with Bohnet and Zeckhauser (2004) which demonstrated that social comparison facilitates recipients' attention to the fairness norm.

The finding of a MFN effect, with more negative-going responses to highly unequal offers than to moderately unequal

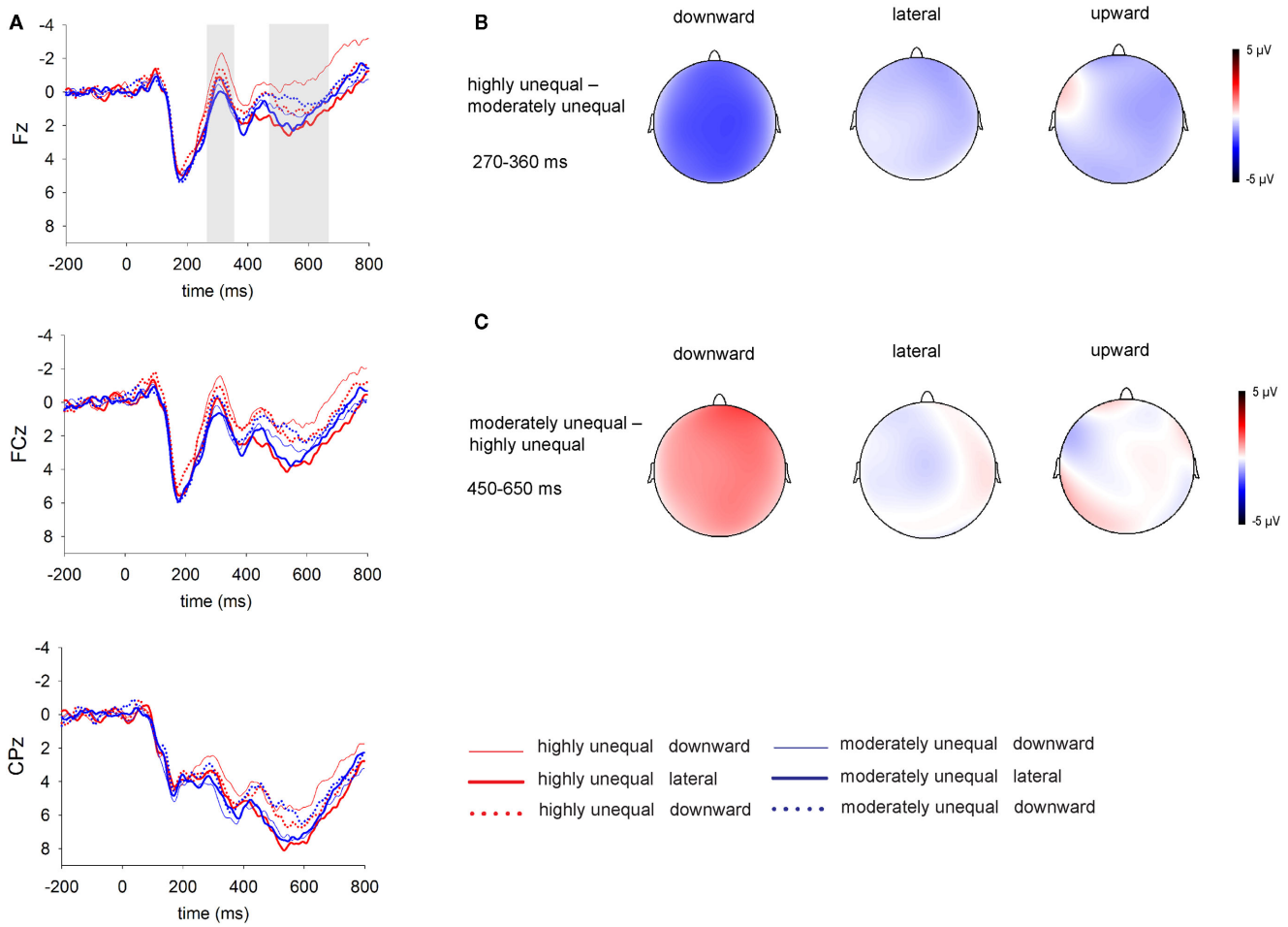


FIGURE 3 | (A) ERP responses time-locked to the onset of different offers at the midline Fz, FCz, and CPz. The shaded 270–360 ms time window was for the calculation of the mean amplitudes of the MFN. The shaded 450–650 ms time window was for the calculation of the mean amplitudes of the LPP; **(B)** topographic maps for the MFN effect; **(C)** topographic maps for the LPP effect.

offers, replicated previous studies in which the MFN effect increased with unfairness in economic games (Polezzi et al., 2008; Boksem and De Cremer, 2010; Hewig et al., 2011; Wu et al., 2011; Van der Veen and Sahibdin, 2011). This effect may reflect the detection of social expectancy violation as egalitarian distribution of assets is an expected social norm (Messick and Sentis, 1983; Fehr and Gächter, 2002; Fehr and Fischbacher, 2004). During evolution, the human brain may have developed specific mechanisms to detect ongoing deviations from social norms (Montague and Lohrenz, 2007). These mechanisms might share the same neural correlates as those engaged in predicting errors during non-social reinforcement learning (Harris and Fiske, 2010). The MFN can therefore reflect not only the encoding of prediction errors for monetary reward or performance feedback but also violations of expectancy toward social norms.

A perhaps surprising finding in this study was that social comparison had no obvious effect on the MFN responses to division schemes. This absence of a social comparison effect appears to be at odds with Boksem et al. (2011) and Qiu et al. (2010) in which social comparison modulated the MFN or MFN-like responses in outcome evaluation. It is possible that the discrepancy between these findings is due to different paradigms employed in the studies. In both Boksem et al. (2011) and Qiu et al. (2010), the participants performed a gambling task in which one's own outcome as well as the other's were presented simultaneously and the outcome for the participant was deterministic. In a recent study, Wu et al. (2011) also found that the social distance between the allocator and the recipient, i.e., being friends vs. strangers, could modulate MFN responses to equal and unequal offers in the dictator game (DG). DG is similar to UG except that the recipient owns no right but has to accept any offer from the allocator. This finding, together with Boksem et al. (2011) and Qiu et al. (2010), suggests that the social context can affect the MFN responses when reward is deterministic. However, in the UG paradigm adopted here, as the participants can decide to either accept or reject the offers, the outcome is negotiable. The (un)certainty of the outcome may affect the extent the participants process the affective/motivational significance of the outcome. The system may adopt a "wait-and-see" strategy and

conduct deeper assessment of offers only at a later stage involving more top-down processes (Cunningham et al., 2003; Leng and Zhou, 2010; Ma et al., 2011). That is, fairness consideration in UG can be differentiated into two stages: an earlier, semi-automatic stage in which the fairness of offers are considered at an abstract level with reference to long-established social norms but without much reference to personal interests; and a later, cognitive appraisal stage in which social factors comes into play (Moore and Loewenstein, 2004; Wu and Zhou, 2009; Leng and Zhou, 2010).

Recent ERP studies employing economic games have indicated that the P300 is sensitive to different offers, with its magnitude less positive to unequal offers (Wu et al., 2011). In the present study, we found the late positivity potential (LPP), rather than the P300, was modulated by social comparison. Although the LPP and the P300 may differ in temporal dynamics and scalp distribution, a number of studies indicated that they share similar functions in social evaluation and attitude categorization (see Hajcak et al., 2010 for a review). The P300 is generally believed to be related to processes of attentional allocation (Gray et al., 2004; Linden, 2005) and/or to high-level motivational/affective evaluation (Yeung and Sanfey, 2004; Nieuwenhuis et al., 2005). Similarly, the LPP has been implicated in the process of social evaluation, with enhanced positive amplitudes reflecting increased motivated attention (van Hooff et al., 2010). For instance, the LPP has been found to be largest for stimuli that are motivationally relevant, receive the highest reports of affective experience, and prompt the largest levels of autonomic arousal (Schupp et al., 2004; Briggs and Martin, 2009).

Moderately unequal offers evoked more positive LPP than highly unequal offers when the recipients in the other dyads were offered less than the participants were. This finding is consistent with the P300/LPP results in the previous ERP studies using economic games (Wu et al., 2011). It may reflect differential distribution of attentional resources to the two types of offers that had different affective/motivational significance. However, the offer type effect on LPP appeared mostly in the downward comparison conditions, not in the lateral or upward comparison conditions (**Figure 3C**). It is possible that in the present setup the LPP is more sensitive to social comparison rather than to offer type. The participants cared more and devoted more attentional resources to the comparison between the offers to themselves and the offers to their peers than to the comparison between the offers to themselves and the amount left to the proposers. This dominance seeking (Rustichini, 2008) captured attention and overshadowed the offer type effect on LPP in the lateral and upward comparison conditions. When the participants were in advantageous positions compared with their peers (e.g., in downward comparison), more attentional resources were left to the comparison between the offers to the participants and the amounts left to the proposers and then the offer type effect emerged.

Note that the pattern of the LPP effect outlined above was inconsistent with the pattern of effect in rejection rates (**Figure 2**), as the offer type effect on LPP was absent in both lateral and upward comparison conditions whereas the rejection rates for the lateral comparison conditions were more similar to the downward comparison conditions, rather than to the upward comparison conditions. The latter pattern of the rejection rates was also observed in another behavioral study in which more levels of offer type were included (unpublished data). The apparent contradiction between the ERP pattern and the behavioral data may be due to different processes underlying the electrophysiological and behavioral responses. The electrophysiological measurement was time-locked to the presentation of division schemes that are evaluated promptly, whereas the behavioral responses may involve more elaborated processes, including weighting pros and cons of rejection (i.e., strategic consideration).

CONCLUSION

By providing participants with highly or moderately unequal schemes of asset division in the UG and by informing them the average offers to their peers in other allocator–recipient dyads, we found that the participants' rejections of the unequal offers were modulated by the comparison between the offers to the participants and the offers to their peers. Participants were more likely to reject division schemes when they were offered less than their peers were, and this was especially the case when the offers to the participants were highly unequal. Electrophysiologically, highly unequal offers elicited more negative-going MFN responses than moderately unequal offers in an early time window (270–360 ms) and this effect was not obviously modulated by social comparison. In a later time window (450–650 ms), the LPP was more positive for moderately unequal offers than for highly unequal offers in downward comparison, whereas this effect was absent in lateral or upward comparison. These findings suggest that the brain responses to the fairness in asset division entails both an earlier (semi-) automatic process in which the brain responds to fairness at an abstract level and a later appraisal process in which factors related to social comparison and fairness norms come into play.

ACKNOWLEDGMENTS

This study was supported by National Basic Research Program (973 Program: 2010CB833904) and by grants from the National Natural Science Foundation of China (31170972, 71172216).

REFERENCES

Babcock, L., Wang, X., and Loewenstein, G. (1996). Choosing the wrong pond: social comparisons in negotiations that reflect a self-serving bias. *Q. J. Econ.* 111, 1–19.

Bohnet, I., and Zeckhauser, R. (2004). Social comparisons in ultimatum bargaining. *Scand. J. Econ.* 106, 495–510.

Boksem, M. A. S., and De Cremer, D. (2010). Fairness concerns predict medial frontal negativity amplitude in ultimatum bargaining. *Soc. Neurosci.* 5, 118–128.

Boksem, M. A. S., Kostermans, E., and De Cremer, D. (2011). Failing where others have succeeded: medial frontal negativity tracks failure in a social context. *Psychophysiology* 48, 973–979.

Briggs, K. E., and Martin, F. H. (2009). Affective picture processing and motivational relevance: arousal and valence effects on ERPs in an oddball task. *Int. J. Psychophysiol.* 72, 299–306.

Camerer, C., and Thaler, R. H. (1995). Anomalies: ultimatums, dictators and manners. *J. Econ. Perspect.* 9, 209–219.

Camerer, C. F. (2003). Behavioural studies of strategic thinking in games. *Trends. Cogn. Sci. (Regul. Ed.)* 7, 225–231.

Cunningham, W. A., Johnson, M. K., Gatenby, J. C., Gore, J. C., and Banaji, M. R. (2003). Neural components of social evaluation. *J. Pers. Soc. Psychol.* 85, 639–649.

Dvash, J., Gilam, G., Ben-Ze'Ev, A., Hendler, T., and Shamay-Tsoory,

S. G. (2010). The envious brain: the neural basis of social comparison. *Hum. Brain Mapp.* 31, 1741–1750.
Fehr, E., and Fischbacher, U. (2004). Third-party punishment and social norms. *Evol. Hum. Behav.* 25, 63–87.
Fehr, E., and Gächter, S. (2002). Altruistic punishment in humans. *Nature* 415, 137–140.
Festinger, L. (1954). A theory of social comparison processes. *Hum. Relat.* 7, 117–140.
Fliessbach, K., Weber, B., Trautner, P., Dohmen, T., Sunde, U., Elger, C. E., and Falk, A. (2007). Social comparison affects reward-related brain activity in the human ventral striatum. *Science* 318, 1305–1308.
Gray, H. M., Ambady, N., Lowenthal, W. T., and Deldin, P. (2004). P300 as an index of attention to self-relevant stimuli. *J. Exp. Soc. Psychol.* 40, 216–224.
Güroglu, B., van den Bos, W., Rombouts, S. A. R. B., and Crone, E. A. (2010). Unfair? It depends: neural correlates of fairness in social context. *Soc. Cogn. Affect. Neurosci.* 5, 414–423.
Güth, W., Schmittberger, R., and Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *J. Econ. Behav. Organ.* 3, 367–388.
Hajcak, G., Holroyd, C. B., Moser, J. S., and Simons, R. F. (2005). Brain potentials associated with expected and unexpected good and bad outcomes. *Psychophysiology* 42, 161–170.
Hajcak, G., MacNamara, A., and Olvet, D. M. (2010). Event-related potentials, emotion, and emotion regulation: an integrative review. *Dev. Neuropsychol.* 35, 129–155.
Hajcak, G., Moser, J. S., Holroyd, C. B., and Simons, R. F. (2007). It's worse than you thought: the feedback negativity and violations of reward prediction in gambling tasks. *Psychophysiology* 44, 905–912.
Handgraaf, M. J. J., van Dijk, E., Wilke, H. A. M., and Vermunt, R. C. (2003). The salience of a recipient's alternatives: inter-and intrapersonal comparison in ultimatum games. *Organ. Behav. Hum. Decis. Process.* 90, 165–177.
Harris, L. T., and Fiske, S. T. (2010). Neural regions that underlie reinforcement learning are also active for social expectancy violations. *Soc. Neurosci.* 5, 76–91.
Haruno, M., and Frith, C. D. (2010). Activity in the amygdala elicited by unfair divisions predicts social value orientation. *Nat. Neurosci.* 13, 160–161.
Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D., and Ziker, J. (2006). Costly punishment across human societies. *Science* 312, 1767–1770.
Hewig, J., Kretschmer, N., Trippe, R. H., Hecht, H., Coles, M., Holroyd, C. B., and Miltner, W. H. (2011). Why humans deviate from rational choice. *Psychophysiology* 48, 507–514.
Hoffman, E., McCabe, K. A., and Smith, V. L. (1996). On expectations and the monetary stakes in ultimatum games. *Int. J. Game Theory* 25, 289–301.
Ito, T. A., Larsen, J. T., Smith, N. K., and Cacioppo, J. T. (1998). Negative information weighs more heavily on the brain: the negativity bias in evaluative categorizations. *J. Pers. Soc. Psychol.* 75, 887–900.
Leng, Y., and Zhou, X. (2010). Modulation of the brain activity in outcome evaluation by interpersonal relationship: an ERP study. *Neuropsychologia* 48, 448–455.
Linden, D. E. J. (2005). The P300: where in the brain is it produced and what does it tell us? *Neuroscientist* 11, 563–576.
Loewenstein, G. F., Thompson, L., and Bazerman, M. H. (1989). Social utility and decision making in interpersonal contexts. *J. Pers. Soc. Psychol.* 57, 426–441.
Ma, Q., Shen, Q., Xu, Q., Li, D., Shu, L., and Weber, B. (2011). Empathic responses to others' gains and losses: an electrophysiological investigation. *Neuroimage* 54, 2472–2480.
Messick, D. M., and Sentis, K. (1983). "Fairness, preference, and fairness biases," in *Equity Theory: Psychological and Sociological Perspectives*, eds D. M. Messick and K. S. Cook (New York: Praeger), 61–94.
Montague, P. R., and Lohrenz, T. (2007). To detect and correct: norm violations and their enforcement. *Neuron* 56, 14–18.
Moore, D. A., and Loewenstein, G. (2004). Self-interest, automaticity, and the psychology of conflict of interest. *Soc. Justice Res.* 17, 189–202.
Nieuwenhuis, S., Aston-Jones, G., and Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychol. Bull.* 131, 510–532.
Polezzi, D., Daum, I., Rubaltelli, E., Lotto, L., Civai, C., Sartori, G., and Rumiati, R. (2008). Mentalizing in economic decision-making. *Behav. Brain Res.* 190, 218–223.
Qiu, J., Yu, C., Li, H., Jou, J., Tu, S., Wang, T., Wei, D., and Zhang, Q. (2010). The impact of social comparison on the neural substrates of reward processing: an event-related potential study. *Neuroimage* 49, 956–962.
Rustichini, A. (2008). Dominance and competition. *J. Eur. Econ. Assoc.* 6, 647–656.
Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., and Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science* 300, 1755–1758.
Sato, A., Yasuda, A., Ohira, H., Miyawaki, K., Nishikawa, M., Kumano, H., and Kuboki, T. (2005). Effects of value and reward magnitude on feedback negativity and P300. *Neuroreport* 16, 407–411.
Schupp, H., Cuthbert, B., Bradley, M., Hillman, C., Hamm, A., and Lang, P. (2004). Brain processes in emotional perception: motivated attention. *Cogn. Emot.* 18, 593–611.
Semlitsch, H. V., Anderer, P., Schuster, P., and Presslich, O. (1986). A solution for reliable and valid reduction of ocular artifacts, applied to the P300 ERP. *Psychophysiology* 23, 695–703.
Tabibnia, G., Satpute, A. B., and Lieberman, M. D. (2008). The sunny side of fairness. *Psychol. Sci.* 19, 339–347.
Takahashi, H., Kato, M., Matsuura, M., Mobbs, D., Suhara, T., and Okubo, Y. (2009). When your gain is my pain and your pain is my gain: neural correlates of envy and schadenfreude. *Science* 323, 937–939.
Tricomi, E., Rangel, A., Camerer, C. F., and O'Doherty, J. P. (2010). Neural evidence for inequality-averse social preferences. *Nature* 463, 1089–1091.
Van der Veen, F. M., and Sahibdin, P. P. (2011). Dissociation between medial frontal negativity and cardiac responses in the ultimatum game: effects of offer size and fairness. *Cogn. Affect. Behav. Neurosci.* doi: 10.3758/s13415-011-0050-1. [Epub ahead of print].
van Hooff, J. C., Crawford, H., and Van Vugt, M. (2010). The wandering mind of men: ERP evidence for gender differences in attention bias towards attractive opposite sex faces. *Soc. Cogn. Affect. Neurosci.* 6, 477–485.
Wood, J. V. (1996). What is social comparison and how should we study it? *Pers. Soc. Psychol. Bull.* 22, 520–537.
Wu, Y., Leliveld, M. C., and Zhou, X. (2011). Social distance modulates recipient's fairness consideration in the dictator game: an ERP study. *Biol. Psychol.* 88, 253–262.
Wu, Y., and Zhou, X. (2009). The P300 and reward valence, magnitude, and expectancy in outcome evaluation. *Brain Res.* 1286, 114–122.
Yeung, N., Holroyd, C. B., and Cohen, J. D. (2005). ERP correlates of feedback and reward processing in the presence and absence of response choice. *Cereb. Cortex* 15, 535–544.
Yeung, N., and Sanfey, A. G. (2004). Independent coding of reward magnitude and valence in the human brain. *J. Neurosci.* 24, 6258–6264.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 21 September 2011; paper pending published: 10 October 2011; accepted: 20 October 2011; published online: 09 November 2011.
Citation: Wu Y, Zhou Y, van Dijk E, Leliveld MC and Zhou X (2011) Social comparison affects brain responses to fairness in asset division: an ERP study with the ultimatum game. Front. Hum. Neurosci. ***5****:131. doi: 10.3389/fnhum.2011.00131*

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 22 February 2012
doi: 10.3389/fnhum.2012.00029

Brain responses in evaluating feedback stimuli with a social dimension

Yuan Zhang[1], Xiang Li[1], Xing Qian[2] and Xiaolin Zhou[1,2,3]*

[1] *Center for Brain and Cognitive Sciences and Department of Psychology, Peking University, Beijing, China*
[2] *Key Laboratory of Child Development and Learning Science (Ministry of Education), Southeast University, Nanjing, China*
[3] *Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing, China*

Edited by:
Christian Bellebaum, Ruhr University Bochum, Germany

Reviewed by:
Benjamin Eppinger, Max Planck Institute for Human Development, Germany
Luca Vizioli, University of Glasgow, UK

***Correspondence:**
Xiaolin Zhou, Department of Psychology, Peking University, Beijing 100871, China.
e-mail: xz104@pku.edu.cn

Previous studies on outcome evaluation and performance monitoring using gambling or simple cognitive tasks have identified two event-related potential (ERP) components that are particularly relevant to the neural responses to decision outcome. The feedback-related negativity (FRN), typically occurring 200–300 ms post-onset of feedback stimuli, encodes mainly the valence of outcome while the P300, which is the most positive peak between 200–600 ms, is related to various aspects of outcome evaluation. This study investigated the extent to which neural correlates of outcome evaluation involving perceptually complex feedback stimuli (i.e., female faces) are similar to those elicited by simple feedback. We asked participants to judge the attractiveness of blurred faces and then showed them unblurred faces as implicit feedback. The FRN effect can be identified in the ERP waveforms, albeit in a delayed 300–380 ms time window, with faces inconsistent with the initial judgment eliciting more negative-going responses than faces consistent with the judgment. However, the ERP waveforms did not show the typical pattern of P300 responses. With the principal component analysis (PCA), a clear pattern of P300 effects *were* revealed, with the P300 being more positive to faces consistent with the initial judgment than to faces inconsistent with the judgment, and more positive to attractive faces than to unattractive ones. The effect of feedback consistency did not interact with the effect of attractiveness in either the FRN or P300 component. These findings suggest that brain responses involved in processing complex feedback stimuli with a social dimension are generally similar to those involved in processing simple feedback stimuli in gambling or cognitive tasks, although appropriate means of data analysis are needed to reveal the typical ERP effects that may have been masked by sophisticated cognitive (and emotional) processes for complex stimuli.

Keywords: outcome evaluation, performance monitoring, facial attractiveness, ERP, FRN, P300, PCA

INTRODUCTION

Over the past decade, there has been increased interest in the neural basis of performance monitoring and outcome evaluation, which plays an important role in decision-making and learning from the environment. Electrophysiological studies on outcome evaluation and performance monitoring have consistently observed two event-related potential (ERP) components (Miltner et al., 1997; Gehring and Willoughby, 2002; Holroyd and Coles, 2002; Nieuwenhuis et al., 2004). The first component, the feedback-related negativity (FRN), is a negative deflection at frontocentral recording sites that typically reaches a maximum amplitude around 250 ms post-onset of the feedback stimulus. It is sensitive to the outcome valence, being more negative-going for negative feedback associated with unfavorable outcomes, such as incorrect responses (Miltner et al., 1997), monetary losses (Gehring and Willoughby, 2002), or violations of expectancy (Heldmann et al., 2008; Wu and Zhou, 2009), than for positive feedback. Another ERP component, the P300, is the most positive peak in the 200–600 ms period post-onset of feedback stimulus and typically increases in magnitude from frontal to parietal sites. The P300 is traditionally believed to reflect processes demanding attentional resources (Polich and Kok, 1995; Polich, 2007) and it has been found to be related to various aspects of outcome evaluation, including the magnitude of reward (Yeung and Sanfey, 2004; Sato et al., 2005), expectancy toward the outcome (Hajcak et al., 2005, 2007; Wu and Zhou, 2009), and the valence of the outcome (Hajcak et al., 2005, 2007; Wu and Zhou, 2009; Leng and Zhou, 2010).

Almost all the previous studies about outcome evaluation and performance monitoring employed relatively simple tasks (e.g., making a gambling choice or estimating time duration elapsed for a simple visual stimulus), with feedback stimuli that demand little perceptual or cognitive processing (e.g., numerals representing monetary reward or symbols indicating the correctness of estimation). However, in our daily life, we often encounter more complex feedback stimuli, which require sophisticated cognitive processing, particularly stimuli that are important in social interactions (e.g., faces). Moreover, we often make judgments or predictions based on limited information. The later outcome evaluation may involve an implicit comparison that checks the

previous judgment or expectancy stored in memory against the newly available information. For example, when we think we recognize a friend from distance we might want to approach and greet him. Only when we come closer do we see details of his face and realize that he is not the expected person. During the time between seeing the individual from afar and approaching him, we form expectations toward meeting a particular person based on partial information; the later implicit comparison between the new information and the initial expectation allows us to see whether our initial expectation or judgment was correct. It is not clear to what extent the pattern of brain responses to the new, complex feedback information is affected by the result of comparison (i.e., the consistency between the initial judgment or expectation and new information) and by the neurocognitive processes associated with the complex feedback stimulus itself.

The main purpose of this study is to investigate to what extent the neural correlates of the (implicit) outcome evaluation involving complex feedback stimuli (e.g., faces) are similar to those revealed for simple feedback in gambling or simple cognitive tasks. Answers to this question may open a new avenue for the use of the ERP technique to investigate the neural basis of outcome evaluation and performance monitoring in more complex, including social, situations. To mimic a situation in which limited information is available for decision-making and the feedback stimuli require sophisticated processing, we blurred photos of female faces and asked participants to judge whether a presented female face was attractive or unattractive. Photos of unblurred faces were presented as feedback and ERPs time-locked to the onset of the feedback were measured. Facial attractiveness is a key feature in social interactions, such as peer and mate choice (Etcoff, 1999; Thornhill and Gangestad, 1999; Johnston, 2006) and inferences about that individual's personality (Dion et al., 1972). If participants did make (implicit) comparisons between newly available information and previous judgments, then the feedback faces would carry information concerning the valence of feedback (consistent vs. inconsistent with the initial judgment), in addition to information concerning facial attractiveness. The processing of the social aspect of the feedback faces (i.e., attractiveness) becomes a necessary step in deciding whether the initial judgment or guess was right. A previous study found that the ventral occipital region including the fusiform face area (FFA) shows differential responses to attractive and unattractive faces even when the task is beauty-irrelevant (e.g., in an identity judgment task), indicating that facial attractiveness may be processed automatically (Chatterjee et al., 2009).

If the neural correlates of outcome evaluation for complex feedback stimuli with a social dimension are similar to those revealed for simple stimuli, inconsistent faces (those whose attractiveness is inconsistent with the initial judgment) should show an FRN effect relative to consistent faces. On the other hand, previous studies demonstrated that the P300 is sensitive to the reward valence in monetary gambling tasks (Hajcak et al., 2005, 2007; Wu and Zhou, 2009; Leng and Zhou, 2010), with more positive amplitudes for positive feedback than for negative feedback. If so, the P300 should be more positive to feedback faces consistent with the initial judgment than to faces inconsistent with the judgment. Moreover, previous studies also showed that ERP responses are more positive to attractive faces than to unattractive faces on late positive potential (LPP) or P300 (Johnston and Oliver-Rodriguez, 1997; Oliver-Rodríguez et al., 1999; Werheid et al., 2007; Schacht et al., 2008). We, therefore, predict a P300 or LPP effect for facial attractiveness in this study.

METHOD

PARTICIPANTS

Sixteen right-handed undergraduate students (eight females, 18–24 years, mean = 21) from Peking University participated in the experiment. All the participants had normal or corrected-to-normal vision, and had no history of neurological, psychiatric, or cognitive disorders. This study was conducted in accordance with the Declaration of Helsinki and was approved by the Academic Committee of the Department of Psychology, Peking University.

DESIGN AND MATERIALS

This experiment had a 2 by 2 factorial design, with the first experimental factor (consistency: consistent vs. inconsistent) referring to whether the attractiveness of feedback faces were consistent with the initial judgment and the second experimental factor (attractiveness: attractive vs. unattractive) referring to the facial attractiveness of the feedback faces. The experiment included pre-test, EEG test, and post-test. In the pre-test, 170 attractive and 170 unattractive gray-scale photos of female faces were selected from the photo pools of Peking University and the Institute of Psychology, Chinese Academy of Sciences, based on a consistent rating of above five or below three on a seven-point scale by 20 participants who did not participate in the EEG test. The experimental stimuli in the EEG test comprised of five attractive and five unattractive blurred female faces and the 170 attractive and 170 unattractive unblurred female faces used in the pre-test. Given that there were variable ratings as to whether particular male faces were attractive, we did not include male faces in this study. The attractiveness of unblurred female faces was confirmed in which the EEG participants were asked to rate these faces in the same way as the pre-test. Both the attractive and unattractive faces were unfamiliar to the EEG participants to exclude the influence of familiarity.

We adjusted all photos to be approximately 218 pixels in width (SD = 10.3 pixels) and 274 pixels in height (SD = 4.6 pixels), centered on a 260 pixels× 280 pixels black background. The brightness of all photos was adjusted to about 97.8 (SD = 6.0) as indicated in the histogram after undergoing the "auto laves" function on Adobe Photoshop CS2. Five attractive and five unattractive faces were blurred by Gaussian filters with $\sigma = 14$ pixels, maintaining all frequencies below four cycles per photo width.

In each trial in the EEG test, participants were instructed to guess the attractiveness of a blurred face, and to simply watch the subsequent unblurred face, which served as implicit feedback to the consistency of the initial judgment. The blurred faces were presented randomly while the 170 attractive and the 170 unattractive feedback faces were presented in pseudo-random orders. For each participant, each of the blurred faces was presented 34 times and each of the attractive or unattractive faces was presented only once, with the restriction that no more than three consecutive

feedback faces were from the same category. Different pseudo-random orders were created for different participants. Unknown to the participants, the blurred face in each trial was not the same one as the feedback face. The purpose of this manipulation was to exclude the potential influence of the blurred faces on the perceptual processing of the subsequent feedback faces as well as to make sure that about half of the trials would constitute "consistent" trials.

PROCEDURES

Participants were seated in a sound-attenuated, electrically shielded chamber approximately 1 m from a computer screen. At the start of each trial, at the center of the computer screen, a white fixation cross (0.6° × 0.6° in visual angle) was presented against a black background for 500 ms. Then a blurred black-and-white face photo was presented (6.3° × 4.6°), and remained on the screen until the participant's response (**Figure 1**). The participants' task was to make a binary attractiveness judgment as quickly as possible, by pressing a key on a joystick using their left or right index finger. Button assignment was counter-balanced between participants. After the response, a fixation cross was presented again for 800 ms. Then a unblurred face photo (6.3° × 4.6°), serving as feedback to the participants' initial judgment, was presented for 800 ms, and participants were asked to simply watch it and wait for the next trial. After the unblurred face, a fixation cross was presented for 700 ms and the screen turned black for 100 ms before the next trial began.

Before the EEG test, participants completed a practice block consisting of 10 trials to make sure they understood the task. The EEG test consisted of four blocks with 85 trials each. After the EEG test, participants completed a post-test, rating the attractiveness of each feedback face that had appeared in the EEG test. For faces which a participant's rating in the post-test was inconsistent with the predefined attractiveness, the corresponding trial with that face in the formal test was excluded from EEG data analysis.

EEG RECORDING

The EEG was recorded continuously from 62 scalp electrodes mounted on an elastic cap (NeuroScan Inc., Herndon, Virginia, USA) according to the extended 10–20 system with the addition of two mastoid electrodes. Signals were referenced online to the left mastoid and were re-referenced offline to the linked mastoids. Eye blinks and vertical eye movements were monitored with electrodes located above and below the left eye. The horizontal electro-oculogram was recorded from electrodes placed 1.5 cm lateral to the left and right external canthi. The electrode impedance was less than 5 kΩ. The EEG was amplified (bandpass 0.05–100 Hz) and digitized at 500 Hz.

DATA ANALYSIS

ERPs were computed for each participant over an epoch from 200 ms before to 800 ms after the onset of the feedback faces, with the 200 ms pre-stimulus EEG activity used for baseline correction. Ocular artifacts were corrected with an eye-movement correction

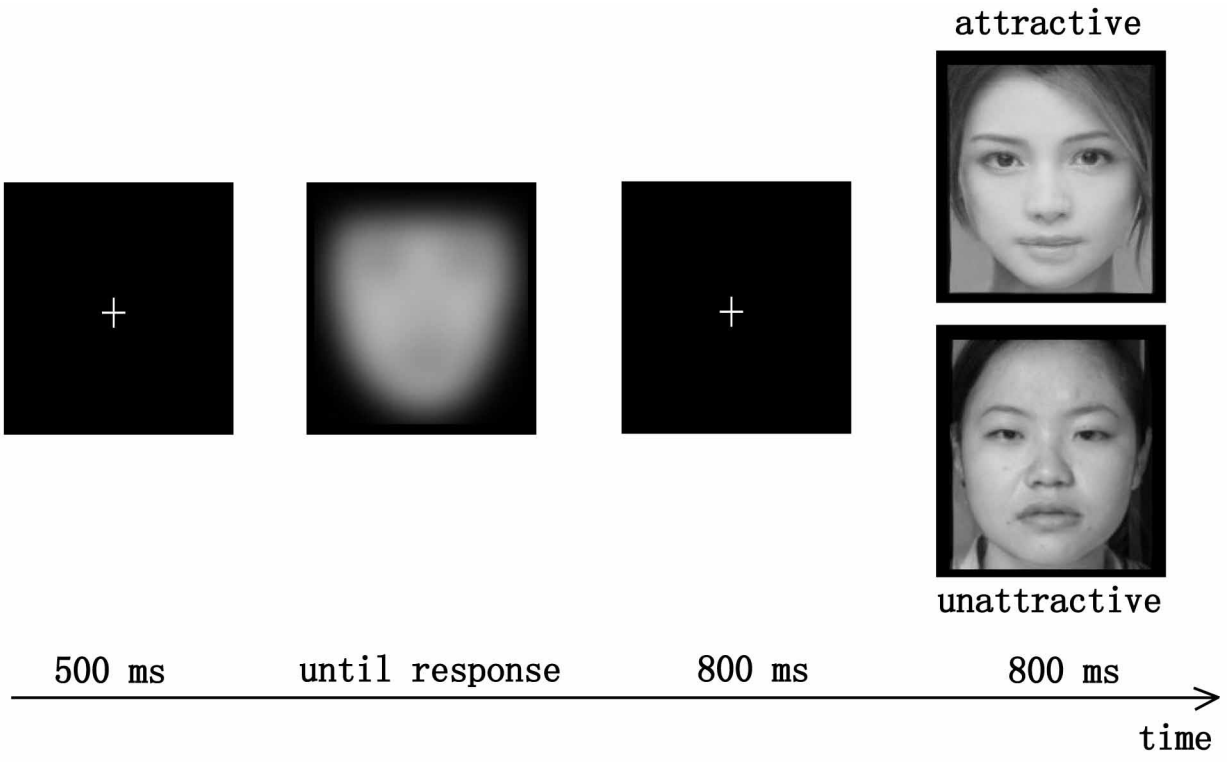


FIGURE 1 | Sequence of events in a single trial. For illustration purpose, the attractive and the unattractive faces in the figure are the ones morphed from several faces used in the experiment.

algorithm which employs a regression analysis in combination with artifact averaging (Semlitsch et al., 1986). Epochs contaminated by blinks and other movement artifacts were excluded from averaging using an 80 μV criterion. The EEG data were low-pass filtered at 30 Hz and were baseline-corrected by subtracting the average activity of that electrode during the baseline period from each trial. After excluding trials with artifacts, each participant had at least 46 trials in each condition.

The grand-average ERP waveforms (**Figure 2**) did not show a typical pattern for ERP responses that were observed for feedback stimuli in gambling or simple cognitive tasks (e.g., no clear P300 component was visible), although it appeared that inconsistent faces elicited negative-going deflections in the 300–380 ms time window. We, therefore, analyzed ERP responses in different conditions in the windows of 150–220 ms (i.e., P200), 300–380 ms (i.e., FRN), and 380–500 ms based on visual inspection. For the purposes of statistical analysis, mean amplitudes for each time window was calculated across 25 electrode locations (F3, F1, Fz, F2, F4, FC3, FC1, FCz, FC2, FC4, C3, C1, Cz, C2, C4, CP3, CP1, CPz, CP2, CP4, P3, P1, Pz, P2, P4) that were chosen to cover scalp areas known from previous studies to be the focus of the FRN and P300. A repeated-measures analysis of variance (ANOVA) was conducted, with attractiveness (attractive vs. unattractive), feedback consistency (consistent vs. inconsistent), anterior-posterior scalp location (frontal, frontocentral, central, centroparietal, parietal), and lateral scalp location (left, left central, midline, right central, right) as four within-subjects experimental factors. The Greenhouse-Geisser correction for violation of the ANOVA assumption of sphericity was applied in all analysis. Bonferroni corrections were used for multiple comparisons.

Given that the processing of the feedback faces and their attractiveness was likely to involve sophisticated neurocognitive processes, it is possible that the FRN and the P300 components were not only overlapping in the time course, but also masked by other cognitive (and emotional) processes associated with the complex feedback stimuli. To examine whether the typical P300 effects that were observed in previous studies for various aspects of the outcome evaluation could also be observed for the more complex feedback faces, we performed principal-component analysis

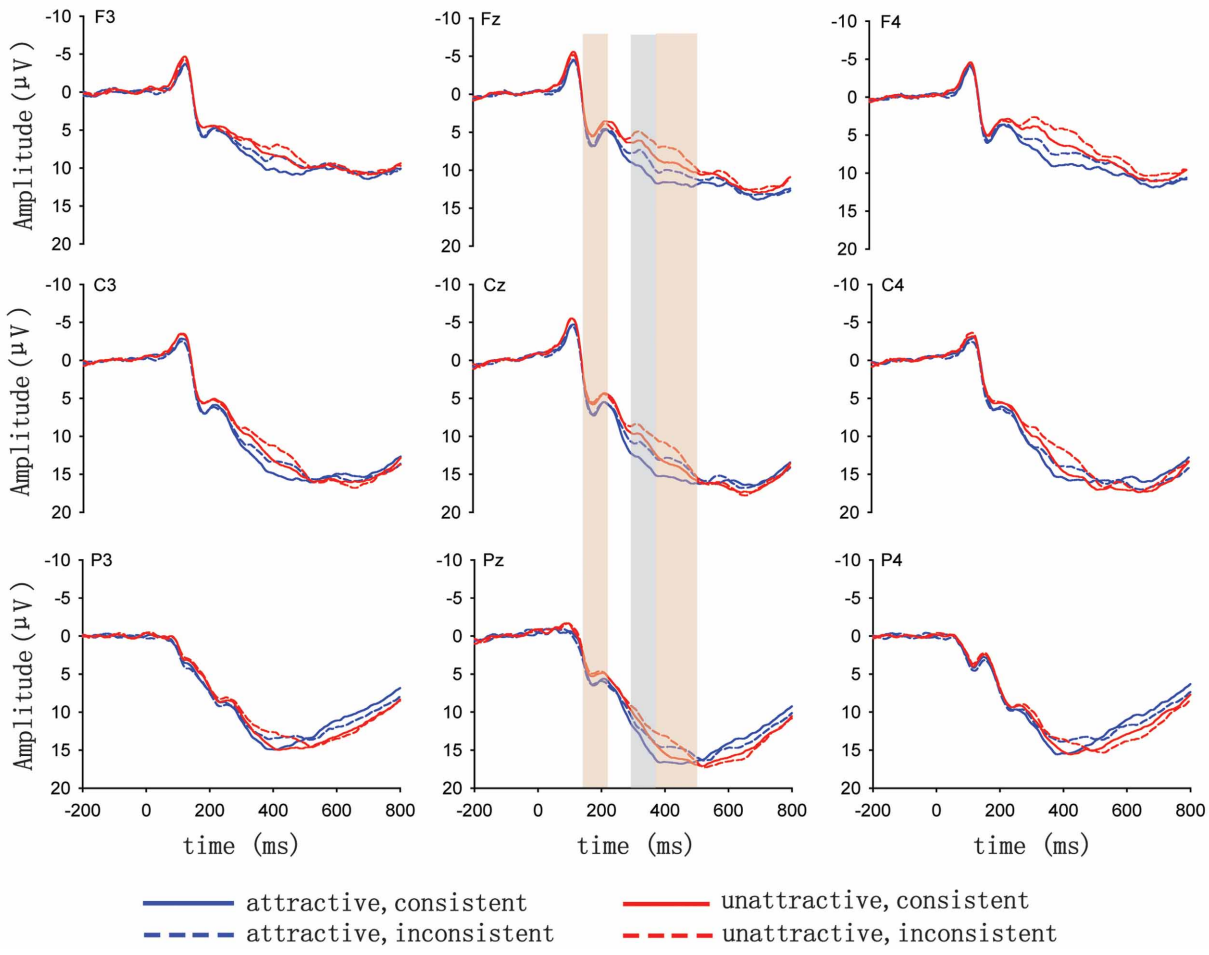


FIGURE 2 | ERP responses to feedback faces at exemplar electrodes as a function of attractiveness and feedback consistency, with 0 ms corresponding to the onset of feedback faces.

(PCA) on the *cleaned* ERP data (i.e., after preprocessing) in order to disentangle the overlapping and/or masked ERP components. PCA has a wide range of applications in ERP analysis, such as cleaning or filtering noises prior to data analysis, or being used in data exploration as a way to detect and summarize features of the dataset. In this study, we applied PCA on the cleaned ERP data to maximize the possibility that the PCA factors represent interpretable signals (i.e., brain activity due to experimental manipulations) as opposed to noise (i.e., artifacts or background EEG).

PCA is a common approach for decomposing an ERP dataset into its constituent factors by summarizing the relationship between variables (such as microvolt recordings at each time point in temporal PCA or at each electrode in spatial PCA; Dien and Frishkoff, 2005; Dien et al., 2005). This process consists of three main steps: (1) computation of the covariance matrix which captures the interrelationships between temporal/spatial variables; (2) extraction and retention of the PCA factors which extract linear combinations of variables (latent factors) to account for patterns of covariance in the ERP data with the fewest PCA factors; and (3) rotation to simple structure, which is used to restructure the allocation of variables to PCA factors to maximize the chance that each PCA factor reflects a single ERP component. These steps yield two matrices, which are useful in further analysis. The first one is a factor loading matrix, representing correlations between the variables and the factor scores (e.g., describing the time course of each of the PCA factors in temporal PCA). The second one is a factor score matrix that indexes the magnitude of the PCA factors for each of the observations, reflecting the contribution of each PCA factor to ERPs.

In this study, we used the spatiotemporal PCA algorithm implemented in the Matlab ERP PCA toolbox 2.23 (Dien and Frishkoff, 2005; http://sourceforge.net/projects/erppcatoolkit/files). In the spatiotemporal PCA, a spatial PCA was performed first on the ERP data from all the electrodes to capture spatial distribution, with the combination of 501 time points (covering the epoch from –200 to 800 ms post-onset of the feedback stimuli), 16 participants and four experimental conditions. Infomax rotation was used, and two spatial factors were extracted based on the resulting Scree plot. For each spatial factor, this analysis yielded factor scores for each combination, representing the amount of activity in the original data captured by that factor. A temporal PCA was then performed for each of the two spatial factors. Promax rotation was used, and four temporal factors were extracted based on the Scree plot, yielding eight unique factors combinations (i.e., two spatial factors by four temporal factors). The covariance relationship matrix and Kaiser normalization were used for each PCA. The waveforms for each factor combination were reconstructed (i.e., converted to microvolts) by multiplying the factor pattern matrix with the standard deviations. A PCA factor may contain one or more known ERP components, or may contain no well-defined or theoretically interesting ERP components. Following the suggestion of Dien et al. (2005) and based on visual inspection of the waveforms associated with each PCA factor, two factors that contained known ERP components involved in outcome evaluation and facial attractiveness (i.e., P300) were selected for further statistical analysis, with attractiveness (attractive vs. unattractive) and feedback consistency (consistent vs. inconsistent) as two within-subjects experimental factors. Here we focused on electrodes that showed the largest effect for each selected PCA factor. The Greenhouse–Geisser correction was applied for violation of the ANOVA assumption of sphericity.

RESULTS

BEHAVIORAL RESULTS

In the post-test on the attractiveness of the feedback faces, 377 trials (6.9%) in total were inconsistent with the categorization of attractiveness based on the pretest (i.e., their rating scores were below five for attractive faces or above three for unattractive faces). After excluding the EEG trials corresponding to these faces, the remaining 5063 trials were categorized into four conditions: attractive-consistent (72 trials per participant on average, ranging from 51 to 117 trials over participants), attractive-inconsistent (78 trials on average, ranging from 48 to 100 trials), unattractive-consistent (91 trials on average, ranging from 61 to 117 trials), and unattractive-inconsistent (76 trials on average, ranging from 48 to 107 trials). Overall, the percentage of trials (out of all the available trials) for each condition was 22.6%, 24.5%, 28.9%, and 24.0%, respectively. No statistically significant differences were found between the conditions.

ERP RESULTS

ERP waveforms time-locked to the onset of feedback faces and topographic distributions of the differences between conditions (e.g., ERP effects) in three time windows are illustrated in **Figures 2** and **3**, respectively.

For the 150–220 ms time window (**Figures 2**, **3**), repeated-measures ANOVA showed a significant main effect of attractiveness, $F(1,15) = 18.39$, $p = 0.01$, with more positive P200 responses to attractive faces (5.95 μV) than to unattractive ones (4.98 μV). The interaction between attractiveness and lateral scalp location was significant, $F(4,60) = 5.17$, $p = 0.001$, indicating that the size of the P200 effect varied over different scalp locations. No other significant effects were found.

For the 300–380 ms window, ANOVA showed a significant main effect of feedback consistency, $F(1,15) = 15.99$, $p = 0.001$, with ERP responses to feedback faces more negative-going following inconsistent judgments (9.97 μV) than following consistent judgments (11.36 μV). The main effect of attractiveness was also significant, $F(1,15) = 20.48$, $p < 0.001$, with more negative-going responses to unattractive faces (9.56 μV) than to attractive ones (11.78 μV). The interaction between attractiveness and feedback consistency was not significant, $F(1,15) < 1$, indicating that the attractiveness effect was not affected by feedback consistency. The interaction between attractiveness and anterior-posterior scalp location was significant, $F(4,60) = 17.66$, $p < 0.001$, as was the interaction between attractiveness and lateral scalp location, $F(4,60) = 8.21$, $p = 0.01$. It is clear from **Figure 3** that the attractiveness effect in this time window was strongest in the frontocentral regions. No other significant effects were found.

To further verify the feedback consistency effect, we carried out statistical analyses based on peak-to-peak measurements of the feedback consistency responses (Sato et al., 2005; Wu and Zhou,

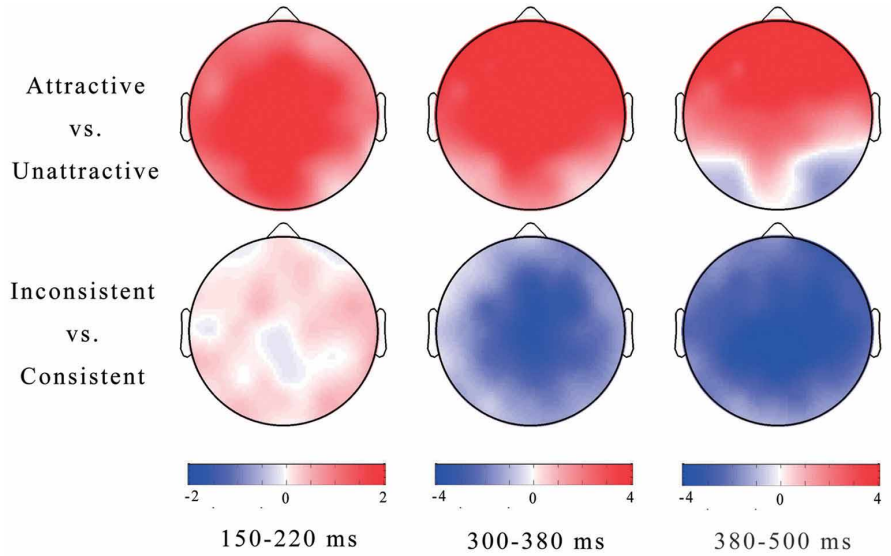


FIGURE 3 | Topographies of ERP effects in three time windows.

2009), with the most negative amplitudes in the 300–380 ms window and the most positive amplitudes in the 220–300 ms window as peaks. Consistent with the above analysis, the main effect of feedback consistency was significant, $F(1,15) = 15.38$, $p = 0.001$, and this effect did not interact with attractiveness, $F(1,15) < 1$. The main effect of attractiveness was significant, $F(1,15) = 20.51$, $p < 0.001$.

For the 380–500 ms window, ANOVA showed a significant main effect of attractiveness, $F(1,15) = 13.31$, $p < 0.01$, with more positive responses to attractive faces (13.62 μV) than to unattractive ones (12.52 μV), and a main effect of feedback consistency, $F(1,15) = 13.15$, $p < 0.01$, with more negative responses to feedback faces following inconsistent judgments (12.27 μV) than following consistent judgments (13.87 μV). However, the interaction between attractiveness and feedback consistency was not significant, $F(1,15) < 1$, suggesting that the ERP responses in this time window may encode the valence and attractiveness of feedback faces independently.

PCA RESULTS

Of the eight factor combinations yielded by the spatiotemporal PCA, two were selected for further statistical analysis based on visual inspection of the shape and the time course of peak responses in the generated waveforms (**Figure 4**). Either of the selected factors might correspond to the P300 component of the ERP. The first factor, maximal at CPz in terms of amplitude, reached its peak in the 200–300 ms window, accounting for 13.7% of the data variance. The second factor, maximal at FCz in terms of amplitude, contained a positive deflection beginning at about 250 ms and lasting to 800 ms, accounting for 27.8% of the data variance.

For the first selected factor (the upper panel of **Figure 4**), repeated-measures ANOVA based on the mean amplitudes in the 200–300 ms window at CPz revealed a significant main effect of attractiveness, $F(1,15) = 29.03$, $p < 0.001$, with more positive responses to attractive faces (6.30 μV) than to unattractive ones (4.91 μV). The main effect of feedback consistency was significant, $F(1,15) = 6.31$, $p < 0.05$, with faces consistent with the initial judgment eliciting more positive responses (5.87 μV) than faces inconsistent with the judgment (5.34 μV). The interaction between attractiveness and feedback consistency was not significant, $F(1,15) < 1$. In addition, we conducted ANOVA based on the peak amplitude in the 200–300 ms window. Consistent with the above analysis, the main effect of attractiveness was significant, $F(1,15) = 27.23$, $p < 0.001$, and this effect did not interact with feedback consistency, $F(1,15) < 1$. The main effect of feedback consistency was significant, $F(1,15) = 6.03$, $p < 0.05$.

For the second selected factor (the lower panel of **Figure 4**), repeated-measures ANOVA based on the mean amplitudes in the 250–800 ms window at FCz revealed a significant main effect of attractiveness, $F(1,15) = 43.00$, $p < 0.001$, with more positive responses to attractive faces (5.37 μV) than to unattractive ones (3.32 μV). The main effect of feedback consistency was significant, $F(1,15) = 10.38$, $p < 0.01$, with faces consistent with the initial judgment eliciting more positive responses (4.76 μV) than faces inconsistent with the judgment (3.93 μV). The interaction between attractiveness and feedback consistency was not significant, $F(1,15) < 1$.

DISCUSSION

The main purpose of this study was to investigate to what extent brain responses to complex feedback stimuli with a social dimension (e.g., faces) in outcome evaluation and performance monitoring are similar to those revealed for simple stimuli used in monetary gambling or simple cognitive tasks. We asked participants to guess the attractiveness of blurred faces and then showed them the unblurred faces. The implicit comparison between the initial judgment stored in memory and the newly acquired

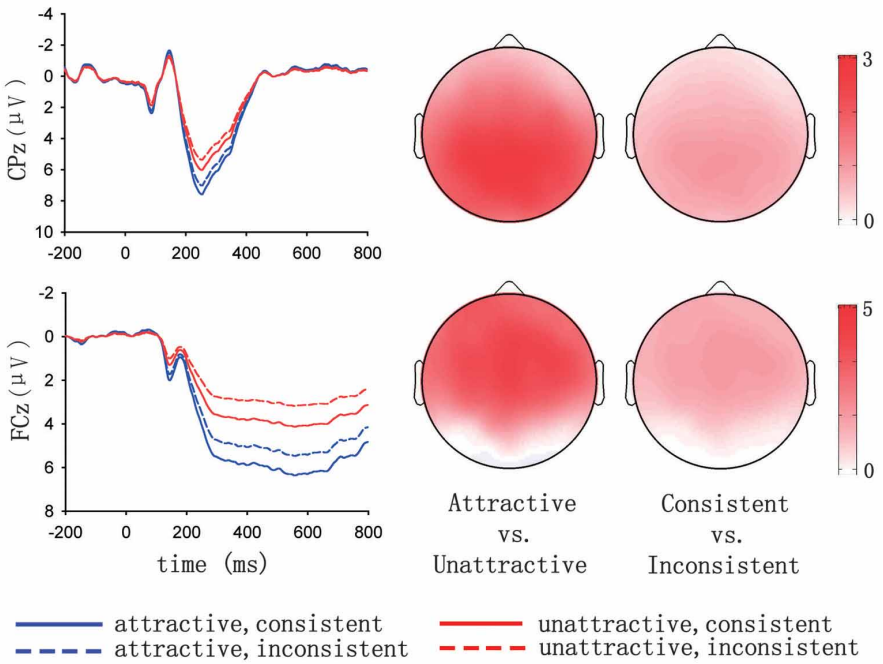


FIGURE 4 | ERP responses on the exemplar electrodes converted from factor scores in the spatiotemporal PCA and depicted as a function of attractiveness and feedback consistency. Topographic maps on the right side present the differential responses corresponding to the factors depicted on the left. Sketches in the upper and lower panels correspond to the first and the second selected spatiotemporal factors, respectively.

information would indicate whether the feedback stimuli had properties consistent or inconsistent with the initial judgment.

An FRN effect was observed for inconsistent feedback faces in ERP waveforms, in accordance with many earlier studies (Gehring and Willoughby, 2002; Holroyd and Coles, 2002; Yeung et al., 2005; Leng and Zhou, 2010; Zhou et al., 2010; Long et al., 2012), albeit in a delayed 300–380 ms time window. Faces inconsistent with the initial judgment elicited more negative-going responses than faces consistent with the judgment. The delay of the FRN effect was likely due to the complexity of neurocognitive processes involved in facial attractiveness and feedback consistency processing. To know whether the initial judgment or guess was a correct one, the system has to first process the perceptual structure of the feedback face, and evaluate the attractiveness of this face by appealing to experiences or schemas stored in long-term memory, and secondly to compare the result of this evaluation with information concerning the initial judgment stored in working memory. It takes time to accomplish these processes, delaying the appearance of the FRN effect, even though assessing facial attractiveness can be fairly automatic (Chatterjee et al., 2009). A previous study has also shown that the FRN effect was delayed when the sophisticated counter-factual thinking was involved in accessing the outcome valence of gambling choices (Yu and Zhou, 2009).

However, the ERP waveforms did not show the typical pattern of P300 responses; instead, sustained deflections were observed, with significant main effects for attractiveness and feedback consistency. We suspected that the usual P300 responses were masked by the sophisticated cognitive (and emotional) processes involved in processing the feedback faces, and the observed deflections (and possibly the preceding ERP responses) contained P300 components. We, therefore, conducted spatiotemporal PCA, which revealed a clear pattern of P300 effects for both feedback consistency and attractiveness, with more positive responses to those faces consistent with the initial judgment than to faces inconsistent with the judgment, and more positive responses to attractive faces than to unattractive ones. Moreover, the P300 effects centered on centroparietal areas lasted for a short time period while the P300 effects centered on frontocentral area sustained for a long time period, indicating that facial attractiveness and feedback consistency could be processed parallel in different brain areas.

The P300 effect for feedback consistency or valence replicated previous studies using gambling tasks (Hajcak et al., 2005, 2007; Wu and Zhou, 2009; Leng and Zhou, 2010). The P300 effect for facial attractiveness also replicated many previous studies (Johnston and Oliver-Rodriguez, 1997; Oliver-Rodríguez et al., 1999; Werheid et al., 2007; Schacht et al., 2008). As attractive faces activate brain areas dedicated to reward processing such as the orbitofrontal cortex, the nucleus accumbens or the ventral striatum (Aharon et al., 2001; Kampe et al., 2001; O'Doherty et al., 2003; Bray and O'Doherty, 2007; Ishai, 2007; Winston et al., 2007; Cloutier et al., 2008), the attractive feedback faces in this study might be considered to be a form of reward, having stronger motivational significance, and capturing more attentional resources than unattractive faces (Sabatinelli et al., 2004). A large number of studies have already demonstrated that the P300 is sensitive to processes that demand attentional resources

(Polich and Kok, 1995; Polich, 2007). The independence between the P300 effect for feedback consistency and the effect for facial attractiveness, while replicating previous studies that simultaneously manipulated different dimensions of feedback stimuli (Hajcak et al., 2005, 2007; Wu and Zhou, 2009; Long et al., 2012), indicate that the evaluative processes for different dimensions of the outcome can be conducted in parallel.

Note that we also found a significant main effect of attractiveness in the 300–380 ms window. However, we do not interpret it as a kind of FRN effect because we believe this effect was very likely due to the spillover of the P300 effects for facial attractiveness.

To conclude, this study provides the first demonstration of electrophysiological responses in outcome evaluation with complex feedback stimuli (e.g., faces) that need complicated neurocognitive processing. We demonstrated that the ERP correlates of processing complex feedback stimuli with a social dimension are generally similar to those involved in processing simplex feedback stimuli; however, appropriate means of ERP data analysis, including PCA, may be needed to uncover the underlying patterns of effects that might be masked by the sophisticated processes associated with the complex stimuli themselves. Consistent with previous studies, the present study demonstrates again that the FRN reflects early assessment of outcome valence while the P300 can encode different attributes of feedback simultaneously in outcome evaluation and performance monitoring.

ACKNOWLEDGMENTS

This study was supported by National Basic Research Program of China (973 Program: 2010CB833904) and by a grant from the Natural Science Foundation of China (30110972). We thank Mr. Zhiheng Zhou and Ms. Chaohui Guo for their help during the preparation for the experiment, Mr. Stijn Massar, Mr. Stephen Politzer-Ahles, and the two reviewers for their comments on the earlier versions of the manuscript, and Ms. Sarah Dinces for English correction.

REFERENCES

Aharon, I., Etcoff, N., Ariely, D., Chabris, C. F., O'Connor, E., and Breiter, H. C. (2001). Beautiful faces have variable reward value: fMRI and behavioral evidence. *Neuron* 32, 537–551.

Bray, S., and O'Doherty, J. (2007). Neural coding of reward-prediction error signals during classical conditioning with attractive faces. *J. Physiol.* 98, 3036–3045.

Chatterjee, A., Thomas, A., Smith, S. E., and Aguirre, G. K. (2009). The neural response to facial attractiveness. *Neuropsychology* 23, 135–143.

Cloutier, J., Heatherton, T. F., Whalen, P. J., and Kelley, W. M. (2008). Are attractive people rewarding? Sex differences in the neural substrates of facial attractiveness. *J. Cogn. Neurosci.* 20, 941–951.

Dien, J., Beal, D. J., and Berg, P. (2005). Optimizing principal components analysis of event-related potentials: matrix type, factor loading weighting, extraction, and rotations. *Clin. Neurophysiol.* 116, 1808–1825.

Dien, J., and Frishkoff, G. A. (2005). "Introduction to principal components analysis of event-related potentials," in *Event-related Potentials: A Methods Handbook*, ed T. Handy (Cambridge, MA: MIT Press), 189–208.

Dion, K., Berscheid, E., and Walster, E. (1972). What is beautiful is good. *J. Pers. Soc. Psychol.* 24, 285–290.

Etcoff, N. (1999). *Survival of the Prettiest: The Science of Beauty.* New York, NY: Doubleday.

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

Hajcak, G., Holroyd, C. B., Moser, J. S., and Simons, R. F. (2005). Brain potentials associated with expected and unexpected good and bad outcomes. *Psychophysiology* 42, 161–170.

Hajcak, G., Moser, J. S., Holroyd, C. B., and Simons, R. F. (2007). It's worse than you thought: the feedback negativity and violations of reward prediction in gambling tasks. *Psychophysiology* 44, 905–912.

Heldmann, M., Russeler, J., and Munte, T. (2008). Internal and external information in error processing. *BMC Neurosci.* 9, 33–40.

Holroyd, C. B., and Coles, M. G. H. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychol. Rev.* 109, 679–709.

Ishai, A. (2007). Sex, beauty and the orbitofrontal cortex. *Int. J. Psychophysiol.* 63, 181–185.

Johnston, V. S. (2006). Mate choice decisions: the role of facial beauty. *Trends Cogn. Sci.* 10, 9–13.

Johnston, V. S., and Oliver-Rodriguez, J. C. (1997). Facial beauty and the late positive component of event-related potentials. *J. Sex Res.* 34, 188–198.

Kampe, K. K. W., Frith, C. C., Dolan, R. J., and Frith, U. (2001). Reward value of attractiveness and gaze. *Nature* 413, 589–590.

Leng, Y., and Zhou, X. (2010). Modulation of the brain activity in outcome evaluation by interpersonal relationship: an ERP study. *Neuropsychologia* 48, 448–455.

Long, Y., Jiang, X., and Zhou, X. (2012). To believe or not to believe: trust choice modulates brain responses in outcome evaluation. *Neuroscience* 200, 50–58.

Miltner, W. H. R., Braun, C. H., and Coles, M. G. H. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task: evidence for a "generic" neural system for error detection. *J. Cogn. Neurosci.* 9, 787–796.

Nieuwenhuis, S., Holroyd, C. B., Mol, N., and Coles, M. G. H. (2004). Reinforcement-related brain potentials from medial frontal cortex: origins and functional significance. *Neurosci. Biobehav. Rev.* 28, 441–448.

O'Doherty, J., Winston, J., Critchley, H., Perrett, D., Burt, D. M., and Dolan, R. J. (2003). Beauty in a smile: the role of medial orbitofrontal cortex in facial attractiveness. *Neuropsychologia* 41, 147–155.

Oliver-Rodríguez, J. C., Guan, Z., and Johnston, V. S. (1999). Gender differences in late positive components evoked by human faces. *Psychophysiology* 36, 176–185.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol.* 118, 2128–2148.

Polich, J., and Kok, A. (1995). Cognitive and biological determinants of P300: an integrative review. *Biol. Psychol.* 41, 103–146.

Sabatinelli, D., Flaisch, T. C., Bradley, M. M., Fitzsimmons, J. R., and Lang, P. J. (2004). Affective picture perception: gender differences in visual cortex? *Neuroreport* 15, 1109–1112.

Sato, A., Yausda, A., Ohira, H., Miyawaki, K., Nishikawa, M., Kumano, H., and Kuboki, T. (2005). Effects of value and reward magnitude on feedback negativity and P300. *Neuroreport* 16, 407–411.

Schacht, A., Werheid, K., and Sommer, W. (2008). The appraisal of facial beauty is rapid but not mandatory. *Cogn. Affect. Behav. Neurosci.* 8, 132–142.

Semlitsch, H. V., Anderer, P., Schuster, P., and Presslich, O. (1986). A solution for reliable and valid reduction of ocular artifacts, applied to the P300 ERP. *Psychophysiology* 23, 695–703.

Thornhill, R., and Gangestad, S. W. (1999). Facial attractiveness. *Trends Cogn. Sci.* 3, 452–460.

Werheid, K., Schacht, A., and Sommer, W. (2007). Facial attractiveness modulates early and late event-related brain potentials. *Biol. Psychol.* 76, 100–108.

Winston, J. S., O'Doherty, J., Kilner, J. M., Perrett, D. I., and Dolan, R. J. (2007). Brain systems for assessing facial attractiveness. *Neuropsychologia* 45, 195–206.

Wu, Y., and Zhou, X. (2009). The P300 and reward valence, magnitude, and expectancy in outcome evaluation. *Brain Res.* 1286, 114–122.

Yeung, N., Holroyd, C. B., and Cohen, J. D. (2005). ERP correlates of

feedback and reward processing in the presence and absence of response choice. *Cereb. Cortex* 15, 535–544.

Yeung, N., and Sanfey, A. G. (2004). Independent coding of reward magnitude and valence in the human brain. *J. Neurosci.* 24, 6258–6264.

Yu, R., and Zhou, X. (2009). To bet or not to bet? The error negativity or error-related negativity associated with risk-taking choices. *J. Cogn. Neurosci.* 21, 684–696.

Zhou, Z., Yu, R., and Zhou, X. (2010). To do or not to do? Action enlarges the FRN and P300 effects in outcome evaluation. *Neuropsychologia* 48, 3606–3613.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 22 September 2011; accepted: 08 February 2012; published online: 22 February 2012.

*Citation: Zhang Y, Li X, Qian X and Zhou X (2012) Brain responses in evaluating feedback stimuli with a social dimension. Front. Hum. Neurosci. **6**:29. doi: 10.3389/fnhum.2012.00029*

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 06 July 2012
doi: 10.3389/fnhum.2012.00204

Neural correlates of feedback processing in decision-making under risk

Beate Schuermann*, Tanja Endrass and Norbert Kathmann

Department of Psychology, Humboldt-Universität zu Berlin, Berlin, Germany

Edited by:
Christian Bellebaum, Ruhr University Bochum, Germany

Reviewed by:
Markus Ullsperger, Max Planck Institute for Neurological Research, Germany
Olave E. Krigolson, Dalhousie University, Canada

****Correspondence:***
Beate Schuermann, Department of Psychology, Institut für Psychologie, Humboldt-Universität zu Berlin, Rudower Chaussee 18, 12489 Berlin, Germany.
e-mail: beate.schuermann@hu-berlin.de

Introduction: Event-related brain potentials (ERPs) provide important information about the sensitivity of the brain to process varying risks. The aim of the present study was to determine how different risk levels are reflected in decision-related ERPs, namely the feedback-related negativity (FRN) and the P300. **Materials and Methods:** Twenty participants conducted a probabilistic two-choice gambling task while an electroencephalogram (EEG) was recorded. Choices were provided between a low-risk option yielding low rewards and low losses and a high-risk option yielding high rewards and high losses. While options differed in expected risks, they were equal in expected values and in feedback probabilities. **Results:** At the behavioral level, participants were generally risk-averse but modulated their risk-taking behavior according to reward history. An early positivity (P200) was enhanced on negative feedbacks in high-risk compared to low-risk choices. With regard to the FRN, there were significant amplitude differences between positive and negative feedbacks on high-risk choices, but not on low-risk choices. While the FRN on negative feedbacks did not vary with decision riskiness, reduced amplitudes were found for positive feedbacks in high-risk relative to low-risk choices. P300 amplitudes were larger in high-risk decisions, and in an additive way, after negative compared to positive feedback. **Discussion:** The present study revealed significant influences of risk and valence processing on ERPs. FRN findings suggest that the reward prediction error signal is increased after high-risk decisions. The increased P200 on negative feedback in risky decisions suggests that large negative prediction errors are already processed in the P200 time range. The later P300 amplitude is sensitive to feedback valence as well as to the risk associated with a decision. Thus, the P300 carries additional information for reward processing, mainly the enhanced motivational significance of risky decisions.

Keywords: decision-making, feedback processing, P200, P300, FRN

INTRODUCTION

A significant function of the human brain is to assess the riskiness of decisions in order to prevent negative outcomes. Brain imaging studies indicate that frontolimbic brain circuits involving the ventromedial prefrontal cortex, amygdala, insula, ventral striatum, and anterior cingulate cortex (ACC), are implicated in risk processing. In particular, the ACC is important for detecting and evaluating unfavorable outcomes (Bush et al., 2000; Luu et al., 2000), and for risk assessment (Ernst et al., 2004; Fukui et al., 2005; McCoy and Platt, 2005). Greater ACC activity predicts enhanced error avoidance (Johansen and Fields, 2004; Frank et al., 2005) and less risk-taking behavior (Paulus and Frank, 2006). An influential model of decision-making under risk is the prospect theory (Tversky and Kahneman, 1981). It proposes that human decision makers are generally risk avoiding when choosing between alternatives. Nevertheless, it has been shown that risk-taking behavior may also depend on the context (Tversky and Kahneman, 1981), i.e., risk aversion increases after gains and decreases after losses.

Studies using event-related brain potentials (ERPs) have revealed that the human brain is able to evaluate the outcome of actions within a few 100 ms. Specific brain potentials are elicited by self-generated responses and performance feedback (Holroyd and Coles, 2002). The error-related negativity (ERN; Falkenstein et al., 1990; Gehring et al., 1990) and the feedback-related negativity (FRN; Miltner et al., 1997) are elicited by erroneous responses and by negative feedback or losses, respectively. ERN and FRN are assumed to originate from the anterior midcingulate cortex (Gehring and Willoughby, 2002; Debener et al., 2005). Therefore, ERN and FRN may reflect similar mechanisms of monitoring and controlling behavior. It has been suggested that the ACC uses reinforcement learning (RL) signals conveyed by the midbrain dopamine system to optimize future decision-making behavior (Holroyd and Coles, 2002). According to the RL theory, ERN and FRN reflect a reward prediction error signal in the ACC that occurs when ongoing events are worse than expected. Subsequently, the ACC triggers an adaptive modification of behavior by relating actions with their consequences (Holroyd and Coles, 2002; Rushworth et al., 2004). Another ERP component that has been shown to carry important information for reward processing is the feedback-related P300, a parietally distributed positivity (Yeung and Sanfey, 2004; Polezzi et al., 2009). It has been suggested that the feedback-related P300 may reflect the extent to which information is motivationally

significant or salient (for a review, see Nieuwenhuis et al., 2005). In line with that, the P300 amplitude varies with the motivational significance of feedback information (Yeung and Sanfey, 2004; Polezzi et al., 2009) and is increased in individuals who attributed more meaning to feedback (de Bruijn et al., 2004).

Economic decision theories presume that risk depends on potential losses and increases with its probability and magnitude (Tversky and Kahneman, 1981; Brown and Braver, 2007). In this regard, rational decisions are made on the basis of the expected value, which is a multiplicative combination of the two components (Machina, 1982). Recent studies investigated the different components of risk-taking by assessing the influences of feedback valence, magnitude, and probability on ERP amplitudes. Research on the impact of the probability of feedback has generally shown that both the FRN and the P300 are modulated by this variable, with unexpected feedback being associated with enhanced amplitudes (Holroyd et al., 2004; Hajcak et al., 2007). Furthermore, it was demonstrated that amplitudes are modulated by an interaction between feedback valence and expectancy: unexpected negative feedback is associated with larger amplitudes compared to unexpected positive feedback (Frank et al., 2005; Moser and Simons, 2009). In gambling paradigms, an additional important variable associated with decision-making under risk is outcome magnitude. Yeung and Sanfey (2004) studied the effects of winning or losing large or small amounts of money on the FRN and P300 and concluded that only the P300 was affected by the amount of monetary loss, whereas the FRN was insensitive to outcome magnitude. In line with this, Toyomaki and Murohashi (2005) reported effects of magnitude on the participants' subjective assessment of losses, but no effects on FRN amplitudes (see also Sato et al., 2005; Hajcak et al., 2006). Other studies reported significant magnitude effects on the FRN (Goyer et al., 2008; Wu and Zhou, 2009). However, tasks in these studies required participants to choose from alternatives without having any information about reward magnitude. To conclude, FRN and P300 seem to reflect different aspects of risk processing in economic decision-making, valence and magnitude processing, respectively.

A limitation of most previous studies is that they did not independently control for the effects of probability, magnitude, and expected value. Some studies focusing on neural correlates of feedback processing used different expected values of choices to determine learning (van der Helden et al., 2010; Schuermann et al., 2011). Furthermore, sometimes participants were unaware of possible outcome magnitudes prior to receiving feedback, and thus could not make informed choices (Goyer et al., 2008; Wu and Zhou, 2009). Finally, for gambling tasks, choices often differed in outcome probability (Yeung and Sanfey, 2004; Cohen et al., 2007). To overcome some of these limitations, we designed a gambling task in which expected risk was independently manipulated from expected values and reward probability. Specifically, participants were requested to select between a low-risk option yielding low rewards and low losses and a high-risk option yielding high rewards and high losses. Unlike traditional RL tasks used in ERP research, participants in the present task were not required to learn outcome contingencies throughout the course of the task. In this study, expected values were equal for both options. There was also no difference in reward probabilities between the low-risk and the high-risk option. Examining risk effects also requires that probabilities involved in a decision are explicitly known (Brand et al., 2006; Brown and Braver, 2007). Therefore, in the present task participants were informed about the outcome probabilities. In sum, the present task should provide a better account to assess pure risk preference and to evaluate the influence of risk parameters on ERPs.

The aim of the present study was to determine how different risk levels are reflected in decision-related ERPs, namely the FRN and the feedback-related P300. Therefore, we developed and tested a novel two-choice gambling task allowing for the examination of risk-taking in unambiguous situations (Pilot experiment). The associated electrocortical indicators of risk-taking behavior were examined in the main experiment. Considering that expected values of high-risk and low-risk options were equal, we predicted that participants are predominantly risk-averse, namely that they are less willing to choose risky options (Tversky and Kahneman, 1981). Furthermore, we assumed that participants are more risk-averse following gains and relatively more risk-seeking following losses (Tversky and Kahneman, 1981). According to the RL theory, which states that the FRN responds to the difference between experienced and anticipated rewards, we predicted enhanced FRN amplitudes for high-risk compared to low-risk decisions. We also assumed that P300 amplitudes would be enhanced for high-risk decisions compared to low-risk ones due to an enhanced motivation of risky decisions (Nieuwenhuis et al., 2005).

PILOT EXPERIMENT

MATERIALS AND METHODS

Participants

Fifty participants (30 women and 20 men) took part in the pilot experiment. Their mean age was 30.5 years (SD: 11.4; range: 18–50). Three of the participants were left-handed. Participants had no history of neurological or psychiatric diseases. All participants received verbal and written explanations of the purpose and procedures of the study, and gave written informed consent in accordance with the Declaration of Helsinki.

Task and procedure

A computerized probabilistic two-choice gambling task was administered, which involved low-risk and high-risk options. On each trial participants were asked to choose between two options that were presented on a computer screen (see **Figure 1**). The colors of the stimuli indicated the relative probability of winning (green), which was always 75%, and the relative probability of losing (red), which was always 25%. Reward magnitudes associated with choice options were displayed in each stimulus. Choices were made by pressing one of two corresponding response buttons. After 700 ms, participants were shown the outcome associated with the selected option for 1100 ms. A red frowny face together with a negative amount indicated negative feedback, while a green smiley face together with a positive amount indicated positive feedback. In addition, the total account balance across trials was presented below the feedback stimuli. Choices had to be made within 2300 ms, otherwise participants were

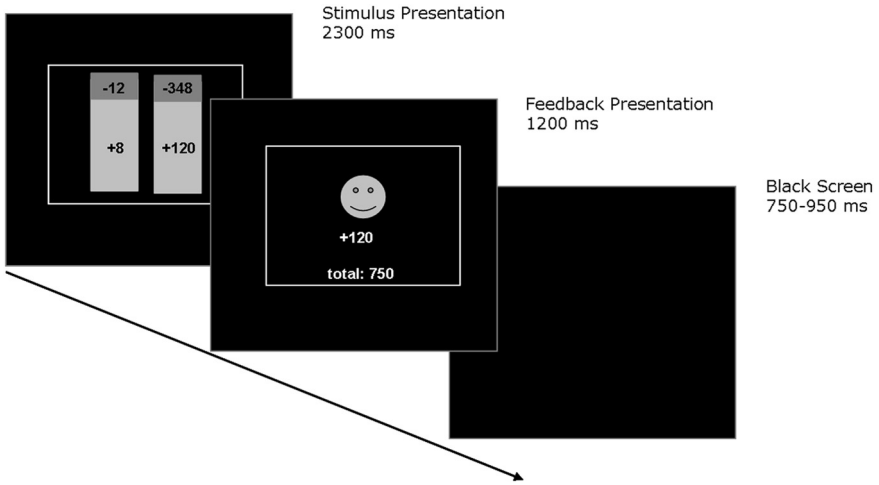


FIGURE 1 | Schematic depiction of the probabilistic two-choice gambling task. During each trial, participants were asked to choose between two options that were represented visually by a histogram. The colors of the histogram indicated the relative probability of gaining (colored green) which was always 75% and the relative probability of losing (colored red) which was always 25%. The current amount of gains and losses associated with each option were displayed in each histogram. Choices were made by pressing one of two corresponding response buttons. After 700 ms, participants were shown the outcome associated with their choice for 1100 ms. A red frowny face together with a negative amount indicated negative feedback, while a green smiley face together with a positive amount indicated positive feedback. In addition, the current total amount was presented below the feedback stimuli.

prompted to respond more quickly. The next trial was presented after an intertrial interval of 750–950 ms. Following standardized written instructions, participants performed two practice trials. The pilot experiment consisted of 112 total trials and lasted about 5 min. Participants were instructed to earn as many points as possible and were told that each point corresponds to one Euro cent. Participants received on average 4.50€ in the pilot experiment. **Table 1** presents an overview of the reinforcement schedule. In each trial, participants always had to choose between options A and B (56 trials) or options C and D (56 trials). The options with the larger maximum outcomes were termed as high-risk (options B and D), and the options with the smaller maximum outcomes were termed as low-risk (options A and C). Positions of options on the computer screen changed across trials in pseudo-random order. At the beginning of the pilot experiment, participants were informed that presented options differed in expected risks, while the expected values were equal for high-risk and low-risk options. According to Brown and Braver (2007), expected risk of each option was defined as [loss probability × (rewards – losses)], expected value of each option was defined as [(reward probability × rewards) + (loss probability × losses)].

Table 1 | Reinforcement schedule in the probabilistic two-choice gambling task.

	Low-risk options (A)	High-risk options (B)	Low-risk options (C)	High-risk options (D)
Reward (75%)	+7	+43	+8	+120
Losses (25%)	−9	−117	−12	−348
Expected risk	−0, 5	−18, 5	−1	−57
Expected value	3	3	3	3

Data analyses

To assess risk-taking behavior, percentages (relative to the total amount of choices) of low-risk (options A and C) and high-risk choices (options B and D) were determined and analyzed using two-tailed *t*-tests. Percentages of low-risk and high-risk choices were further analyzed as a function of total account balance (positive account balance; i.e., >0€ vs. negative account balance; i.e., <0€), performing two-tailed *t*-tests. Moreover, we analyzed whether the probability of high-risk choices on a given trial varied as a function of prior feedback valence and prior risk-taking behavior. This was done with an ANOVA with the within-subject factors previous feedback valence (gains vs. losses on the previous trial) and previous risk-taking (low-risk options vs. high-risk options on the previous trial). Statistical analysis was carried out with the Predictive Analytic Software (PASW) 19.0 for Windows.

RESULTS

Table 2 presents the behavioral results. A significant preference for the low-risk options over the high-risk options was found throughout the task, [$t_{(49)} = 2.84$, $p = 0.007$]. The analysis of risk-taking as a function of actual account balance revealed that participants avoided high-risk options when their current balance was positive, [$t_{(49)} = 4.71, p < 0.001$]. By contrast, high-risk options were preferred when the current balance was negative, [$t_{(49)} = 4.05$, $p < 0.001$]. Further, it was shown that risk preference varied as a function of prior feedback valence, [$F_{(1,\ 49)} = 25.62$, $p < 0.001$], and prior risk-taking, [$F_{(1,\ 49)} = 14.85$, $p < 0.001$]. The interaction of feedback valence and prior risk-taking

Table 2 | Behavioral results of the pilot experiment ($N = 50$) and the main experiment ($N = 20$) presenting mean (M) and standard deviation (SD).

	Low-risk options		High-risk options		*T*	df	*p*
	M (%)	SD (%)	M (%)	SD (%)			
PILOT EXPERIMENT							
–	56.9	17.2	43.1	17.2	2.84	49	0.007
Positive balance	49.1	18.1	29.5	13.9	4.71	49	0.000
Negative balance	7.7	8.4	13.6	8.4	−4.05	49	0.000
MAIN EXPERIMENT							
–	59.2	19.7	40.8	19.7	2.09	19	0.050
Positive balance	49.0	20.4	26.3	13.5	3.15	19	0.005
Negative balance	10.1	8.3	14.5	8.4	−1.75	19	0.096

Note: Percentages of low-risk and high-risk choices were analyzed as a function of total account balance (positive account balance; i.e., > 0€ vs. negative account balance; i.e., < 0€), performing two-tailed t-tests.

was not significant, [$F_{(1,\ 49)} < 1, p = 0.970$]. These effects reflect that participants preferred higher risks following losses than following gains, as well as following a high-risk decision as compared to a low-risk decision.

DISCUSSION

With this pilot experiment we aimed to explore risk-taking behavior using a probabilistic two-choice gambling task. During each trial, participants were required to choose between options associated with two different risk levels. As expected, participants preferred the low-risk options over the high-risk options, although options did not differ with respect to expected values. Results are consistent with previous findings of Polezzi et al. (2008). In that study, participants had to choose between a predictable option (which was always associated with a gain of 10€) and an unpredictable option (which was associated with a gain of 30€ or a loss of 10€). The results showed a clear preference for options associated with a predictable outcome, although the expected value of both options was identical. Analysis of the choice history also revealed a loss avoidance tendency among participants. Participants strongly avoided the high-risk options following gains and when they had positive balances. This was not the case after losses and with negative account balances. When faced with rewarding feedback, participants were possibly more willing to protect the money they had and thus showed more conservative behavior. By contrast, the increase in risk proclivity might occur due to an anticipation of larger monetary rewards in order to reduce negative consequences (in terms of corrective actions). These findings are in line with previous studies (Gehring and Willoughby, 2002; Goyer et al., 2008), showing that participants are more likely to engage in risky choices following losses. In summary, the pilot experiment demonstrated the usefulness of the two-choice gambling task as a suitable test for examining risk-taking behavior in unambiguous situations.

MAIN EXPERIMENT

MATERIALS AND METHODS

Participants

20 participants (five men) attended in the ERP study. Their mean age was 29.5 years (SD: 8.9; range, 21–52). All participants were right-handed and had no history of neurological or psychiatric disorders. All participants received verbal and written explanations of the purpose and procedures of the study, and gave written informed consent in accordance with the Declaration of Helsinki. Participants gave written informed consent in accordance with the Declaration of Helsinki.

Task and procedure

Task and procedure for Experiment 2 were identical to the pilot experiment except that it comprised of 640 trials that were divided into three blocks with short breaks between blocks. Again, positions of high and low-risk options (left or right on the screen) varied across trials in pseudo-random order. In each trial, participants had to chose between options A and B (320 trials), or options C and D (320 trials). The increase in trial number was necessary to obtain a sufficient number of trials for ERP analyses in all conditions. The experiment lasted about 40 min. All participants were paid 15€ for their participation. To ensure ecological validity of the task and to enhance motivation, participants were informed that they would additionally receive the highest amount they earned in one of the three blocks. The average earning was 6.70€ in this experiment.

EEG recording and data analyses

The electroencephalogram (EEG) was recorded from 64 electrodes sites including Cz as a recording reference, using an equidistant electrode system (EASYCAP GmbH, Herrsching-Breitbrunn, Germany). The montage also included additional electrodes that were placed on external locations: below the left and right eye (IO1 and IO2) and in the neck. The ground electrode was located below T1. Electrode impedances were kept below 5 kΩ. Electrical activity was sampled digitally at a rate of 500 Hz, using a time constant of 10 s and a low-pass filter of 250 Hz. Individual electrode positions were digitized based on the run-time measurement of ultrasonic pulses using ELPOS (zebris Medical GmbH, Isny i. Allgäu, Germany). Offline, the EEG data were re-referenced to average reference and corrected for eye-movement artifacts using the multiple source eye correction method as implemented in BESA 5.1 (Brain Electrical Source Analysis, MEGIS Software GmbH, Gräfelfing, Germany).

For the FRN and P300 analyses, raw data were filtered offline with a low-pass filter of 40 Hz and an additional notch filter at 50 Hz. Feedback-locked epochs were obtained for each trial, starting 200 ms prior to feedback onset and continuing for 1000 ms post-feedback. Individual averages were baseline-corrected to an average activity between –200 and 0 ms before feedback onset. Feedback-locked epochs were excluded from further analyses if they still contained artifacts. For each participant, ERPs were averaged separately for feedback valance (positive vs. negative) and risk levels (low-risk options vs. high-risk options).

Three components were analyzed. The P200 is a positive component peaking at around 200 ms after feedback onset and measured over frontal areas (Carretié et al., 2001; Polezzi et al., 2008). In the present study, the P200 was determined as the most positive peak between the time window of 100 to 300 ms at electrodes Fz and FCz. FRN amplitudes were computed as the difference between the most negative peak following feedback onset in a 200 to 400 ms time window and the preceding positive peak in the 100 to 300 ms time window at electrodes Fz and FCz. Prior to peak detection, ERPs were filtered with a 15 Hz low-pass filter. The P300 was quantified at CPz and Pz and defined as the mean amplitude in the time range between 300 and 400 ms after feedback presentation. ERP time windows were based on the visual inspection of the grand-average waveforms (for P200 and FRN: Fz, for P300: Pz). Repeated-measurement ANOVAs were computed for the analysis of ERP data, with the within-subject factors electrode (for P200 and FRN: Fz and FCz, for P300: CPz and Pz), feedback valence (gains vs. losses) and risk option (low-risk options vs. high-risk options). Analyses of the behavioral data were identical to the pilot experiment. Correlation coefficients (Pearson r) were used to examine associations between FRN and P300 magnitude and percentage of low-risk choices (relative to the percentage of the high-risk choices). All statistical tests were two-tailed.

RESULTS

Behavioral results

As in the pilot experiment, we found a preference for low-risk options over high-risk options, [$t_{(19)} = 2.09$, $p = 0.050$] (cf. **Table 2**). The analysis of risk-taking as a function of account balance revealed that participants avoided high-risk options more when their current balance was positive compared to when it was negative, [$t_{(19)} = 3.15$, $p = 0.005$]. Unlike the pilot data, participants only tended to increase risk-taking behavior when their current balance was negative compared to when it was positive, [$t_{(19)} = 1.76$, $p = 0.096$]. Furthermore, it was found that the proportion of chosen high-risk options varied as a function of previous outcome valence and previous risk-taking. Consistent with the pilot experiment, we found main effects for feedback valence, [$F_{(1,\ 19)} = 16.98$, $p = 0.001$], and for risk-taking, [$F_{(1,\ 19)} = 11.57$, $p = 0.003$]. Following losses, participants made more risky decisions than after gains. Following risky choices, participants were more likely to choose a high-risk option than after having made a low-risk choice. Finally, a significant interaction was found between previous feedback valence and previous risk-taking, indicating that most high-risk choices were made after high-risk choices followed by negative feedback, [$t_{(19)} = 4.37$, $p < 0.001$].

ERP results

Figure 2 presents feedback-locked ERP waveforms on positive (dashed line) and on negative (solid line) feedback trials, separately for selected high-risk and low-risk options. **Figure 3** displays ERPs for positive and negative feedback trials for a comparison of the high-risk (solid line) and low-risk options (dashed line). Inspection of ERPs indicated three distinct components related to risk processing. The first component is the P200, an early positive wave peaking at a latency of approximately 200 ms. The FRN is the negative deflection between two positive components. The third component is the P300, peaking approximately between 300 and 400 ms following feedback onset. Losses elicited large P300 amplitudes which may overlap with the FRN effect. It is noteworthy that ERP waveforms indicate that FRN peak amplitudes were lower for losses than for gains which are mainly due to the variation of P200 and P300 amplitudes. Therefore, FRN amplitudes were calculated as the difference between the most negative peak amplitude minus the preceding positive peak amplitude. With regard the P200, a significant main effect of valence was found, [$F_{(1,\ 19)} = 67.56$, $p < 0.001$], indicating larger amplitudes for losses compared to gains. There was also a significant main effect of risk, [$F_{(1,\ 19)} = 43.87$, $p < 0.001$], revealing enhanced amplitudes in high-risk options compared to low-risk options. Furthermore, a significant interaction of risk and valence was found, [$F_{(1,\ 19)} = 19.36$, $p < 0.001$]. In high-risk decisions, larger amplitude differences between positive and negative feedbacks were found which was due to enhanced P200 amplitudes on negative feedbacks in high-risk options. The main effect of electrode was also significant, [$F_{(1,\ 19)} = 34.59$, $p < 0.001$], with larger P200 at FCz compared to Fz.

Consistent with previous findings, a main effect of valence was found for FRN amplitudes, indicating enhanced FRN amplitudes on losses compared to gains, [$F_{(1,\ 19)} = 11.08$, $p = 0.004$]. Furthermore, the FRN was modulated by risk as evidenced by a significant interaction between valence and risk option, [$F_{(1,\ 19)} = 10.12$, $p = 0.005$]. After choosing the high risk option, large amplitude differences between losses and gains were found, [$t_{(19)} = 4.03$, $p = 0.001$]. By contrast, there was no significant amplitude difference between losses and gains in the low-risk condition, [$t_{(19)} = 1.08$, $p = 0.293$]. While the FRN on losses did not differ with respect to the riskiness of options, [$t_{(19)} = 0.65$, $p = 0.522$], amplitudes on gains were reduced in the high-risk compared to the low-risk option, [$t_{(19)} = -3.55$, $p = 0.002$]. The main effect of electrode also approached significance, [$F_{(1,\ 19)} = 3.46$, $p = 0.079$], due to larger FRN amplitudes at Fz relative to FCz.

The P300 is also depicted in **Figures 2** and **3**. A main effect of valence was found, [$F_{(1,\ 19)} = 18.06$, $p < 0.001$], indicating that the P300 was larger on losses compared to gains. There was a significant main effect of risk option, [$F_{(1,\ 19)} = 36.08$, $p < 0.001$], showing enhanced amplitudes in high-risk compared to low-risk choices. No significant interaction between feedback valence and risk option was observed ($F < 1$). The P300 tended to be larger at CPz compared to Pz, [$F_{(1,\ 19)} = 3.48$, $p = 0.078$].

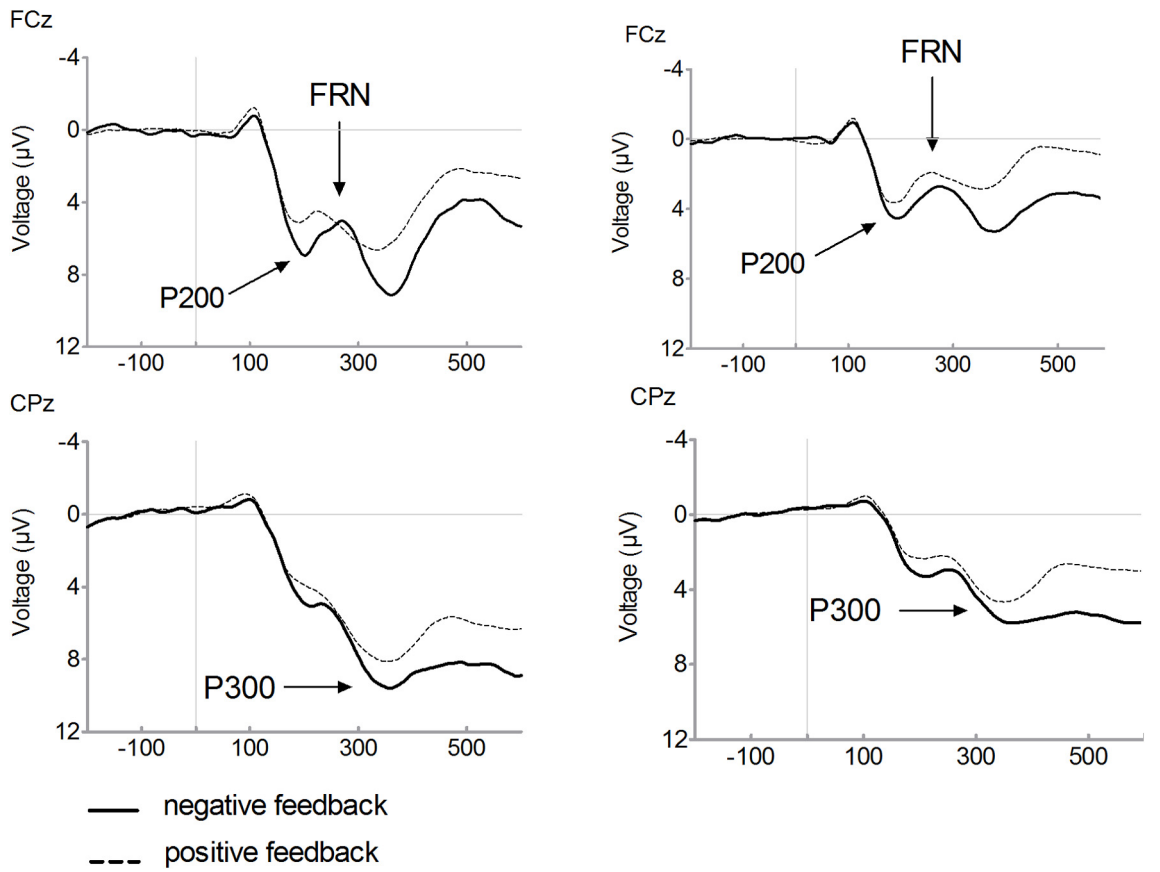


FIGURE 2 | Feedback-related brain potential waveforms for high-risk and low-risk decisions. Averaged feedback-locked event-related brain potential (ERP) waveforms are presented for recording sites FCz and CPz. ERPs for high-risk decisions are depicted in the left columns and for low-risk decisions in the right column. Waveforms for negative feedback (solid line) and for positive feedback (dashed line) are overlaid. The P200 was determined as the most positive peak between 100 and 300 ms. The FRN was computed as the difference between the most negative peak following feedback onset in a 200 to 400 ms time window and the preceding positive peak between 100 and 300 ms. The P300 was defined as the mean amplitude in the time range between 300 and 400 ms after feedback presentation.

Correlational findings

Bivariate correlations were computed relating ERPs following losses and following gains (averaged across FCz and Fz for FRN analyses, and across CPz and Pz for P300 analyses) to the percentage of low-risk choices. Negative correlations for FRN amplitudes indicate an increase in FRN (i.e., more negative amplitudes) with risk aversion.

We found significant correlations between risk avoidance and FRN amplitudes on gains ($r = -0.47$, $p = 0.039$), the FRN on losses at a trend level ($r = -0.41$, $p = 0.073$), and P300 amplitudes on losses ($r = 0.52$, $p = 0.018$). No significant correlations were found between risk-avoidance and P300 amplitudes after gains. Note that FRN and P300 amplitudes were not correlated ($r = -0.02$, $p = 0.91$).

GENERAL DISCUSSION

The current study focused on decision-making and its neural correlates using a monetarily motivated probabilistic two-choice gambling task. In accordance with previous findings, participants were generally risk-averse but modulated their risk-taking behavior according to reward history (Tversky and Kahneman, 1981). ERP research on decision-making has left the question unanswered as to how electrophysiological indicators are specifically affected by different risk levels. In order to address this issue, we independently controlled for different risk parameters.

Differential processing of risky decisions was reflected in the FRN amplitudes. The FRN was modulated by the well-known distinction between gains and losses (Gehring and Willoughby, 2002; Hajcak et al., 2006), but only in high-risk options. Considered in the framework of the RL theory of the FRN (Holroyd and Coles, 2002), it appears that the reward prediction error signal is increased after high-risk decisions, where larger potential positive and negative consequences were expected. The null effect under the low-risk condition might be explained by generally smaller reward prediction errors generated in the

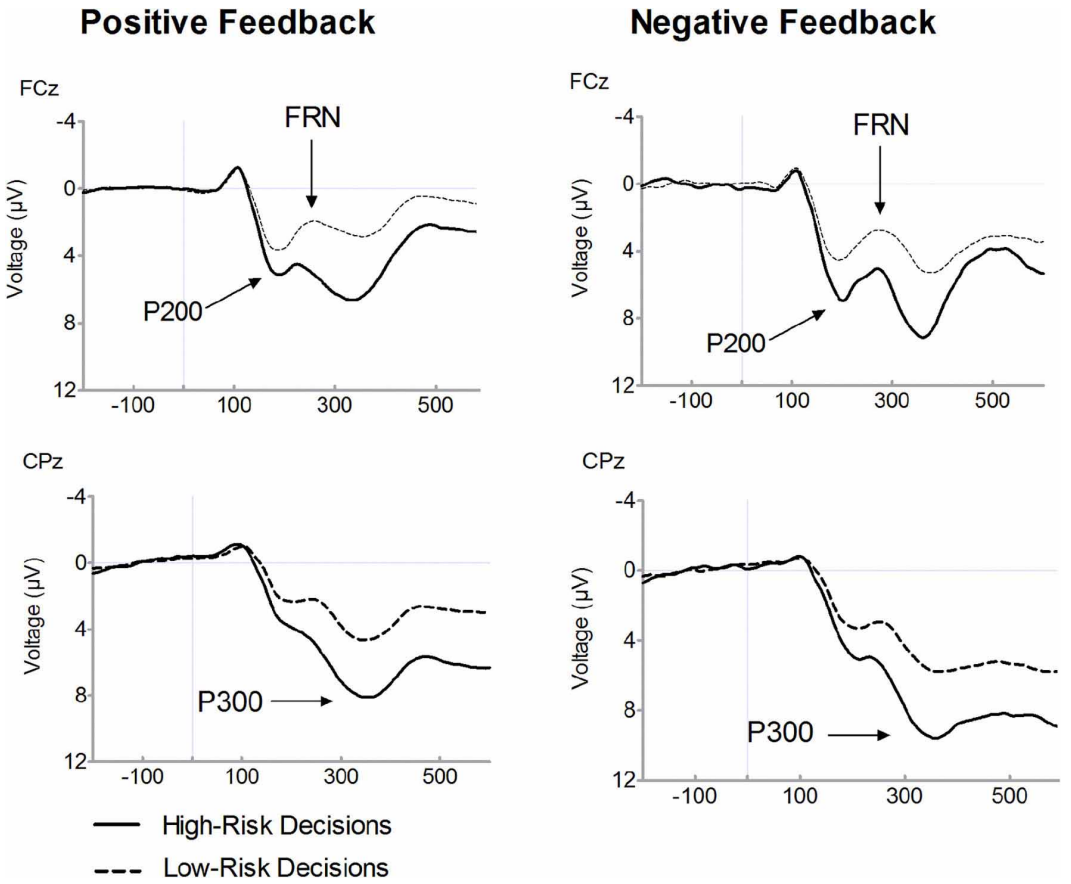


FIGURE 3 | Feedback-related brain potential waveforms for positive (i.e., gains) and negative feedback (i.e., losses). Averaged feedback-locked event-related brain potential (ERP) waveforms are presented for recording sites FCz and CPz. ERPs for positive feedback (i.e., gains) are depicted in the left columns and for negative feedback (i.e., losses) in the right column. Waveforms for high-risk decisions (solid line) and for low-risk-decisions (dashed line) are overlaid.

low-risk condition, characterized by small positive and small negative outcomes. Results are also in line with Holroyd et al. (2004), showing that the FRN reflects loss sizes in relation to what was expected. Whereas Holroyd and Coles (2002) interpreted the FRN purely as a reinforcement signal, Gehring and Willoughby (2002) suggested that the FRN might also reflect the motivational impact of ongoing events. Thus, modulations in high-risk decisions may also reflect the motivational or emotional significance of high-risk decisions compared with low-risk decisions. Due to the potential negative consequences of high-risk choices, discriminating between losses and gains seems highly important for optimizing future decision-making. Consistent with this interpretation, brain imaging studies emphasized the role of the ACC in evaluating unfavorable outcomes (Bush et al., 2000; Luu et al., 2000) and in risk assessment (Ernst et al., 2004; Fukui et al., 2005). Also, cingulate lesions in monkeys have been shown to impair the ability to use previous reinforcements to guide future decision-making behavior (McCoy and Platt, 2005; Kennerley et al., 2006). Alternatively, the FRN pattern in high-risk decisions might be due to variations of the P200 which was also affected by outcome valence. However, the P200 was enhanced on negative feedback trials in high-risk options whereas the FRN on negative feedbacks did not vary with risk. Interestingly, the interaction of feedback valence and risk was mainly caused by FRN amplitude variations on positive feedbacks. In the present study, reduced amplitudes were found for positive feedbacks in high-risk relative to low-risk choices. Larger amplitudes were found for lower outcomes and for smaller positive reward prediction errors. Our results contribute to a growing debate about the relevance of positive feedback for reward-related processes and several studies found greater modulation of amplitudes on gain trials compared to loss trials (Cohen et al., 2007; Holroyd et al., 2008). It has been shown that the amplitude following positive feedbacks varied with the probability of reward. Thus it was argued that it may represent the magnitude of the positive reward prediction error (Cohen et al., 2007). Possibly, the current result of reduced FRN amplitudes in the high-risk condition may be a consequence of the larger positive reward prediction error associated with gains in that condition. However, few studies examined the effects on positive feedbacks, and consistent patterns of results have not been observed yet. Therefore, these results have to be interpreted with caution, and further studies

are needed to examine the effects of reward prediction error on different risk parameters.

Decision riskiness also affected the P200, a component that has previously been associated with the attention processing of emotional stimuli, such as faces. The more negative the valence, the larger the P200 amplitude (Carretié et al., 2001, 2005). Consistent with these findings, in the present study negative feedback (i.e., losses) and high-risk decisions induced larger P200s. The present results indicate that amplitudes were modulated by outcome magnitude, especially in high-risk gambles. Polezzi et al. (2008) reported that unpredictable outcomes are associated with larger P200 amplitudes compared to predictable outcomes, which is consistent with the current results. But, in that study the P200 was not sensitive to the distinction between positive and negative outcome. Our data suggest an early processing of negative feedback and of high-risk decisions. Also, Bellebaum et al. (2010) revealed that P200 amplitudes are larger under a reward outcome condition compared to a non-reward outcome condition both in active and observational gambling tasks. Moreover, the active execution induced a larger discrepancy of P200 amplitudes between reward and non-reward than that of passive observation. The present findings suggest an enhanced sensitivity in risky decisions to the gain- and loss-outcome difference at a very early stage. Importantly, we found large reward prediction errors in high-risk options as early as in the P200 time range. Possibly, the P200 codes the most relevant features of a context, especially when risky decisions have to be made in order to avoid future negative consequences.

In agreement with previous studies (Frank et al., 2005; Schuermann et al., 2011), P300 amplitudes were enhanced on negative compared to positive feedbacks. In addition to valence effects, enhanced P300 amplitudes were found in high-risk options relative to low-risk options. Both results are in accordance with the finding that the feedback-related P300 is sensitive to reward probability (Bellebaum and Daum, 2008; Pfabigan et al., 2011) since negative outcomes were less probable in the current study and with the finding that the P300 is sensitive to reward magnitude since negative outcomes in the high-risk condition were of greater magnitude (Yeung and Sanfey, 2004; Hajcak et al., 2005, 2007). The P300 enhancement in high-risk choices may reflect enhanced motivational significance of risky decisions. This is supported by the hypothesis that the P300 may reflect motivational processes linked to noradrenergic transmission (Nieuwenhuis et al., 2005).

Interestingly, FRN and P300 appear to reflect risk-taking behavior, but they might reflect different aspects of risky decision-making. Whereas FRN amplitudes are reduced by large positive prediction errors, P300 amplitudes are enhanced due to larger negative outcomes. Importantly, risk-avoidance behavior was associated with enhanced FRN and P300 amplitudes. Possibly, increased FRN amplitudes reflect enhanced cognitive control that is essential for the avoidance of risky decisions. The association between FRN and risk-avoidance is in line with previous findings, revealing an inverse relation between ERN/FRN amplitudes and risk-taking behavior in healthy individuals (e.g., Hewig et al., 2007; Santesso and Segalowitz, 2009) and in patients with borderline personality disorder (Ruchsow et al., 2006; Schuermann et al., 2011). FRN findings complement brain imaging results, suggesting that greater ACC activity predicts less risk-seeking behavior (Paulus and Frank, 2006). Nonetheless, interpretation of correlation analyses of the present study should be cautiously interpreted due to the relatively small sample size.

To our knowledge, this is the first study that independently controlled for different risk parameters. Nevertheless, there are possible confounds that should be discussed. First, the comparison between positive and negative feedbacks is confounded with feedback probability, i.e., positive feedbacks were more frequent than negative feedbacks. Therefore, reward probability may have influenced the difference between positive and negative feedbacks (Holroyd et al., 2003; but: Hajcak et al., 2005). While this should have affected the distinction between positive and negative feedbacks in both conditions, an amplitude difference was only found in the high-risk condition. In addition, feedback probability should not affect the comparison of feedback types between high- and low-risk gambles, since respective feedback probabilities were equal in both conditions. Second, although we aimed to disentangle expected risk from feedback probability, we could not independently manipulate outcome magnitude and reward prediction errors. Future studies on decision-making behavior under risk should further examine the influence of valence, magnitude, and expected risk on behavioral and ERP parameters to describe the underlying neural mechanisms more precisely. In particular, future research could parametrically vary risk parameters such that they vary from trial to trial in a decorrelated fashion.

To conclude, the present findings indicate that the processes underlying human decision-making are significantly affected by decision riskiness when controlling for reward probility and expected value. The increased P200 on negative feedback in high-risk decisions suggests that large reward prediction errors are processed as early as in the P200 time range. The FRN is affected by feedback valence depending on decision riskiness. Considered in the framework of the RL theory of the FRN, it has been suggested that the reward prediction error signal is increased after high-risk decisions compared to low-risk decisions. The later P300 amplitude is sensitive to feedback valence as well as to the risk associated with a decision. Thus, the P300 carries additional information for reward processing, mainly the enhanced motivational significance of risky decisions. Due to the potential negative consequences of high-risk choices, rapidly processing the relevant and informative features of a context when decisions have to be made seemed highly important for optimizing future decision-making. Risk-taking is a central cognitive-motivational construct accounting for many everyday decisions. In addition, understanding the neurocognitive basis of risk-taking behavior might also be central to explaining certain symptoms of psychopathological conditions, e.g., borderline personality disorder, patients with bipolar disorders or patients with substance dependency.

ACKNOWLEDGMENTS

The authors thank T. Pinkpank, Dipl.-Psych, R. Kniesche, Dipl.-Ing., and Robert M. Blasiak, M.Sc. for their helpful support, and S. Kessler-Scheil, Dipl.-Psych. for her help in data collection.

REFERENCES

Bellebaum, C., and Daum, I. (2008). Learning-related changes in reward expectancy are reflected in the feedback-related negativity. *Eur. J. Neurosci.* 27, 1823–1835.

Bellebaum, C., Kobza, S., Thiele, S., and Daum, I. (2010). It was not MY fault: event-related brain potentials in active and observational learning from feedback. *Cereb. Cortex* 20, 2874–2883.

Brand, M., Labudda, K., and Markowitsch, H. J. (2006). Neuropsychological correlates of decision-making in ambiguous and risky situations. *Neural Netw.* 19, 1266–1276.

Brown, J. W., and Braver, T. S. (2007). Risk prediction and aversion by anterior cingulate cortex. *Cogn. Affect. Behav. Neurosci.* 7, 266–277.

Bush, G., Luu, P., and Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends Cogn. Sci.* 4, 215–222.

Carretié, L., Hinojosa, J. A., Mercado, J. A. F., and Tapia, M. (2005). Cortical response to subjectively unconscious danger. *Neuroimage* 24, 615–623.

Carretié, L., Mercado, J. A. F., Tapia, M., and Hinojosa, J. A. (2001). Emotion, attention, and the 'negative bias', studied through event-related potentials. *Int. J. Psychophysiol.* 41, 75–85.

Cohen, M. X., Elger, C. E., and Ranganath, C. (2007). Reward expectation modulates feedback-related negativity and EEG spectra. *Neuroimage* 35, 968–978.

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

de Bruijn, E. R. A., Mars, R. B., and Hulstijn, W. (2004). "It wasn't me ... or was it?" How false feedback affects performance," in *Errors, Conflicts, and the Brain Current Opinions on Performance Monitoring*, eds M. Ullsperger and M. Falkenstein (Leipzig: MPI of Cognitive Neuroscience), 118–124.

Eppinger, B., Kray, J., Mock, B., and Mecklinger, A. (2008). Better or worse than expected? Aging, learning and the ERN. *Neuropsychologia* 46, 521–529.

Ernst, M., Nelson, E. E., McClure, E. B., Monk, C. S., Munson, S., Eshel, N., Zarahn, E., Leibenluft, E., Zametkin, A., Towbin, K., Blair, J., Charney, D., and Pine, D. S. (2004). Choice selection and reward anticipation: an fMRI study. *Neuropsychologia* 42, 1585–1597.

Falkenstein, M., Hohnsbein, J., Hoormann, J., and Blanke, L. (1990). "Effects of errors in choice reaction tasks on the ERP under focused and devided attention," in *Psychophysiological Brain Research*, eds C. H. M. Brunia, A. W. K. Gaillard, and A. Kok (Tilburg: Tilburg University Press), 192–195.

Frank, M. J., Woroch, B. S., and Curran, T. (2005). Error-related negativity predicts reinforcement learning and conflict biases. *Neuron* 47, 495–501.

Fukui, H., Murai, T., Fukuyama, H., Hayashi, T., and Hanakawa, T. (2005). Functional activity related to risk anticipation during performance of the Iowa Gambling Task. *Neuroimage* 24, 253–259.

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

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

Goyer, J. P., Woldorff, M. G., and Huettel, S. A. (2008). Rapid electrophysiological brain responses are influenced by both valence and magnitude of monetary rewards. *J. Cogn. Neurosci.* 20, 2058–2069.

Hajcak, G., Holroyd, C. B., Moser, J. S., and Simons, R. F. (2005). Brain potentials associated with expected and unexpected bad outcomes. *Psychophysiology* 42, 161–170.

Hajcak, G., Moser, J. S., Holroyd, C. B., and Simons, R. F. (2006). The feedback-related negativity reflects the binary evaluation of good versus bad outcomes. *Biol. Psychol.* 71, 148–154.

Hajcak, G., Moser, J. S., Holroyd, C. B., and Simons, R. F. (2007). It's worse than you thought: the feedback negativity and violations of reward prediction in gambling tasks. *Psychophysiology* 44, 905–912.

Hewig, J., Trippe, R., Hecht, H., Coles, M. G., Holroyd, C. B., and Miltner, W. H. (2007). Decision-making in Blackjack: an electrophysiological analysis. *Cereb. Cortex* 17, 865–877.

Holroyd, C. B., and Coles, M. G. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychol. Rev.* 109, 679–709.

Holroyd, C. B., Nieuwenhuis, S., Yeung, N., and Cohen, J. D. (2003). Errors in reward prediction are reflected in the event-related brain potential. *Neuroreport* 14, 2481–2484.

Holroyd, C. B., Larsen, J. T., and Cohen, J. D. (2004). Context dependence of the event-related brain potential associated with reward and punishment. *Psychophysiology* 41, 245–253.

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

Johansen, J. P., and Fields, H. L. (2004). Glutamatergic activation of anterior cingulate cortex produces an aversive teaching signal. *Nat. Neurosci.* 7, 398–403.

Kennerley, S. W., Walton, M. E., Behrens, T. E., Buckley, M. J., and Rushworth, M. F. (2006). Optimal decision making and the anterior cingulate cortex. *Nat. Neurosci.* 9, 940–947.

Luu, P., Flaisch, T., and Tucker, D. M. (2000). Medial frontal cortex in action monitoring. *J. Neurosci.* 20, 464–469.

Machina, M. J. (1982). Expected utility analysis without the independence axiom. *J. Math. Psychol.* 30, 188–205.

McCoy, A. N., and Platt, M. L. (2005). Expectations and outcomes: decision-making in the primate brain. *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* 191, 201–211.

Miltner, W. H. R., Braun, C. H., and Coles, M. G. H. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task: evidence for a 'generic' neural system for error detection. *J. Cogn. Neurosci.* 9, 788–798.

Moser, J. S., and Simons, R. F. (2009). The neural consequences of flip-flopping: the feedback-related negativity and salience of reward prediction. *Psychophysiology* 46, 313–320.

Nieuwenhuis, S., Aston-Jones, G., and Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychol. Bull.* 131, 510–532.

Paulus, M. P., and Frank, L. R. (2006). Anterior cingulate activity modulates nonlinear decision weight function of uncertain prospects. *Neuroimage* 30, 668–677.

Pfabigan, D. M., Alexopoulos, J., Bauer, H., and Sailer, U. (2011). Manipulation of feedback expectancy and valence induces negative and positive reward prediction error signals manifest in event-related brain potentials. *Psychophysiology* 48, 656–664.

Polezzi, D., Lotto, L., Daum, I., Sartori, G., and Rumiati, R. (2008). Predicting outcomes of decisions in the brain. *Behav. Brain Res.* 187, 116–122.

Polezzi, D., Sartori, G., Rumiati, R., Vidotto, G., and Daum, I. (2009). Brain correlates of risky decision-making. *Neuroimage* 49, 1886–1894.

Ruchsow, M., Walter, H., Buchheim, A., Martius, P., Spitzer, M., Kachele, H., Gron, G., and Kiefer, M. (2006). Electrophysiological correlates of error processing in borderline personality disorder. *Biol. Psychol.* 72, 133–140.

Rushworth, M. F., Walton, M. E., Kennerley, S. W., and Bannerman, D. M. (2004). Action sets and decisions in the medial frontal cortex. *Trends Cogn. Sci.* 8, 410–417.

Santesso, D. L., and Segalowitz, S. J. (2009). The error-related negativity is related to risk taking and empathy in young men. *Psychophysiology* 46, 143–152.

Sato, A., Yasuda, A., Ohira, H., Miyawaki, K., Nishikawa, M., Kumano, H., and Kuboki, T. (2005). Effects of value and reward magnitude on feedback negativity and P300. *Neuroreport* 16, 407–411.

Schuermann, B., Kathmann, N., Stiglmayr, C., Renneberg, B., and Endrass, T. (2011). Impaired decision making and feedback evaluation in borderline personality disorder. *Psychol. Med.* 41, 1–11.

Toyomaki, A., and Murohashi, H. (2005). Discrepancy between feedback negativity and subjective evaluation in gambling. *Neuroreport* 16, 1865–1868.

Tversky, A., and Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science* 211, 453–458.

van der Helden, J., Boksem, M. A., and Blom, J. H. (2010). The importance of failure: feedback-related negativity predicts motor learning efficiency. *Cereb. Cortex* 20, 1596–1603.

Vogt, B. A., Berger, G. R., and Derbyshire, S. W. (2003). Structural and functional dichotomy of human midcingulate cortex. *Eur. J. Neurosci.* 18, 3134–3144.

Wu, Y., and Zhou, X. (2009). The P300 and reward valence, magnitude, and expectancy in outcome evaluation. *Brain Res.* 1286, 114–122.

Xu, Q., Shen, Q., Chen, P., Ma, Q., Sun, D., and Pan, Y. (2011). How an uncertain cue modulates subsequent monetary outcome evaluation: an ERP study. *Neurosci. Lett.* 505, 200–204.

Yeung, N., and Sanfey, A. G. (2004). Independent coding of reward magnitude and valence in the human brain. *J. Neurosci.* 24, 6258–6264.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 28 February 2012; accepted: 21 June 2012; published online: 06 July 2012.

Citation: Schuermann B, Endrass T and Kathmann N (2012) Neural correlates of feedback processing in decision-making under risk. Front. Hum. Neurosci. ***6****:204. doi: 10.3389/fnhum.2012.00204*

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 20 September 2011
doi: 10.3389/fnhum.2011.00100

All about the money – external performance monitoring is affected by monetary, but not by socially conveyed feedback cues in more antisocial individuals

Daniela Melitta Pfabigan*, Johanna Alexopoulos, Herbert Bauer, Claus Lamm and Uta Sailer

Social, Cognitive and Affective Neuroscience Unit, Faculty of Psychology, University of Vienna, Vienna, Austria

Edited by:
Patrizia Thoma, Ruhr-University Bochum, Germany

Reviewed by:
David Polezzi, Ruhr-University of Bochum, Germany
Inti Brazil, Donders Institute for Brain Cognition and Behaviour, Netherlands

****Correspondence:***
Daniela Melitta Pfabigan, Social, Cognitive and Affective Neuroscience Unit, Faculty of Psychology, University of Vienna, Liebiggasse 5, A-1010 Vienna, Austria.
e-mail: daniela.pfabigan@univie.ac.at

This study investigated the relationship between feedback processing and antisocial personality traits measured by the PSSI questionnaire (Kuhl and Kazén, 1997) in a healthy undergraduate sample. While event-related potentials [feedback related negativity (FRN), P300] were recorded, participants encountered expected and unexpected feedback during a gambling task. As recent findings suggest learning problems and deficiencies during feedback processing in clinical populations of antisocial individuals, we performed two experiments with different healthy participants in which feedback about monetary gains or losses consisted either of social–emotional (facial emotion displays) or non-social cues (numerical stimuli). Since the FRN and P300 are both sensitive to different aspects of feedback processing we hypothesized that they might help to differentiate between individuals scoring high and low on an antisocial trait measure. In line with previous evidence FRN amplitudes were enhanced after negative and after unexpected feedback stimuli. Crucially, participants scoring high on antisocial traits displayed larger FRN amplitudes than those scoring low only in response to expected and unexpected negative numerical feedback, but not in response to social–emotional feedback – irrespective of expectancy. P300 amplitudes were not modulated by antisocial traits at all, but by subjective reward probabilities. The present findings indicate that individuals scoring high on antisociality attribute higher motivational salience to monetary compared to emotional–social feedback which is reflected in FRN amplitude enhancement. Contrary to recent findings, however, no processing deficiencies concerning social–emotional feedback stimuli were apparent in those individuals. This indicates that stimulus salience is an important aspect in learning and feedback processes in individuals with antisocial traits which has potential implications for therapeutic interventions in clinical populations.

Keywords: antisocial personality, feedback processing, FRN, P300

INTRODUCTION

Individual behavior lacking consideration of others, no matter whether intentional or not, is known as antisocial behavior (Berger, 2003). The pathological manifestation of antisocial behavior is the so-called antisocial personality disorder (ASP). The DSM-IV classification (American Psychiatric Association, 1994) of ASP includes diagnostic characteristics such as lack of respect for social norms, reckless and aggressive behavior, irresponsibility, and lack of remorse and guilt (Rodrigo et al., 2010). The corresponding diagnosis of the ICD-10 (World Health Organization, 1992) classification scheme, the so-called dissocial personality disorder, adds another important diagnostic characteristic: the inability to learn from experience, in particular from punishment. Dinn and Harris (2000) suggested that these learning deficits might be triggered by an inability to effectively process negative and positive feedback stimuli. Considering a dimensional account of the distribution of personality characteristics, non-clinical manifestations of ASP symptoms should also be observable in healthy individuals (Walters, 2009). In particular, the characteristic deficits in learning from experience give rise to the question whether or not healthy individuals with antisocial tendencies process and respond to external feedback in a comparable way to healthy individuals without these behavioral tendencies.

Event-related potentials (ERPs) are a useful tool to investigate neural processes related to feedback processing, in particular as their high temporal resolution allows detecting early differences in processing between individuals. Therefore, the main objective of the present study was to investigate the relationship of individual differences in antisocial personality traits with two ERP components related to external feedback processing, the feedback related negativity (FRN) and the P300, respectively.

The FRN is a negative-going deflection over frontal electrode sites which can be determined within 200–300 ms after negative external feedback, such as the indication of an incorrect response or of monetary loss (Miltner et al., 1997; Nieuwenhuis et al., 2004; Yeung et al., 2004). The FRN is thought to be generated in or near what has been originally labeled as anterior cingulate

cortex (ACC; Miltner et al., 1997; Gehring and Willoughby, 2002; Holroyd and Coles, 2002) but according to later neuroanatomical evidence has been determined as anterior medial cingulate cortex (aMCC; Vogt, 2005). Holroyd and Coles (2002) postulated to view the FRN as a reinforcement signal induced by the mesencephalic dopamine system which is conveyed to the aMCC to optimize new action–outcome relations. Furthermore, the authors assumed that outcomes that are worse than expected would elicit the largest amplitude deflections. Another important aspect of their theory is the account of FRN amplitude back propagation after learning. The better one learns specific action–outcomes – i.e., the less unexpected these outcomes become – the smaller the FRN amplitudes get after successful acquisition of the action–outcome relation. In contrast to this reinforcement learning account, Gehring and Willoughby (2002) stated that the FRN might rather reflect the subjective negative evaluation of self-relevant information than the commission of an error *per se*. Following their hypothesis, the FRN has been proposed to reflect a neuronal signal which detects discrepancies between internal and external representations (i.e., discrepancies between subjective reward expectations and objective reward contingencies) to highlight motivationally salient outcomes (Yeung et al., 2005). This interpretation is in line with recent findings from our group using a similar experimental paradigm as in the present study (Pfabigan et al., 2011). Notably, the learning of action–outcome relations led to more predictable outcomes and decreases motivational salience of these respective outcomes. Thus, the observation of FRN amplitudes becoming smaller in amplitude after these relations have been learned (Holroyd and Coles, 2002; Nieuwenhuis et al., 2004; Sailer et al., 2010) is in line with both the reinforcement learning and the motivational salience accounts.

While the specific link between feedback processing, FRN, and antisocial traits has never been explored, a related study by Von Borries et al. (2010) attempted to establish a relationship between learning, feedback processing, and psychopathy. Psychopathy is a personality construct bearing some conceptual overlap with ASP, with comorbidity of ASP and psychopathy amounting to 30% (Hart and Hare, 1996; Coid and Ullrich, 2010). Nevertheless, ASP and psychopathy may not be equalized since ASP focuses on observable behavior whereas psychopathy emphasizes personality traits. Unfortunately, these two concepts are repeatedly mixed up in literature. Psychopathy is assessed via semi-structured questionnaires in forensic samples [Psychopathy Checklist-Revised (PCL-R); Hare, 2003] and via self-report questionnaires in healthy individuals [e.g., Psychopathic Personality Inventory-Revised (PPI-R), Lilienfeld and Andrews, 1996]. Notably, DSM-IV and ICD-10 do not incorporate all symptoms necessary for a PCL-R-based psychopathy diagnosis, thus the concept of psychopathy is not adequately represented in these diagnostic manuals (World Health Organization, 1992; Widinger, 2007). Von Borries et al. (2010) reported learning deficits as well as altered ERPs related to performance monitoring in a forensic sample scoring high on the PCL-R. The authors suggested that negative feedback cues were not adequately assessed by the psychopathic participants to adapt behavior in subsequent actions. Nevertheless, no significant group differences were reported by Von Borries et al. (2010) regarding FRN amplitudes.

The P300 is another ERP component commonly investigated during feedback processing. It is characterized by a positive deflection peaking around 300–600 ms after stimulus onset at posterior recording sites, and P300 has been shown to be sensitive to the significance and occurrence probability of a stimulus (Duncan-Johnson and Donchin, 1977; Johnson and Donchin, 1980) as well as task complexity (Israel et al., 1980). Increased P300 amplitude is thought to reflect the increased allocation of neural resources and related enhanced stimulus processing (Polich, 2007). Moreover, P300 amplitude modulation was found in decision and outcome evaluation tasks, supposedly reflecting the functional or motivational significance of the feedback stimuli (Yeung and Sanfey, 2004; Hajcak et al., 2005; Yeung et al., 2005; Luu et al., 2009). Ambiguous results have been reported regarding the relationship between P300 amplitude, psychopathy, and antisocial behavior, as different studies report both enhanced (Raine and Venables, 1988) and decreased P300 amplitudes in psychopathic and antisocial individuals (Costa et al., 2000; Bernat et al., 2007). Hicks et al. (2007) suggested that P300 amplitude reduction is in particular associated with antisocial facets of psychopathy. Several studies focused on P300 latency and antisociality, with a recent review by Gao and Raine (2009) suggesting that delayed P300 amplitudes in more antisocial individuals might reflect deficits in stimulus processing speed.

The aim of the present study was to investigate the relationship between individual differences in antisocial personality traits measured in healthy individuals with ERP correlates of external feedback processing. To the best of our knowledge, no study has investigated this relationship yet, which is surprising given the importance of antisocial behavior. Antisocial personality traits are a prevalent problem for individuals and society alike. By identifying their potential mechanisms in healthy individuals we aim to further the knowledge regarding these personality traits and associated clinically relevant manifestations. Based on the observation of FRN amplitude decrease after learning from incorrect responses (FRN amplitude back propagation; Holroyd and Coles, 2002; Nieuwenhuis et al., 2004), and the ICD-10 classification postulating deficits in learning from external cues in individuals with ASP, we expected individuals with distinctive antisocial personality traits to show larger FRN amplitudes after negative feedback than individuals lacking these personality traits. Irrespective of antisociality, we expected larger FRN amplitudes after negative compared to positive feedback (Miltner et al., 1997; Holroyd and Coles, 2002), and after unexpected compared to expected feedback (Nieuwenhuis et al., 2004; Hajcak et al., 2007; Pfabigan et al., 2011). Regarding P300 amplitudes, we expected an amplitude decrease in individuals scoring high on an antisociality measure compared to the ones scoring low (Gao and Raine, 2009) indicating inefficient allocation of neural resources during the processing of task relevant information. In general, we expected larger P300 amplitudes in response to unexpected compared to expected feedback stimuli (Duncan-Johnson and Donchin, 1977). Additionally, high-scoring participants might display delayed P300 latencies (Costa et al., 2000; Bernat et al., 2007). Furthermore, we were interested in whether the dimension of motivational salience of the feedback stimuli (i.e., social–emotional vs. non-social) would affect feedback processing in a comparable way

in both groups or discriminate them. Thus, we conducted two experiments using either numbers indicating monetary gain or loss (experiment 1) or emotional faces (experiment 2) as feedback stimuli. Human facial expressions can be considered as valuable social cues since they incorporate crucial information necessary in social exchange situations (Rolls, 2000). Consequently, antisocial behavior *per se* gives rise to the question whether or not the processing of social–emotional cues is disrupted in antisocial individuals compared to socially oriented ones. Indeed, Marsh and Blair (2008) reported deficits while recognizing fearful faces in antisocial individuals. However, since we did not include fearful facial expressions as feedback stimuli, we had no directional hypothesis regarding group differences for the dimension of the feedback stimuli.

The experimental paradigm applied was a gambling task in which participants encountered expected and unexpected positive and negative feedback outcomes.

EXPERIMENT 1: MONETARY FEEDBACK

METHODS

Participants

Initially, 31 right-handed psychology students of the University of Vienna (16 females) participated in the first study. The data of two male participants had to be excluded due to data acquisition problems. The mean age of the remaining 29 participants was 26.10 ± 3.11 years. Handedness was assessed by the Edinburgh Handedness Inventory (Oldfield, 1971). Participants had normal or corrected-to-normal vision and reported no history of neurological or psychiatric disorders. The study was conducted in accordance with the *Declaration of Helsinki* and local guidelines of the University of Vienna. Informed written consent was obtained from each participant prior to participation. At the end of the experiment participants received an individually adjusted bonus depending on their performance in the experimental task (between 10 and 25 Euros).

Prior to electroencephalogram (EEG) data collection, participants completed the PSSI questionnaire (Kuhl and Kazén, 1997). The PSSI is a self-assessment tool covering the manifestation of 14 non-pathologic personality traits related to personality disorders described in the DSM-IV and the ICD-10 diagnostic criteria. For this study, the so-called antisociality (AS) scale of the PSSI was of particular interest. Its reliability (Cronbach's $\alpha = 0.86$ – AS-scale) and validity are reported to be satisfactory (Kuhl, 2001). Raw scores were transformed into standardized T-values (mean of 50, SD of 10) for all participants. High T-values on the AS subscale, which consists of 10 items (e.g., "If people turn against me, I can wear them down."), characterize people with self-determined and inconsiderate behavior to achieve individual goals. Furthermore, individuals scoring high on the AS subscale are described to act overly self-centered, offending, and humiliating while interacting with others, and to have problems adjusting to social and legal norms. Participants scored on average with a T-value of 49.00 ± 10.95 on the AS-scale, individual T-values ranged from 34 to 72. Based on the distribution of these individual T-scores, participants were separated into three groups; approximately below, above, and within two thirds of the sample's SD. This classification scheme was chosen by the authors particularly for the present study to effectively separate more social from more antisocial individuals. Twelve participants formed the low-trait group (mean 38.33 ± 2.74, range of 34–42; seven females), six participants formed the middle group (mean 48.50 ± 4.59, range of 44–54; three females), and 11 participants constituted the high-trait group (mean 60.91 ± 4.93, range of 55–72; six females). There was no influence of sex on the individual scores on the AS-scale [independent samples t-test: $t(27) = 0.57$, $p > 0.50$]. Only the 12 low-trait ("social group") and the 11 high-trait ("antisocial group") participants were considered for data analysis to enhance the separation effect for antisocial traits. The T-values of these two groups differed significantly from each other [independent samples t-test: $t(21) = 13.74$, $p < 0.001$], indicating that our group categorization was successful.

Experimental procedures

Participants were comfortably seated 70 cm in front of a 21″ cathode ray tube monitor with a 75-Hz refresh rate (Sony GDM-F520). Stimulus presentation was controlled by a Pentium IV 3.00 GHz computer and E-Prime 2.0 software (Psychology Software Tools, Inc., Pittsburgh, PA, USA). The paradigm used was identical to that described in Pfabigan et al. (2011). The experimental session began with a training run of 48 trials where participants learned specific cue–response contingencies for a forthcoming experimental task. Each trial started with a black fixation cross on a gray screen, followed by an imperative cue consisting of a black line drawing of a simple figure (Bates et al., 2000; circle, triangle, or star, each presented 16 times during training; 10.5 cm × 10.5 cm in size). During the subsequent presentation of a black question mark, participants had to choose one of two buttons on a response pad. Feedback was provided afterward. The imperative cue remained on the screen for 500 ms; the question mark appeared immediately following the cue offset and remained on the screen until the participant responded, or 2000 ms had elapsed. Approximately 350 ms after the offset of the question mark which was triggered either by a button press or elapse of 200 ms, the feedback stimulus appeared on the screen for 700 ms. During the inter-trial-interval, the fixation cross was presented again for a randomly varied duration of 2200–2700 ms. In the training run, one of the three imperative cues was associated with 100% reward probability for button one (cue "one"), and another cue was associated with 75% reward probability for button two (cue "two"). Irrespective of button choice the third cue was not rewarded at all (cue "three"). The German word for correct (RICHTIG) was presented after rewarded choices and the one for incorrect (FALSCH) with all other choices (including failure to enter a choice with the allotted time window). The assignment of the three cues to the different reward probabilities was counterbalanced across participants. After having learnt these simple cue–response-mappings the experimental task consisting of 900 trials started. Participants were now asked to search for more complex button press response patterns on the basis of these simple cue–response-mappings (e.g., pressing button one thrice, and button two twice in five consecutive trials). This instruction was chosen to sustain participants' expectations regarding the different reward probabilities for the three cues during the whole experiment. However, unknown to the participants, no such button press response pattern existed.

Although this instruction to search for meta-rules might have induced monitoring and working memory processes during the decision phase, it was indispensable for making the occurrence of unexpected feedback plausible.

After completing 48 training trials, participants started with the first block of the experimental task (150 trials). Here, a correct choice was indicated by the central presentation of the number 15 in green color (2 cm × 1.5 cm in size), announcing a gain of 15 Eurocents. An incorrect choice was indicated by the number 15 in red color, announcing a loss of 15 Eurocents. If participants had missed the response interval they were informed about it and also lost 15 Eurocents; the respective trials were discarded from further analysis. After a block of 150 trials, participants were provided with overall performance feedback about how much money they had won. Afterward, they were instructed to search for a new button press response pattern in the next block. After three blocks, a 5-min break took place, where participants were paid with the amount of money they had already won to maintain their motivation. In contrast to the training run, participants were now provided with positive feedback in 75% of the trials where they selected the previously learned buttons for cue "one" and "two." With cue "three" participants were provided with positive feedback in 25% of these trials. This contrast between the new reward contingencies and those of the training session ensured that participants encountered trials where a gain was highly expected (cue "one"), but a loss occurred, i.e., feedback was worse than expected. Likewise, participants encountered trials where a loss was highly expected (cue "three"), but a gain occurred, i.e., feedback was better than expected (**Table 1**, for details). The data corresponding to cue "two" were not further analyzed since subjective expectation levels had not changed with this cue (75% probability for gain during the training and the experimental session). Nevertheless, cue "two" was essential in this experimental paradigm – otherwise the occurrence of unexpected feedback stimuli would not have been plausible to the participants.

The experiment ended after six blocks. Afterward, participants were asked to estimate the subjectively perceived reward frequencies of the three cues in a brief questionnaire. Finally, they were rewarded with the remaining money won in the last three blocks. Including a seed capital of 5 Euros, participants gained on average 18.58 ± 4.34 Euros. Finally, participants were debriefed about the external feedback manipulation. The whole experiment took about 70 min.

EEG acquisition and preprocessing

The EEG was recorded via 61 Ag/AgCl ring electrodes, arranged equidistantly in an elastic electrode cap (EASYCAP GmbH, Herrsching, Germany; model M10). A balanced non-cephalic sterno-vertebral reference was used (Stephenson and Gibbs, 1951). Vertical and horizontal electrooculograms (EOG) were recorded bipolarly with electrodes placed 1 cm above and below the left eye and on the outer canthi, respectively, to enable off-line eye movement artifact correction. During two pre-experimental calibration trials, participants performed vertical and horizontal eye movements. These data were used to calculate subject- and channel-specific coefficients for eye movement correction (Bauer and Lauber, 1979). Skin abrasion at each recording site (Picton and Hillyard, 1972) and degassed conductance gel ensured electrode impedances below 2 kΩ. Signals were amplified using an AC amplifier set-up with a time constant of 10 s (Ing. Kurt Zickler GmbH, Pfaffstätten, Austria). All signals were recorded within a frequency range of 0.016–125 Hz and sampled at 250 Hz for digital storage.

Off-line and prior to analysis the weighted EOG signals were subtracted from the EEG signals. Subsequently, blink coefficients were calculated using a template matching procedure and blink artifacts were also subtracted from the EEG signals (Lamm et al., 2005, for details). EEGLAB 6.03b (Delorme and Makeig, 2004) was used for further analysis. A low-pass filter (cut-off frequency 30 Hz, roll-off 6 dB per octave) was applied to the EEG data. For ERP analysis signal epochs started 200 ms before feedback onset and lasted 900 ms, with the mean of the first 200 ms serving as the baseline. Before applying extended (infomax) independent component analysis (ICA; Bell and Sejnowski, 1995; Lee et al., 1999) trials contaminated by muscular or movement artifacts were rejected based on visual inspection. ICA was performed to remove residual ocular artifacts, as described in Delorme et al. (2007), and afterward a semi-automatic artifact removal procedure was employed to eliminate epochs containing voltage values exceeding ±75 μV in any channel. Due to the experimental set-up the data sets per subject consisted of three times more expected feedback trials than unexpected feedback trials. Therefore, numbers of trials per condition were equalized per subject in order to adjust for the signal-to-noise ratio of the ERPs. For each participant, we randomly drew the same number of trials that were available for unexpected positive feedback trials out of all expected positive feedback trials (surviving artifact screening). The same procedure

Table 1 | Reward probabilities in training and experimental sessions, classification of conditions, and probability of occurrence in both studies.

	Probability of positive feedback				
Cue–response-combination	**Training (%)**	**Experiment (%)**	**Condition**	**Number of trials**	**Probability of occurrence (%)**
Cue 1 + button 1	100	75	Exp-pos	225/900	25
			Unexp-neg	75/900	8.3
Cue 2 + button 2	75	75	–		
Cue 3 + button 1/2	0	25	Unexp-pos	75/900	8.3
			Exp-neg	225/900	25

The assignment of the three visual cues to the experimental conditions was counterbalanced across participants.

was applied to the unexpected and expected negative feedback trials.

Data analysis

Artifact-free epochs were averaged separately for each subject and each of the following four conditions: (1) expected positive feedback (exp-pos; cue "one"), (2) expected negative feedback (exp-neg; cue "three"), (3) unexpected positive feedback (unexp-pos; cue "three"), and (4) unexpected negative feedback (unexp-neg; cue "one"). FRN amplitudes were assessed at electrode site FCz which was chosen based upon existing literature (Gehring and Willoughby, 2002; Holroyd and Coles, 2002) and the visual inspection of the data. The peak-to-peak voltage difference between the most negative peak between 200 and 400 ms after feedback onset (FRN) and the preceding positive peak (P2) was calculated (Holroyd et al., 2003). P300 amplitudes were obtained by searching for local positive maxima (in relation to baseline) between 300 and 600 ms after feedback onset at electrode site Pz where the P300 was most prominent. P300 latency was measured from feedback onset to the corresponding positive maximum.

Feedback related negativity amplitude differences were analyzed by means of a mixed-design $2 \times 2 \times 2$ ANOVA with the between-subjects factor *group* (low-trait, high-trait), and the within-subjects factors *expectation* (expected, unexpected) and *valence* (positive, negative). The same ANOVA model was applied to P300 peak amplitudes and P300 peak latencies. Regarding the *a priori* FRN hypothesis on group differences after negative feedback stimuli, we calculated a linear contrast. Furthermore, significant interaction effects without *a priori* hypotheses were explored with Tukey's HSD *post hoc* test To demonstrate the effect size of the ANOVA results, partial eta-squared (η_p^2) is reported (Cohen, 1973). All statistical analyses were performed using SPSS 15 (IBM SPSS Statistics 15, Somer, NY, USA).

EXPERIMENT 2: EMOTIONAL FACES FEEDBACK

METHODS

Participants

Initially, 28 right-handed female psychology students of the University of Vienna participated in the second experiment. We included only women in the second study since no gender differences emerged in experiment 1 and because of easier participant recruitment. The data of two participants had to be excluded from further analysis due to data acquisition artifacts. The mean age of the remaining 26 participants was 23.38 ± 3.41 years. The study was conducted in accordance with the *Declaration of Helsinki* and local guidelines of the University of Vienna. Informed written consent was obtained from each participant prior to participation. At the end of the experiment each participant received a fixed bonus of 15 Euros for participation.

Again, the PSSI questionnaire was administered before EEG data collection. The average score on the AS-scale in this sample was 49.69 ± 10.65, ranging from 31 to 72. Participants were divided into three groups based on whether their T-values lay approximately below, above, or within two thirds of the samples' SD (comparable to experiment 1). Ten participants formed the low-trait group (mean 39.20 ± 3.68, range of 31–42), six participants the middle group (mean 49.00 ± 3.69, range of 45–54), and the remaining 10 participants constituted the high-trait group (mean 60.60 ± 6.26, range of 56–72). Only the 10 low-trait ("social group") and the 10 high-trait ("antisocial group") participants were considered for analysis. The T-values of these two groups differed significantly from each other [independent samples t-test: $t(18) = 9.33$, $p < 0.001$], again indicating that our group categorization was successful. No differences of the individual AS-scale scores were observed when comparing both experiments either [independent samples t-test; $t(53) = 0.24$, $p > 0.80$] although experiment 2 comprised only female participants.

Experimental procedures

Experimental procedures were equivalent to those in experiment 1, with the only exception that participants were presented with emotional faces instead of colored numbers depicting positive and negative feedback. In particular, feedback stimuli consisted of pictures of faces with emotional expressions taken from the standardized Ekman series (Ekman and Friesen, 1976; 4 cm × 5 cm in size). Two male and two female faces showing the emotions "happiness" and "anger" were used as positive ("happy" face) and negative ("angry" face) feedback stimuli, with poser gender balanced across experimental trials. Participants were familiarized with the emotional faces during task instruction. Participants were informed that they could earn 10–15 Euros depending on their task performance, i.e., the number of correct responses. After each of the six experimental task blocks participants were given an overall performance feedback in terms of the number of correct responses. Afterward they were informed that they had performed extremely well – and regardless of their points accumulated – all were paid 15 Euros. Finally, they were debriefed about the external feedback manipulation.

EEG acquisition and preprocessing

Data acquisition and preprocessing procedures were identical to experiment 1. Data were recorded from 61 Ag/AgCl ring electrodes. The same eye movement and blink correction algorithms were applied as described in experiment 1.

Data analysis

Subject- and condition-wise averages were calculated for the four conditions (1) expected positive feedback (exp-pos; cue "one"), (2) expected negative feedback (exp-neg; cue "three"), (3) unexpected positive feedback (unexp-pos; cue "three"), and (4) unexpected negative feedback (unexp-neg; cue "one"). Subsequently, FRN and P300 peaks were extracted using the same criteria as in experiment 1. For FRN analysis, data were subjected to a mixed-design $2 \times 2 \times 2$ ANOVA with the between-subjects factor *group* (low-trait, high-trait), and the within-subjects factors *expectation* (expected, unexpected) and *valence* (positive, negative). Regarding the *a priori* FRN hypothesis on group differences after negative feedback stimuli, we calculated a linear contrast. The same ANOVA model was applied for P300 peak and latency analysis.

To address the question whether or not FRN amplitude differences between low-scoring and high-scoring participants in experiment 1 differed significantly from those in experiment 2, we compared the Cohen's d effect sizes of the between-subject factor *group* of both experiments by means of a homogeneity test based

on the Q statistic with CMA v2.2.030 software (Comprehensive Meta-Analysis; Biostat™, Englewood, USA).

RESULTS EXPERIMENT 1: MONETARY FEEDBACK

BEHAVIORAL RESULTS

Participants learned the cue–response contingencies in the training session. Button one was chosen in 90.43% of cue "one" trials, and button two in 79.95% of cue "two" trials. No button preference emerged for cue "three" trials (42.24% button one vs. 49.92% button two).

In the post-experimental questionnaire, participants estimated the probability of occurrence of positive feedback after cue "one" with a median of 70, range of 50–90, after cue "two" with a median of 70, range of 20–85, and after cue "three" with a median of 20, range of 1–70. Positive feedback was expected significantly more often after cue "one" than cue "three" (Wilcoxon signed-ranks test: $Z = -4.79$, $p < 0.001$).

ERP RESULTS

Figure 1 displays feedback-locked average ERPs for expected and unexpected, positive and negative feedback conditions for the low-trait and the high-trait group at electrode site FCz of experiment 1.

Regarding FRN amplitudes, analysis revealed main effects for *expectation* [$F(1,21) = 8.94$, $p = 0.007$, $\eta_p^2 = 0.30$], *valence* [$F(1,21) = 26.38$, $p < 0.001$, $\eta_p^2 = 0.56$], and *group* [$F(1,21) = 7.20$, $p = 0.014$, $\eta_p^2 = 0.26$]. FRN amplitudes were more pronounced after unexpected compared to expected feedback, and after negative compared to positive feedback. Furthermore, high-scoring participants displayed enhanced FRN amplitudes compared to low-scoring ones. A linear contrast testing the *a priori* hypothesis of group differences regarding negative feedback stimuli revealed significantly larger FRN amplitudes after negative feedback in the high-trait group compared to the low-trait group ($p = 0.015$). No significant interaction effects emerged (all $ps > 0.123$).

Regarding P300 amplitudes, analysis revealed a main effect for the factor *expectation* [$F(1,21) = 65.37$, $p < 0.001$, $\eta_p^2 = 0.76$], indicating that P300 amplitudes were largest after unexpected compared to expected feedback. The *expectation x valence* interaction showed a trend toward significance [$F(1,21) = 3.80$, $p = 0.065$, $\eta_p^2 = 0.15$], thereby pointing toward largest P300 amplitudes after unexpected positive feedback. Regarding P300 latency, analysis revealed main effects for *expectation* [$F(1,21) = 22.85$, $p < 0.001$, $\eta_p^2 = 0.52$] and *valence* [$F(1,21) = 13.84$, $p = 0.001$, $\eta_p^2 = 0.40$], but no interaction effects (all $ps > 0.151$). P300 latencies were prolonged after unexpected as well as negative feedback stimuli. No effects of group emerged for P300 amplitude ($p > 0.758$) and latency analyses ($p > 0.881$).

RESULTS EXPERIMENT 2: EMOTIONAL FACES FEEDBACK

BEHAVIORAL RESULTS

Again, participants learned the cue–response contingencies in the training session. Button one was chosen in 90.60% of cue "one" trials, and button two in 77.34% of cue "two" trials. No button preference emerged for cue "three"-trials (44.53% button one vs. 54.68% button two).

In the post-experimental questionnaire, participants estimated the probability of occurrence of positive feedback after cue "one" with a median of 70, range of 60–90, after cue "two" with a median of 70, range of 50–80, and after cue "three" with a median of 30, range 2–40. Again, positive feedback was expected significantly more often after cue "one" than cue "three" (Wilcoxon signed-ranks test: $Z = -4.48$, $p < 0.001$).

ERP RESULTS

Figure 2 displays feedback-locked average ERPs for expected and unexpected, positive and negative feedback conditions for the low-trait and the high-trait group at electrode site FCz for experiment 2.

Regarding FRN amplitudes, statistical analysis revealed main effects for *expectation* [$F(1,18) = 6.93$, $p = 0.017$, $\eta_p^2 = 0.28$] and *valence* [$F(1,18) = 13.53$, $p = 0.002$, $\eta_p^2 = 0.43$]. FRN amplitudes were larger after unexpected compared to expected feedback, as well as after negative compared to positive feedback. The linear contrast testing the *a priori* hypothesis of group differences regarding negative feedback stimuli revealed no amplitude differences between the low-trait and the high-trait group ($p > 0.641$). Otherwise, no significant main effect for *group* ($p > 0.877$), nor any significant interaction effects were observed (all $ps > 0.205$).

For P300 amplitudes, we observed a main effect of *expectation* [$F(1,18) = 27.49$, $p < 0.001$, $\eta_p^2 = 0.60$], and a significant interaction of *expectation x valence* [$F(1,18) = 17.56$, $p = 0.001$, $\eta_p^2 = 0.49$]. Significantly smaller P300 amplitudes were found after expected positive compared to the remaining three feedback conditions (all $ps < 0.012$), and after expected negative compared to unexpected positive feedback ($p < 0.004$). No expectation effect was present for negative feedback conditions ($p > 0.383$), and no valence effect was present for unexpected feedback conditions ($p > 0.115$). Regarding P300 latency, the ANOVA revealed main effects of *expectation* [$F(1,18) = 4.93$, $p = 0.039$, $\eta_p^2 = 0.22$] and *valence* [$F(1,18) = 11.25$, $p = 0.004$, $\eta_p^2 = 0.39$], but no interaction effects (all $ps > 0.109$). P300 latencies were longer after unexpected as well as negative feedback stimuli. No effects of *group* emerged for the P300 amplitude ($p > 0.826$) and latency analyses ($p > 0.403$). P300 peak amplitudes and latencies of both experiments are depicted in **Table 2**.

COMPARISON OF EXPERIMENT 1 AND 2

The Cohen's *d* effect size for the *group* factor was −1.502 in experiment 1, and −0.147 in experiment 2. The comparison of both effect sizes by means of a homogeneity test corroborated our previous findings. FRN amplitude differences between low-scoring and high-scoring individuals were only significantly different from each other when monetary feedback was provided ($\chi^2_{(1)} = 8.68$, $p = 0.003$).

DISCUSSION

The main objective of the present study was to investigate neuronal correlates of feedback processing in healthy individuals scoring high or low on an antisociality measure by applying a gambling task with two different types of feedback stimuli. No group differences were observed between low-scoring and high-scoring participants when administering emotional faces as feedback stimuli.

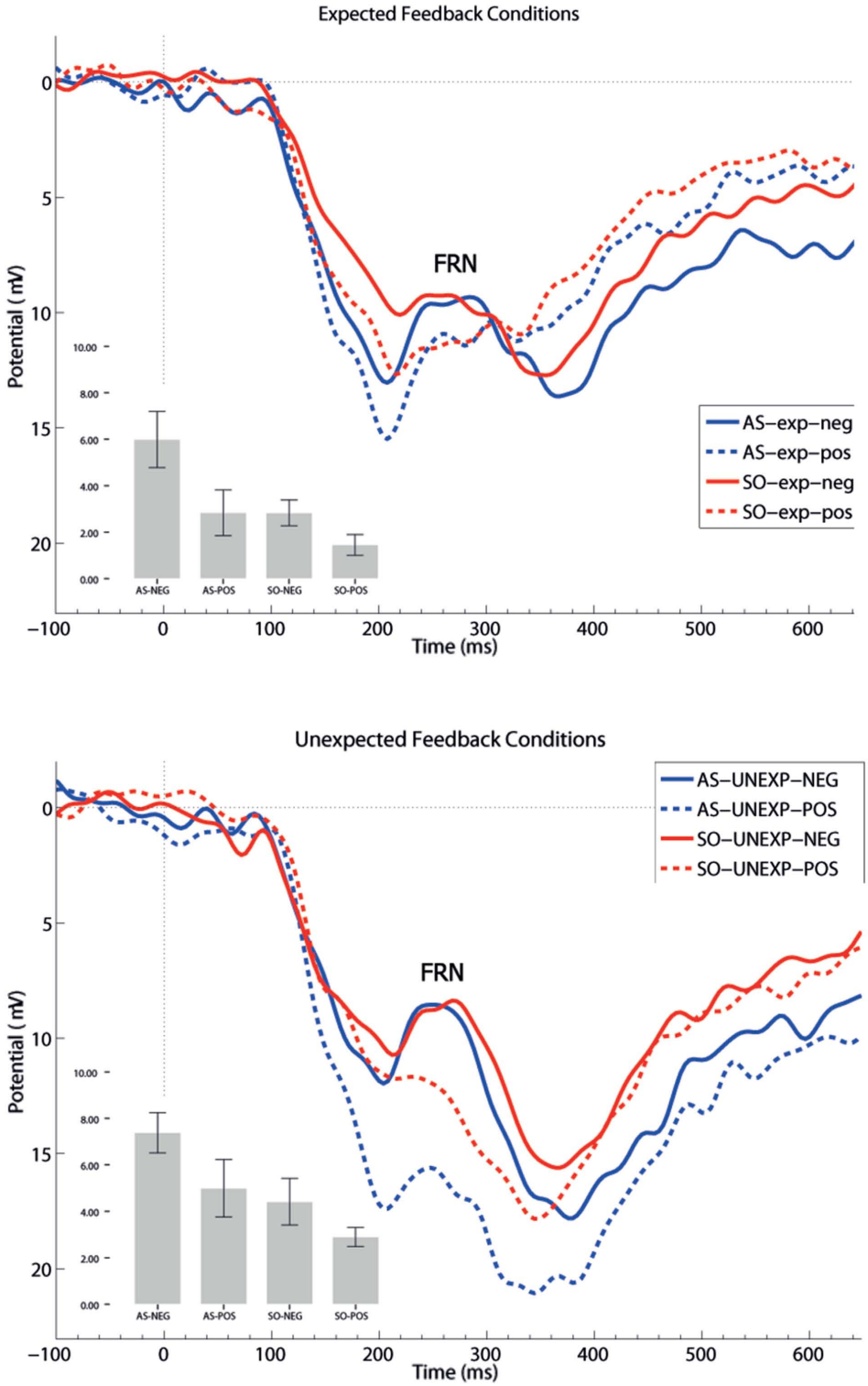


FIGURE 1 | Grand average ERPs of experiment 1. Grand averages at electrode sites FCz for expected (upper panel) and unexpected (lower panel) positive (POS) and negative (NEG) feedback conditions differentiating low-trait (SO) and high-trait (AS) participants for experiment 1. Negative is drawn upward per convention; feedback presentation started at 0 ms. The bar chart depicts the respective peak-to-peak mean FRN amplitude values. Error bars indicate 1 SE.

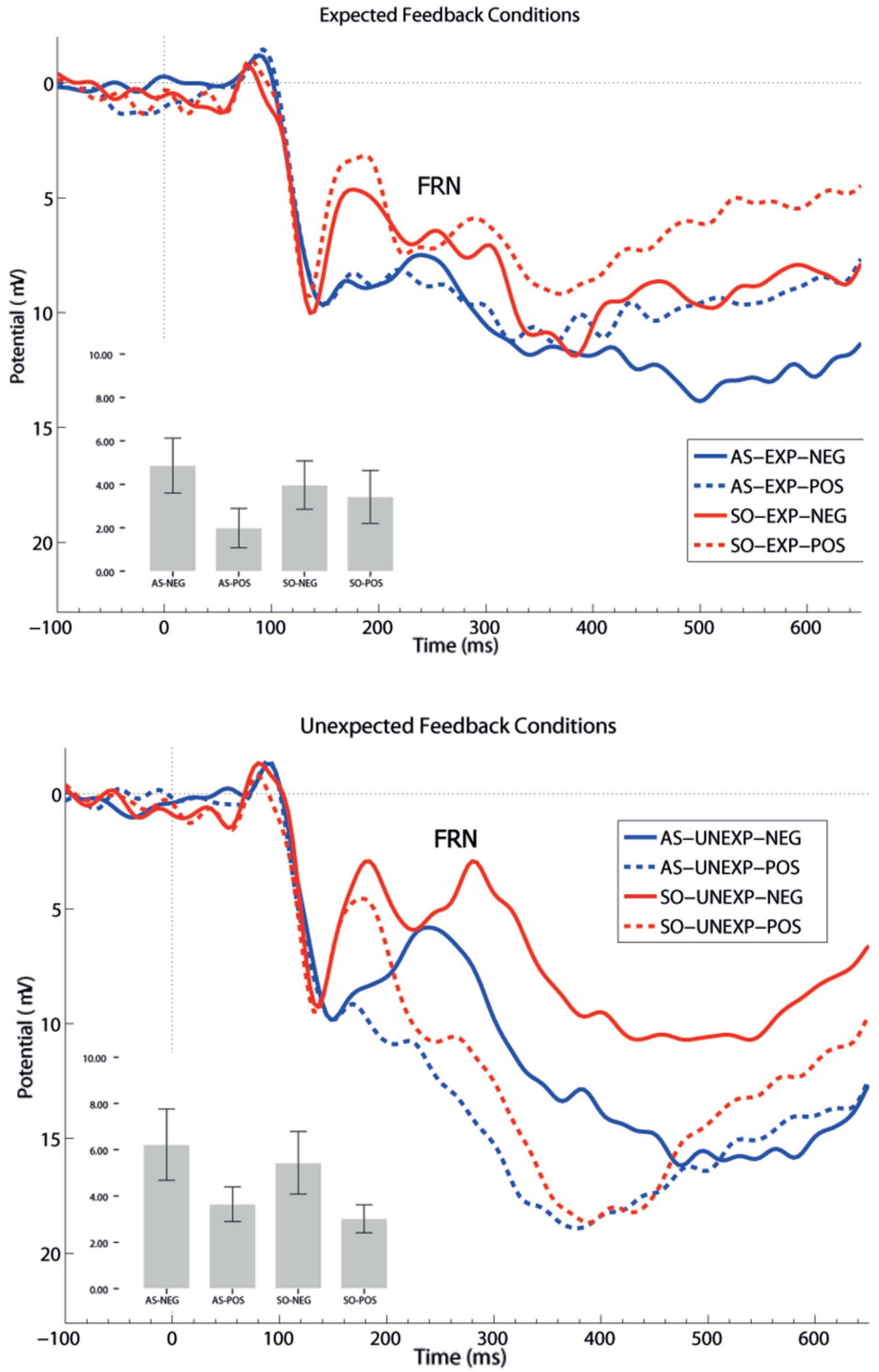


FIGURE 2 | Grand average ERPs of experiment 2. Grand averages at electrode sites FCz for expected (upper panel) and unexpected (lower panel) positive (POS) and negative (NEG) feedback conditions differentiating low-trait (SO) and high-trait (AS) participants for experiment 2. Negative is drawn upward per convention; feedback presentation started at 0 ms. The bar chart depicts the respective peak-to-peak mean FRN amplitude values. Error bars indicate 1 SE.

However, when administering numbers directly indicating monetary gain or loss as feedback stimuli, the high-trait group displayed enhanced FRN amplitudes compared to the low-trait group. In particular, FRN amplitudes were larger after expected and unexpected negative feedback indicating monetary loss in high-scoring individuals, but not in low-scoring ones. This is the main finding

Table 2 | Mean base-to-peak amplitude values and mean latencies and corresponding SE values of the P300 at Pz for the high-trait group and the low-trait group for experiment 1 (money FB) and experiment 2 (facial FB).

	High-trait group				Low-trait group			
	Mean amplitudes	**SE**	**Mean latency**	**SE**	**Mean amplitudes**	**SE**	**Mean latency**	**SE**
MONEY FB								
Exp-pos	15.65	2.42	391	26.92	15.52	1.43	395	35.72
Exp-neg	15.94	2.12	463	36.04	16.03	1.53	445	41.21
Unexp-pos	23.18	2.92	441	39.96	21.64	2.07	463	33.22
Unexp-neg	20.15	1.85	476	32.47	18.59	1.31	498	41.27
Facial FB								
Exp-pos	16.33	1.55	415	31.54	16.32	1.37	392	19.31
Exp-neg	20.20	1.83	454	26.41	19.68	2.07	435	23.27
Unexp-pos	23.33	2.09	422	26.32	24.63	2.59	435	14.30
Unexp-neg	23.15	2.52	502	26.44	19.96	2.60	434	23.55

of the present experiment. To be more explicit, it is the monetary gambling task which successfully discriminates individuals based on their scores on the PSSI antisociality scale. P300 amplitudes and latencies were not affected by antisociality.

Feedback related negativity enhancement after negative compared to positive feedback stimuli, and after unexpected compared to expected feedback stimuli is in line with recent literature (Miltner et al., 1997; Gehring and Willoughby, 2002; Holroyd and Coles, 2002). A general FRN amplitude enhancement can be interpreted as error signal (Miltner et al., 1997), as response conflict signal (Botvinick et al., 2001), or as indicator for outcomes worse than expected (Holroyd and Coles, 2002). The response conflict account would imply that high-scoring participants experienced more cognitive conflict after monetary feedback presentation, no matter whether they won or lost money and irrespective of expectancy. On the contrary, the error signal account would imply that each possible feedback outcome was worse than expected for the high-scoring participants. However, both assumptions do not seem plausible since neither susceptibility for enhanced cognitive conflict nor a general negativity bias have been reported in antisocials. Further theories on FRN modulation emphasize subjective stimulus evaluation and motivational significance of the depicted stimuli (Gehring and Willoughby, 2002; Yeung et al., 2005). Both accounts would indicate that the feedback stimuli comprising in particular monetary losses, but not the feedback stimuli comprising emotional faces, were more salient to high-scoring than to low-scoring participants. Thus, one might argue that the high-scoring individuals process emotional faces adequately and comparable to the low-scoring individuals. It is most likely that the additional monetary incentive triggers an increase in motivational salience which then yields to neuronal processing differences between the two groups.

To our knowledge, no other study has investigated the relationship between antisociality in a community sample and feedback processing yet. Therefore, we will discuss studies investigating constructs related to antisociality and ASP, namely psychopathy and the concept of externalizing psychopathology.

Regarding psychopathy, Von Borries et al. (2010) conducted a feedback processing study in a forensic sample. The authors applied a probabilistic gambling task to incarcerated psychopathic violent offenders and to a free and healthy control group. Von Borries et al. (2010) found reduced error-related activity in the psychopathic group which was discussed as being an indicator for a disability in forming an internal template of the presented learning rule. This observation of reduced neuronal activity regarding internal performance monitoring in psychopathy has already been reported. Brazil et al. (2009) observed psychopaths to show deficits in later stages of error processing and subsequent behavioral adaption. Munro et al. (2007) reported diminished error-related brain activity during a face flanker task compared to a letter flanker task. External performance monitoring (as indexed by learning rate and FRN amplitudes) was statistically comparable in psychopaths and healthy controls in the study of Von Borries et al. (2010). This non-significant result might be explainable by the fact that the authors did not distinguish between different aspects of psychopathy. Recently, dual deficit models propose two sub dimensions of psychopathy (Lykken, 1995; Fowles and Dindo, 2009). The so-called primary psychopathy is associated with deficits in emotional and interpersonal domains (Hare, 2003). Contrary, the so-called secondary psychopathy is associated with impulsive and antisocial behavior (Hare, 2003). Unfortunately, Von Borries et al. (2010) as well as studies investigating neural correlates of error processing did not distinguish between these two facets of psychopathy. Thus, the lack of group differences regarding feedback processing might be attributed to participants rather scoring high on primary than on secondary psychopathy, which is not considered to be associated with antisocial behavior (Hare, 2003).

The concept of externalizing psychopathology describes another personality facet and is considered to reflect an underlying vulnerability factor for impulse control deficits which can be found in conduct disorder, substance-use disorders, and adult ASP (Krueger, 1999; Krueger et al., 2001). Furthermore, personality dimensions such as aggression and impulsivity were suggested to be basic markers of an externalizing vulnerability (Krueger et al., 2001). To cross-reference proneness to externalizing and

psychopathy, Patrick (2007) suggested a relation of secondary psychopathy with externalizing psychopathology. Apart from the incorporation of antisocial behavior, externalizing psychopathology shares another relevant theoretical assumption with ASP, namely a failure to learn from experience, which can be found in the components of externalizing psychopathology (Hall et al., 2007). For example, reduction in neuronal activity has been reported in relation to internal performance monitoring in highly impulsive individuals (Pailing et al., 2002; Potts et al., 2006), as well as in individuals scoring high on the externalizing construct (Hall et al., 2007). Recently, Bernat et al. (2011) investigated the neuronal correlates of gain/loss feedback and externalizing psychopathology. Applying time frequency decomposition measures, the authors found no relation between FRN time frequency measures and proneness to externalizing. Thus, Bernat et al. (2011) assumed that performance monitoring deficits in highly externalizing individuals were limited to internal performance monitoring processes, whereas external performance monitoring processes reflected by FRN amplitudes were not affected. Their results are partly in line with data on psychopathy (Von Borries et al., 2010) and with the results of our second experiment where no group differences due to antisociality were observed. However, Bernat et al. (2011) did not reward their participants based on their task performance in comparison to our first experiment; although they also presented their participants with numbers indicating gain or loss. The monetary reward cues of our first experiment might hold responsible for the difference between the results of Bernat et al. (2011) and ours. Indeed, Bernat et al. (2011) stated that individuals prone to externalizing might be extremely sensitive to immediate and concrete reward, whereas abstract and more symbolic reward cues like the ones used in their study might have decreased individual reactivity to rewards. This might be also applicable to healthy antisocial individuals. Thus, the data of our first experiment point toward the assumption that the prospect of tangible monetary reward or loss is crucial when investigating external performance monitoring and its association with antisociality.

P300 enhancement after unexpected compared to expected feedback stimuli is in line with previous findings (Duncan-Johnson and Donchin, 1977; Johnson and Donchin, 1980). P300 amplitude enhancement can be interpreted as indicator of subjective reward probability. The observation of slightly larger P300 amplitudes after unexpected positive feedback might index subjective stimulus salience in all participants (Hajcak et al., 2005; Yeung et al., 2005). P300 latencies were prolonged after unexpected compared to expected, and after negative compared to positive feedback stimuli in both experiments. This observation might indicate that these feedback stimuli were more difficult to classify as the expected and positively valenced ones (Polich, 2007). The healthy student sample might be the main reason why the present data did not yield any group differences between low-scoring and high-scoring individuals regarding P300 amplitudes and latencies. It is possible that the proposed P300 decrement in antisocials (Gao and Raine, 2009) is only observable in clinical populations with known resource allocation or attention deficits. Our results indicate that both groups allocated a comparable amount of cognitive resources to the processing of the feedback stimuli and that they experienced subjective reward probability alike.

The rather small sample size of the low- and high-trait groups in both experiments poses a limitation of the present study. Thus, the present results have to be considered preliminarily. Although no gender differences were apparent for FRN or P300 analysis in the first experiment (all Fs < 1), we are aware that prevalence rates of ASP are typically higher in men than in women (Grant et al., 2004). Consequently, future studies should emphasize homogeneous samples of participants. Furthermore, future studies should investigate external feedback processing in healthy antisocial controls in relation to individuals suffering from ASP, primary, and secondary psychopathy.

To summarize, the present findings indicate that only individuals scoring high on antisocial traits attribute higher motivational salience to concrete and monetary compared to more abstract and social reinforcers. This is reflected in FRN amplitude enhancement after expected and unexpected negative feedback. No processing deficiencies concerning emotional feedback stimuli were apparent in those individuals in our study since comparable neuronal responses were observed in both participating groups. Thus, we propose to consider stimulus salience as an important aspect in feedback processes in individuals scoring high on antisocial traits. Since feedback processing is essential in learning processes, potential implications for therapeutic interventions in these individuals arise. Antisocials might profit the most from therapeutic programs including concrete and economically valid reinforcers.

ACKNOWLEDGMENTS

Participants' financial remuneration was funded by a scholarship of the University of Vienna to Daniela Melitta Pfabigan (Förderstipendium StudFG). We thank Michael Zeiler for his assistance concerning the CMA effect size analysis.

REFERENCES

American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders – DSM-IV*. Washington, DC: American Psychiatric Press.

Bates, E., Federmeier, K., Herron, D., Iyer, G., Jacobsen, T., Pechmann, T., D'Amico, S., Devescovi, A., Wicha, N., Orozco-Figueroa, A., Kohnert, K., Gutierrez, G., Lu, C. C., Hung, D., Hsu, J., Tzeng, O., Andonova, E., Gerdjikova, I., Mehotcheva, T., Székely, A., and Pléh, C. (2000). Introducing the CRL International Picture-Naming Project (CRL-JPNP). *Center Res. Lang. Newslett.* 12, 1–14.

Bauer, H., and Lauber, W. (1979). Operant conditioning of brain steady potential shifts in man. *Biofeedback Self Regul.* 4, 145–154.

Bell, A. J., and Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.* 7, 1129–1159.

Berger, K. S. (2003). *The Developing Person Through Childhood and Adolescence*. New York, NY: Worth Publishers.

Bernat, E. M., Hall, J. R., Steffen, B. V., and Patrick, C. J. (2007). Violent offending predicts P300 amplitude. *Int. J. Psychophysiol.* 66, 161–167.

Bernat, E. M., Nelson, L. D., Steele, V. R., Gehring, W. J., and Patrick, C. J. (2011). Externalizing psychopathology and gain-loss feedback in a simulated gambling task: dissociable components of brain response revealed by time-frequency analysis. *J. Abnorm. Psychol.* 120, 352–364.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C S., and Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychol. Rev.* 108, 624–652.

Brazil, I. A., de Bruijn, E. R. A., Bulten, B. H., von Borries, A. K. L.,

van Lankveld, J. J. D. M., Buitelaar, J. K., and Verkes, R. J. (2009). Early and late components of error monitoring in violent offenders with psychopathy. *Biol. Psychiatry* 65, 137–143.

Cohen, J. (1973). Eta-squared and partial eta-squared in fixed factor ANOVA designs. *Edu. Psychol. Measurements* 33, 107–111.

Coid, J., and Ullrich, S. (2010). Antisocial personality disorder is on a continuum with psychopathy. *Compr. Psychiatry* 51, 426–433.

Costa, L., Bauer, L., Kuperman, S., Porjesz, B., O'Conner, S., Hesselbrock, V., Rohrbaugh, J., and Begleiter, H. (2000). Frontal P300 decrements, alcohol dependence, and antisocial personality disorder. *Biol. Psychiatry* 47, 1064–1071.

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21.

Delorme, A., Sejnowski, T., and Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *Neuroimage* 34 1443–1449.

Dinn, W. M., and Harris, C. L. (2000). Neurocognitive function in antisocial personality disorder. *Psychiatry Res.* 97, 173–190.

Duncan-Johnson, C. C., and Donchin, E. (1977). On quantifying surprise: the variation of event-related potentials with subjective probability. *Psychophysiology* 14, 456–467.

Ekman, P., and Friesen, W. V. (1976). *Pictures of Facial Affect*. Palo Alto, CA: Consulting Psychologists Press.

Fowles, D. C., and Dindo, L. (2009). Temperament and psychopathy. A dual-pathway model. *Curr. Opin. Psychol. Sci.* 18, 179–183.

Gao, Y., and Raine, A. (2009). P3 event-related potential impairments in antisocial and psychopathic individuals: a meta-analysis. *Biol. Psychol.* 82, 199–210.

Gehring, W. J., and Willoughby, A. R. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. *Science* 14, 385–602.

Grant, B. F., Hasin, D. S., Stinson, F. S., Dawson, D. A., Chou, S. P., Ruan, W. J., and Pickering, R. P. (2004). Prevalence, correlates, and disability of personality disorders in the united states: results from the national epidemiologic survey on alcohol and related conditions. *J. Clin. Psychiatry* 65, 948–958.

Hajcak, G., Holroyd, C. B., Moser, J. S., and Simons, R. F. (2005). Brain potentials associated with expected and unexpected good and bad outcomes. *Psychophysiology* 42, 161–170.

Hajcak, G., Moser, J. S., Holroyd, C. B., and Simons, R. F. (2007). It's worse than you thought: the feedback negativity and violations of reward prediction in gambling tasks. *Psychophysiology* 44, 905–912.

Hall, J. R., Bernat, E. M., and Patrick, J. C. (2007). Externalizing psychopathology and the error-related-negativity. *Psychol. Sci.* 18, 326–333.

Hare, R. D. (2003). *The Hare Psychopathy Checklist-Revised*, 2nd Edn. Toronto: Multi-Health Systems.

Hart, S. D., and Hare, R. D. (1996). Psychopathy and antisocial personality disorder. *Curr. Opin. Psychiatry* 9, 129–132.

Hicks, B. M., Bernat, E., Malone, S. M., Iacono, W. G., Patrick, C. J., Krueger, R. F., and McGue, M. (2007). Genes mediate the association between P3 amplitude and externalizing disorders. *Psychophysiology* 44, 98–105.

Holroyd, C. B., and Coles, M. G. H. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychol. Rev.* 109, 679–709.

Holroyd, C. B., Nieuwenhuis, S., Yeung, N., and Cohen, J. D. (2003). Errors in reward prediction are reflected in the event-related brain potential. *Neuroreport* 14, 2481–2484.

Israel, J. B., Chesney, G. L., Wickens, C. D., and Donchin, E. (1980). P300 and tracking difficulty: evidence for multiple resources in dual-task performance. *Psychophysiology* 27, 259–273.

Johnson, R., and Donchin, E. (1980). P300 and stimulus categorization: two plus one is not so different from one plus one. *Psychophysiology* 17, 167–178.

Krueger, R. F. (1999). The structure of common mental disorders. *Arch. Gen. Psychiatry* 56, 921–926.

Krueger, R. F., McGue, M., and Iacono, W. G. (2001). The higher-order structure of common DSM mental disorders: internalization, externalization, and their connections to personality. *Pers. Individ. Diff.* 30, 1245–1259.

Kuhl, J. (2001). *Motivation und Persönlichkeit. Interaktionen psychischer Systeme*. Göttingen: Hogrefe.

Kuhl, J., and Kazén, M. (1997). *Persönlichkeits-Stil und Störungs-Inventar (PSSI)*. Göttingen: Hogrefe Verlag.

Lamm, C., Fischmeister, F. P., and Bauer, H. (2005). Individual differences in brain activity during visuo-spatial processing assessed by slow cortical potentials and low resolution electromagnetic tomography. *Brain Res. Cogn. Brain Res.* 25, 900–912.

Lee, T. W., Girolami, M., and Sejnowski, T. J. (1999). Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural Comput.* 11, 417–441.

Lilienfeld, S. O., and Andrews, B. P. (1996). Development and preliminary validation of a self report measure of psychopathic personality traits in noncriminal populations. *J. Pers. Assess.* 66, 488–524.

Luu, P., Shane, M., Pratt, N. L., and Tucker, D. M. (2009). Corticolimbic mechanisms in the control of trial and error learning. *Brain Res.* 1247, 100–113.

Lykken, D. T. (1995). *The Antisocial Personalities*. Mahaw: L. Erlbaum.

Marsh, A. A., and Blair, R. J. R. (2008). Deficits in facial affect recognition among antisocial populations: a meta-analysis. *Neurosci. Biobehav. Rev.* 32, 454–465.

Miltner, W. H. R., Braun, C. H., and Coles, M. G. H. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task: evidence for a "generic" neural system for error detection. *J. Cogn. Neurosci.* 9, 788–798.

Munro, G. E. S., Dywan, J., Harris, G. T., McKee, S., Unsal, A., and Segalowitz, S. J. (2007). ERN varies with degree of psychopathy in an emotion discrimination task. *Biol. Psychol.* 76, 31–42.

Nieuwenhuis, S., Holroyd, C. B., Mol, N., and Coles, M. G. H. (2004). Reinforcement-related brain potentials from medial frontal cortex: origins and functional significance. *Neurosci. Biobehav. Rev.* 28, 441–448.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113.

Pailing, P. E., Segalowitz, S. J., Dywan, J., and Davies, P. L. (2002). Error negativity and response control. *Psychophysiology* 39, 198–206.

Patrick, C. J. (2007). "Antisocial personality disorder and psychopathy," in *Handbook of Personality Disorders,* eds W. O'Donohue, K. A. Fowler, and S. O. Lilienfeld (New York: Sage), 109–166.

Pfabigan, D. M., Alexopoulos, J., Bauer, H., and Sailer, U. (2011). Manipulation of feedback expectancy and valence induces negative and positive reward prediction error signals manifest in event-related brain potentials. *Psychophysiology* 48, 656–664.

Picton, T. W., and Hillyard, S. A. (1972). Cephalic skin potentials in electroencephalography. *Electroencephalogr. Clin. Neurophysiol.* 33, 419–424.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol.* 118, 2128–2148.

Potts, G. F., Martin, L. E., Burton, P., and Montague, P. R. (2006). When things are better or worse than expected: the medial frontal cortex and the allocation of processing resources. *J. Cogn. Neurosci.* 18, 1112–1119.

Raine, A., and Venables, P. H. (1988). Enhanced P3 evoked potentials and longer P3 recovery times in psychopaths. *Psychophysiology* 25, 30–38.

Rodrigo, C., Pajapakse, S., and Jayananda, G. (2010). The "antisocial" person: an insight in to biology, classification and current evidence on treatment. *Ann. Gen. Psychiatry* 9, 1–31.

Rolls, E. T. (2000). The orbitofrontal cortex and reward. *Cereb. Cortex* 10, 284–294.

Sailer, U., Fischmeister, F. P., and Bauer, H. (2010). Effects of learning on feedback-related brain potentials in a decision-making task. *Brain Res.* 1342, 85–93.

Stephenson, W., and Gibbs, F. A. (1951). A balanced non-cephalic reference electrode. *Electroencephalogr. Clin. Neurophysiol.* 3, 237–240.

Vogt, B. A. (2005). Pain and emotion. Interactions in subregions of the cingulate gyrus. *Nat. Rev. Neurosci.* 6, 533–544.

Von Borries, A. K. L., Brazil, I. A., Bulten, B. H., Buitelaar, J. K., Verkes, R. J., and de Bruijn, E. R. A. (2010). Neural correlates of error-related learning deficits in individuals with psychopathy. *Psychol. Med.* 40, 1559–1568.

Walters, G. D. (2009). Latent structure of a two-dimensional model of antisocial personality disorder: construct validation and taxometric analysis. *J. Pers. Disord.* 23, 647–660.

Widinger, T. A. (2007). "Psychopathy and DSM-IV psychopathology," in *Handbook of Psychopathy*, ed. C. J. Patrick (New York, NY: Guilford Press), 156–171.

World Health Organization. (1992). *The ICD-10 Classification of Mental and Behavioural Disorders*. Geneva: World Health Organization.

Yeung, N., Botvinick, M. M., and Cohen, J. D. (2004). The neural basis of error detection: conflict monitoring and the error-related negativity. *Psychol. Rev.* 111, 931–959.

Yeung, N., Holroyd, C. B., and Cohen, J. D. (2005). ERP correlates of feedback and reward processing in the presence and absence of response choice. *Cereb. Cortex* 15, 535–544.

Yeung, N., and Sanfey, A. G. (2004). Independent coding of reward magnitude and valence in the human brain. *J. Neurosci.* 24, 6258–6264.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 04 July 2011; accepted: 30 August 2011; published online: 20 September 2011.

Citation: Pfabigan DM, Alexopoulos J, Bauer H, Lamm C and Sailer U (2011) All about the money – external performance monitoring is affected by monetary, but not by socially conveyed feedback cues in more antisocial individuals. Front. Hum. Neurosci. ***5****:100. doi: 10.3389/fnhum.2011.00100*

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 28 October 2011
doi: 10.3389/fnhum.2011.00123

Dissociation of response and feedback negativity in schizophrenia: electrophysiological and computational evidence for a deficit in the representation of value

Sarah E. Morris[1,2]*, Clay B. Holroyd[3], Monica C. Mann-Wrobel[1†] and James M. Gold[4]

[1] *VISN 5 Mental Illness Research, Education, and Clinical Center, Baltimore, MD, USA*
[2] *Department of Psychiatry, University of Maryland School of Medicine, Baltimore, MD, USA*
[3] *Department of Psychology, University of Victoria, Victoria, BC, Canada*
[4] *Maryland Psychiatric Research Center, University of Maryland School of Medicine, Catonsville, MD, USA*

Edited by:
Patrizia Thoma, Ruhr-University Bochum, Germany

Reviewed by:
Martin Brüne, Ruhr University Bochum, Germany
Christian Bellebaum, Ruhr University Bochum, Germany

****Correspondence:***
Sarah E. Morris, Division of Adult Translational Research, National Institute of Mental Health, Room 7107, 6001 Executive Blvd, Bethesda, MD 20892, USA.
e-mail: sarah.morris@nih.gov

[†]Present address:
Monica C. Mann-Wrobel, VISN 6 Mental Illness Research, Education, and Clinical Center, Durham VA Medical Center, 508 Fulton Street, Durham, NC 27705, USA.

Contrasting theories of schizophrenia propose that the disorder is characterized by a deficit in phasic changes in dopamine activity in response to ongoing events or, alternatively, by a weakness in the representation of the value of responses. Schizophrenia patients have reliably reduced brain activity following incorrect responses but other research suggests that they may have intact feedback-related potentials, indicating that the impairment may be specifically response-related. We used event-related brain potentials and computational modeling to examine this issue by comparing the neural response to outcomes with the neural response to behaviors that predict outcomes in patients with schizophrenia and psychiatrically healthy comparison subjects. We recorded feedback-related activity in a passive gambling task and a time estimation task and error-related activity in a flanker task. Patients' brain activity following an erroneous response was reduced compared to comparison subjects but feedback-related activity did not differ between groups. To test hypotheses about the possible causes of this pattern of results, we used computational modeling of the electrophysiological data to simulate the effects of an overall reduction in patients' sensitivity to feedback, selective insensitivity to positive or negative feedback, reduced learning rate, and a decreased representation of the value of the response given the stimulus on each trial. The results of the computational modeling suggest that schizophrenia patients exhibit weakened representation of response values, possibly due to failure of the basal ganglia to strongly associate stimuli with appropriate response alternatives.

Keywords: schizophrenia, error-related negativity, feedback, reward, dopamine

INTRODUCTION

For more than 30 years, pharmacological, neurophysiological, and neuroimaging studies have documented that the dopamine (DA) system is disrupted in schizophrenia (see Davis et al., 1991 for a review). Although the initial formulation of the dopamine hypothesis of schizophrenia, which proposed that the illness was the result of hyperdominergia (Matthysse, 1973) has been refined, the dopamine system has remained central to the study of schizophrenia. The antipsychotic effects of DA-blocking medications provide evidence of a relationship between tonic DA levels and the symptoms of schizophrenia but recent theoretical and empirical advances in the study of phasic DA activity (Schultz, 1998, 2002) allow new understanding of some of the most persistent cognitive and motivational deficits characteristic of the illness (Ziauddeen and Murray, 2010).

Many of these advances are based upon findings from studies of transient changes in mesencephalic DA neurons in primates (Schultz, 1998, 2002). This work describes phasic increases and decreases in firing of these neurons that can be understood as coding an error signal associated with a reinforcement learning algorithm (see Suri, 2002 for a review). In neural network models, reward prediction error signals (RPEs) are computed by an "adaptive critic" that attributes a value to ongoing events and outputs an error when it changes its own prediction. Positive (+) RPEs indicate that ongoing events are "better" than expected, and negative (−) RPEs indicate that ongoing events are "worse" than expected. This RPE signal may be used as a learning signal by DA target areas in order to optimize performance (Schultz et al., 1995). RPE signals are based on neural representations of value associated with different response alternatives (Montague and Sejnowski, 1994; Niv, 2009) and these values appear to be represented in the basal ganglia (Samejima et al., 2005; Lau and Glimcher, 2008) where the critic is thought to reside (O'Doherty et al., 2004).

This understanding of the functioning of phasic DA has given rise to contrasting models of schizophrenia and motivational impairment. One hypothesis is based on the idea that there is a primary impairment in the ability to signal prediction errors (Corlett et al., 2007; Frank, 2008; Fletcher and Frith, 2009), a hypothesis supported by evidence of reduced brain activity following RPEs in individuals with schizophrenia (Waltz et al., 2009) and disrupted frontal activity following RPEs in individuals with ketamine-induced delusions and perceptual aberrations (Corlett et al., 2006).

This hypothesis is also consistent with behavioral studies showing learning impairments and decreased reward-related response speeding (Waltz et al., 2007; Murray et al., 2008) in schizophrenia. By this account, disruption of RPE-related DA signaling interferes with reinforcement and response selection in schizophrenia, such that behavioral motivation is based on faulty associations and contingencies (Kapur, 2003; Smith et al., 2006). An alternative hypothesis proposes that schizophrenia patients experience decreased motivational drive due to impaired representation of response value despite normative evoked emotional experiences (Gold et al., 2008). This model is based on observations that associations between subjective valuation of stimuli and subsequent action selection are weaker in schizophrenia patients than psychiatrically healthy comparison subjects (Heerey and Gold, 2007) and that patients appear to fail to fully represent the full range of possible outcomes when choosing among gambling options (Heerey et al., 2008). This model is consistent with recent evidence of poor internal representation of motivational information (Ursu et al., 2011) and disruption of error likelihood predictions (Krawitz et al., 2011) in schizophrenia. This pattern of findings suggest that schizophrenia patients experience a deficit in the ability to simultaneously represent and consider the various cognitive and affective attributes associated with different response options, resulting in a selective impairment in motivation to seek out rewarding activities which is especially apparent in patients with high levels of negative symptoms (Gold et al., in press).

Reinforcement learning models can provide insight into these alternative mechanisms because they provide formal accounts of the relationship between predictive values associated with behavior and subsequent outcomes that do or do not violate those predictions (Niv, 2009). As a case in point, the reinforcement learning theory of the error-related negativity (RL–ERN; Holroyd and Coles, 2002) provides a framework for evaluating whether impairments in the representation of response Value (as mediated by the basal ganglia) and/or RPE signals (as mediated by the DA system) are apparent in schizophrenia. The RL–ERN model links the properties of the phasic DA system to learning-related changes in event-related brain potentials (ERPs) that follow correct vs. incorrect responses (the response ERN; Falkenstein, 1990; Gehring et al., 1993) and rewarding vs. non-rewarding outcomes (the feedback ERN; Miltner et al., 1997; Gehring and Willoughby, 2002; Ruchsow et al., 2002). In this model, changes in the response and feedback ERN that accompany learning reflect the functioning of a dopamine-mediated reward system in which motor neurons in the ACC use signals carried by the DA system for the adaptive modification of behavior. As with other reinforcement learning models, the RPEs are driven by changes in predictive Value implemented by the critic, which is hypothesized to lie in the basal ganglia (Holroyd and Coles, 2002). Thus, for example, the basal ganglia may contain units that represent whether a left or a right button press predicts reward. The RL–ERN theory holds that the generation of the ERN is associated with the impact on ACC of phasic decreases in DA activity on error trials (−RPEs) when events are worse than predicted vs. phasic positive increases of DA activity on correct trials (+TDEs; Holroyd et al., 2008) when events are better than predicted. In this way, the ERN is elicited when the system first determines the outcome of the trial such that a response ERN is elicited when the correctness of the response is detected immediately following the response and a feedback ERN is elicited when it is detected because of the feedback. In trial-and-error learning tasks, participants gradually learn the stimulus–response mappings and become able to judge their accuracy. Concomitantly, the performance feedback becomes redundant and the ERN propagates from the time of feedback presentation to the time of response generation (Holroyd and Coles, 2002).

Schizophrenia patients reliably exhibit diminished response ERN amplitude relative to healthy subjects across task types (Kopp and Rist, 1999; Alain et al., 2002; Bates et al., 2002, 2004; Mathalon et al., 2002; Morris et al., 2006) but the integrity of the feedback ERN in schizophrenia is less clear. Previously, to examine the integrity of the –RPE as an indicator of the function of the phasic DA system, we studied schizophrenia patients while they completed a probabilistic learning task (PLT) in which the validity of the accuracy feedback varied (Morris et al., 2008). Schizophrenia patients exhibited reduced response ERN amplitude in all conditions and reduced feedback ERN amplitude in the valid feedback condition during early trials when the feedback was most informative for stimulus–response learning. In the other feedback conditions, however, group differences in feedback ERN amplitude were equivocal. To date, this is the only study that has examined the feedback ERN in schizophrenia patients and the finding that this activity was only selectively impaired was unexpected. Thus, we were motivated to examine this question more closely.

Here, we examined whether schizophrenia patients exhibit impairment in both of these putatively DA-related ERP components (the response and feedback ERNs) or whether they exhibit a selective deficit in the response ERN only. Further, we used computational modeling to distinguish whether the results stem from an impairment in the transmission of RPE signals vs. a deficit in predictive value associated with response generation; because the RPEs by definition constitute violations of predictive value, without a formal account of their relationship the two possibilities would otherwise be difficult to disentangle. We examined the ERN on two tasks that elicit a feedback ERN, a passive gambling task and a time estimation task, and a third task that elicits a response ERN, a flanker task. The passive gambling task, modeled after Potts et al. (2006), was selected because it elicits a feedback ERN in the absence of a response, thus removing between-group variability due to differences in perceived task difficulty or response speed or accuracy. The time estimation task was selected because it elicits a robust feedback ERN (Miltner et al., 1997; Holroyd et al., 2006; Holroyd and Krigolson, 2007) and the parameters can be adjusted so that individuals with disparate response accuracy receive the various types of feedback with similar frequency. These tasks include a sufficient number of trials (more than 50) to obtain a reliable feedback ERN (Marco-Pallares et al., 2011). The Eriksen flanker task (EFT; Eriksen and Eriksen, 1974) was used for eliciting the response ERN because it induces speeded motor errors and does not require memorization of response rules that might prove differentially challenging for the patient group. It has also been used in previous studies of the ERN in schizophrenia (Kopp and Rist, 1999; Morris et al., 2006) so it serves as a good benchmark for evaluating the findings of the current study. In addition to these three tasks, we also re-analyzed the ERP data from our

previous PLT study (Morris et al., 2008) using a difference wave approach (Holroyd and Krigolson, 2007) in order to measure both the response and feedback ERN in a single task for the purposes of computational modeling.

To preview the results of these studies, the response ERN deficit in schizophrenia was replicated but the same patients showed normal feedback ERN amplitude on both feedback tasks. These results, however, do not unambiguously support either the disrupted RPE model or the impaired response value model of schizophrenia, as it is possible that the response ERN was reduced due to an abnormality in RPE signaling or to a weakened representation of response value. Therefore, we used computational modeling to simulate the ERP results and test alternative hypotheses about the origin of this dissociation between response and feedback ERN abnormalities. Specifically, we utilized a formal instantiation of the RL–ERN theory to parametrically and systematically vary two parameters related to the neural computation of Value (as expressed by the basal ganglia) and to the change in Value (as expressed by the dopamine system) to explore how changes to these parameters would affect the ERN and behavior.

MATERIALS AND METHODS

PARTICIPANTS

Thirty-two schizophrenia outpatients and 23 healthy comparison subjects completed the time estimation task. All subjects except for two control subjects completed the passive gambling task. A subset of 20 patients and 15 comparison subjects (the final participants recruited into the study) completed the flanker task in addition to the other two tasks. Demographic and clinical characteristics are summarized in **Table 1**. The groups did not differ in age, $t(53) = 0.15$, $p = 0.88$, gender, χ^2 $(1, N = 55) = 0.075$, $p = 0.78$, or ethnicity, χ^2 $(2, N = 55) = 2.00$, $p = 0.37$. Schizophrenia patients had fewer years of education than comparison subjects, $t(53) = -2.79$, $p = 0.007$, but did not differ in parental education, $t(53) = 0.28$. The results of these comparisons in age, gender, ethnicity, education, and parental education did not change when only the participants who completed the flanker task were compared.

Table 1 | Demographic and symptom rating data for control subjects and schizophrenia patients.

	Control subjects ($n = 23$)		Schizophrenia patients ($n = 32$)	
	M	SD	*M*	SD
Age (years)	46.74	11.01	47.09	6.74
Education (years)[a]	14.48	2.25	12.98	1.72
Parent's highest education (years)	13.52	2.25	13.74[b]	3.26
Gender				
Male	18		26	
Female	5		6	
Ethnicity				
European American	13		14	
African American	10		16	
Multiracial			2	
BPRS 20-item total score			36.06	11.25
SANS 22-item total score			32.22	16.24

BPRS, brief psychiatric rating scale; SANS, scale for the assessment of negative symptoms.

[a] *Group difference $p = 0.007$.*

[b] *$N = 31$.*

Patients were recruited from outpatient psychiatric clinics at the Maryland Psychiatric Research Center and the Baltimore Veterans Affairs Medical Center. They were diagnosed using a best-estimate approach combining information from medical records, collateral information (when available), and the structured clinical interview for DSM-IV (SCID; First et al., 1994). Twenty-seven of the patients were diagnosed with schizophrenia and five were diagnosed with schizoaffective disorder. Patients were medicated with second-generation antipsychotic medication(s) (APM; $n = 31$) or both a second-generation antipsychotic and a traditional APM ($n = 1$). On the day of testing, symptom ratings were obtained using the brief psychiatric rating scale (BPRS; Overall and Gorham, 1962) and the scale for the assessment of negative symptoms (SANS; Andreasen, 1982).

Healthy comparison subjects were recruited via newspaper advertisements, fliers, or random-digit dialing of local phone numbers. They were assessed with the SCID and had no personal or family history of schizophrenia or schizoaffective disorder and no personal lifetime history of bipolar disorder or major depressive disorder. A lifetime history of head injury or neurological illness and alcohol or substance dependence in the last 6 months were exclusionary criteria for all participants. All participants provided written informed consent for the protocol which was approved by the University of Maryland School of Medicine IRB and the VA Maryland Healthcare System Research and Development Committee.

TASKS

Testing procedures took place in a sound-attenuated, dimly lit room in which participants were seated approximately 1 m from a video monitor. The order of tasks was counterbalanced.

PASSIVE GAMBLING TASK

Participants completed a passive gambling task modeled after Potts et al. (2006) in which participants viewed pairs of pictures presented sequentially (see also Holroyd et al., 2011). Each picture depicted either a lemon or a gold bar. On 80% of trials, pairs consisted of the same stimulus (i.e., lemon followed by lemon or gold followed by gold, with equal probability). On the remaining trials, pairs consisted of one of each stimulus (i.e., lemon followed by gold or gold followed by lemon, with equal probability). When a gold bar was presented as the second image, it was always followed by feedback indicating a 50¢ bonus. When a lemon was presented as the second image, it was always followed by feedback indicating no bonus. Thus four feedback conditions were created: unexpected non-reward (bar followed by lemon), expected non-reward (lemon followed by lemon), unexpected reward (lemon followed by bar), and expected reward (bar followed by bar). Participants viewed 7 blocks of 48 trials and the total bonus earned during each block was displayed following each block. Participants

were informed that one of the bonus amounts (ranging from 9 to 15$) would be selected at random and added to their payment. Trials were ordered in a constrained random sequence such that stimulus pairs were not repeated on more than three consecutive trials. Pictures were displayed for 600 ms. The interval between pictures in a pair was 400 ms, the feedback was displayed for 2200 ms, and the interval between feedback offset and onset of the next image was 400 ms.

TIME ESTIMATION TASK

Participants completed a task modeled after Holroyd et al. (2006) modification of Miltner et al. (1997) time estimation task in which they were instructed to press a button when they estimated that 1 s had elapsed after the presentation of a tone. Following each response, feedback indicating whether the response was on-time (a plus sign) or not on-time (a zero) was displayed. An adjustable response window was used which decreased the allowable RT deviation as participants became more accurate. The window was initially set to 900–1100 ms and then narrowed by 20 ms following each on-time response and widened by 20 ms following each not on-time response. The maximum window was 0–2000 ms. In light of Holroyd et al. (2006) finding that uninformative feedback elicited a feedback ERN similar in amplitude to that elicited by negative feedback, uninformative/neutral feedback (a question mark) was provided on one-third of trials (selected at random). The tone was 80 Hz, 80 dB presented for 50 ms via ear inserts. Feedback was presented for 1000 ms beginning 3050 ms after tone offset. The interval between feedback offset and the onset of the next tone was 2000 ms. Participants completed 210 trials with brief rest pauses after 70 and 140 trials.

FLANKER TASK

Subjects performed a modified version of the flanker task (Eriksen and Eriksen, 1974). Each trial began with the display of flanker stimuli which were two pairs of equilateral triangles or squares appearing in a vertical array. Flanker stimuli were displayed for 100 ms before the middle triangle, the target, appeared. Participants were instructed to respond with the hand that corresponded to the direction in which the target was pointing. The flanking triangles were oriented either in the same (congruent condition) or opposite (incongruent condition) direction as the target or flanking squares were used instead of triangles (neutral condition). The six different types of target/flanker combinations were presented with equal frequency and in a constrained random sequence such that no trial type was repeated on more than three consecutive trials. The flanker/target array was displayed for 70 ms. Beginning 2000 ms after the offset of the target array, feedback was displayed for 1000 ms. The delay between the offset of the feedback and the onset of the subsequent flankers was 1950 ms.

Before beginning the flanker task, subjects were instructed to respond quickly and accurately and were penalized 2¢ for incorrect responses, rewarded 2¢ for correct responses and penalized 5¢ for slow responses (RT > 1100 ms) regardless of accuracy. Participants completed 24 practice trials followed by 6 blocks of 54 trials. All flanker stimuli were white presented on a black background.

PSYCHOPHYSIOLOGICAL RECORDING, DATA REDUCTION, AND ANALYSES

General procedures

Electroencephalography (EEG) recordings were obtained using 32 Ag/AgCl electrodes in International 10/20 system positions. Electrooculographic activity was recorded from electrodes placed above and below the left eye and at the outer edge of both eyes. Physiological signals were recorded using a Synamps amplifier and Scan 4.3 software (Compumedics/Neuroscan, Charlotte, NC, USA). Scalp EEG data were recorded at a rate of 500 Hz and referenced to averaged earlobe electrodes. After epoching, vertical and horizontal eye movement artifacts were corrected offline (Gratton et al., 1983; Miller et al., 1988), a 0.1- to 20-Hz 24 dB filter was applied and a 200-ms baseline was subtracted from each epoch. The Greenhouse–Geisser adjustment for repeated measures and an alpha level of 0.05 were used. Corrected F, p, and effect size (partial eta squared, or η_p^2) values and uncorrected degrees of freedom are reported. Simple-effects ANOVAs with the Bonferroni correction were used for *post hoc* comparisons on between-group measures.

Although we have previously (Morris et al., 2008) examined the ERN using a "base-to-peak" approach, recent investigations have indicated that a "difference wave" approach may be more appropriate for extracting this ERP component (Holroyd and Krigolson, 2007), first because it minimizes overlap with other interfering ERP components (Luck, 2005), and second because recent evidence indicates that unexpected positive feedback may elicit a positive-going deflection in the ERP (Holroyd et al., 2008; See also: Potts et al., 2006; Cohen et al., 2007; Eppinger et al., 2008; Baker and Holroyd, 2011; Holroyd et al., 2011) that the base-to-peak approach may overlook. Procedures used for computing difference waves are provided below. Group means and SE for the difference wave amplitude from for the feedback ERN (Passive gambling task and time estimation task) and the response ERN from the flanker task are provided in **Table 2**.

Passive gambling task

For the second stimulus of each pair, "Expected" difference waves were computed by subtracting the Expected Bonus waveforms

Table 2 | Mean (and SE) feedback and response ERN difference wave amplitudes for schizophrenia patients and control subjects.

Task	Condition	Control subjects ($n = 23$)	Schizophrenia patients ($n = 32$)
Passive gambling[a]	Expected	−2.59 (0.51)	−3.17 (0.31)
	Unexpected	−3.59 (0.66)	−4.02 (0.55)
Time estimation	Neutral	−3.63 (0.45)	−4.53 (0.68)
	Zero	−4.05 (0.55)	−4.98 (0.70)
Flanker[b]		−11.46 (2.24)	−4.66 (1.28)

Data are from Cz (passive gambling task) and FCz (time estimation and flanker); feedback ERN: passive gambling task and time estimation task; response ERN: flanker task.

[a] *N = 21 control subjects.*

[b] *N = 15 control subjects and 20 patients.*

from the Expected Non-Bonus waveforms. "Unexpected" difference waveforms were computed by subtracting the Unexpected Bonus waveforms from the Unexpected Non-Bonus waveforms (Holroyd and Krigolson, 2007). Examination of the grand-average difference waveforms indicated substantial component overlap characterized by early, frontally distributed activity (the feedback ERN) followed by posteriorly distributed activity (P300). To isolate the feedback ERN from this overlapping component, the latency of the waveforms was adjusted at all channels using the following procedure. The latency of the maximal negativity between 170 and 290 ms was determined for the FCz channel in the expected non-bonus average for each participant. The expected non-bonus condition was used because the feedback ERN was maximal in this condition in the group averages. This latency was then set to 240 ms (the approximate latency of peak in the unadjusted group average waveforms) by adjusting the start point of the waveform by the difference between the latency of the peak and 240 ms. The conditional waveforms for all four feedback conditions in all channels were then adjusted by the same degree and difference waves were re-computed using these adjusted averages. The amplitude of the maximal negativity occurring between 180 and 300 ms was identified in the difference waves and was analyzed using 2 (Group) × 2 (Expectedness) × 5 (Site: Fz, FCz, Cz, CPz, Pz) mixed-model ANOVA. To check whether the latency adjustment obscured effects of interest, difference wave amplitudes from the non-adjusted waveforms were analyzed and the results were unchanged.

Time estimation task

Difference waves for the neutral and "zero" feedback conditions were created by subtracting activity following "plus" feedback from the activity following neutral and "zero" feedback, respectively (Holroyd and Krigolson, 2007). Examination of the grand-average difference waves indicated component overlap similar to that observed in the data from the time estimation task, so a similar latency adjustment was made. For each participant, the latency of the maximum negativity within a narrow window (200–290 ms following FB onset) was identified in the "zero" feedback average in the FCz channel. This latency was then set to 240 ms (the approximate latency of peak in the unadjusted group average waveforms) by adjusting the start point of the waveform by the difference between the latency of the peak and 240 ms. The conditional waveforms for all three feedback conditions in all channels were then adjusted by the same degree. Difference waves were then re-created and the feedback ERN was scored as the amplitude of the maximum negativity occurring between 200 and 280 ms in the neutral and "zero" difference waves. The feedback ERN was compared in a 2 (Group) × 2 (Difference wave type) × 5 (Site: Fz, FCz, Cz, CPz, Pz) mixed-model ANOVA. The results of this analysis were also unchanged when difference wave amplitudes from the non-adjusted waveforms were analyzed.

Flanker task

Response-locked waveforms were created for correct and error trials and difference waves were created by subtracting the activity following correct responses from that following errors. The response ERN was quantified in these difference waves as the maximal negativity between 0 and 150 ms. Data for one control subject whose response ERN difference wave peak amplitude exceeded the group mean by more than 4 SDs were replaced with the next largest values. These were analyzed in two 2 (Group) × 5 (Site: Fz, FCz, Cz, CPz, Pz) mixed-model ANOVAs.

RESULTS

ERP STUDIES

Passive gambling task

Event-related brain potential waveforms for the second stimulus, which varied in both valence and expectedness, are shown in **Figure 1**. Consistent with the hypothesis that the feedback ERN is relatively intact in schizophrenia, there was no difference between groups in feedback ERN difference wave amplitude, $F(1, 51) = 0.10$, $p = 0.75$, $\eta_p^2 = 0.002$. Although the peak negativity is more prominent for the expected feedback than for the unexpected feedback in the group average waveforms, the group mean difference scores were larger for unexpected than for expected outcomes, although the effect of expectedness was not significant, $F(1, 51) = 2.91$, $p = 0.09$, $\eta_p^2 = 0.05$. Difference wave amplitudes were greatest at Cz but did not differ significantly among sites, $F(4, 204) = 2.32$, $p = 0.08$, $\eta_p^2 = 0.02$. There were no interactions involving group, expectedness, or channel (all p values > 0.3).

Time estimation task

Behavior. Because RTs were not recorded for responses occurring more than 3000 ms after the tone, long RTs could not be distinguished from non-responses. In order to avoid overestimating participants' RT accuracy by omitting these highly inaccurate responses, missing RTs were replaced with the RT from the previous trial (or from the next trial if no response was made on the previous trial). This process may have resulted in a minor overestimation of participants' response accuracy. The average percentage of trials with missing/replaced RTs did not differ between patients (6%) than controls (3%), $F(1, 53) = 1.35$, $p = 0.25$. The absolute deviation of RT from the target RT of 1000 ms was then computed for each trial. Mean RT deviation and a running average of RT deviation over the course of the task are presented in **Figure 2**. As seen in the figure, the controls were slightly more accurate at estimating the 1-s interval but this difference was not statistically reliable: mean RT deviation, $t(53) = 1.83$, $p = 0.073$. Both groups received "zero" feedback on a greater percentage of trials (36%) than "plus" feedback (30%), $F(1, 53) = 25.38$, $p = 0.00$, $\eta_p^2 = 0.32$, consistent with the initially narrow RT window. Response time accuracy improved by an average of 87 ms on trials following "zero" feedback, worsened by an average of 106 ms after "plus" feedback and changed by less than a millisecond on average after neutral feedback [main effect of FB type, $F(2, 106) = 159.41$, $p = 0.00$, $\eta_p^2 = 0.75$]. This main effect was moderated by a Group × FB type interaction, $F(2, 106) = 3.33$, $p = 0.05$, $\eta_p^2 = 0.06$ characterized by control subjects having a smaller increase in deviation following "plus" feedback than patients, 88 vs. 125 ms, $t(53) = 2.09$, $p = 0.04$, suggesting that patients had difficulty sustaining accurate responding following positive feedback.

Event-related brain potentials. Group average latency-adjusted waveforms for the feedback-locked ERPs are presented in **Figure 3**.

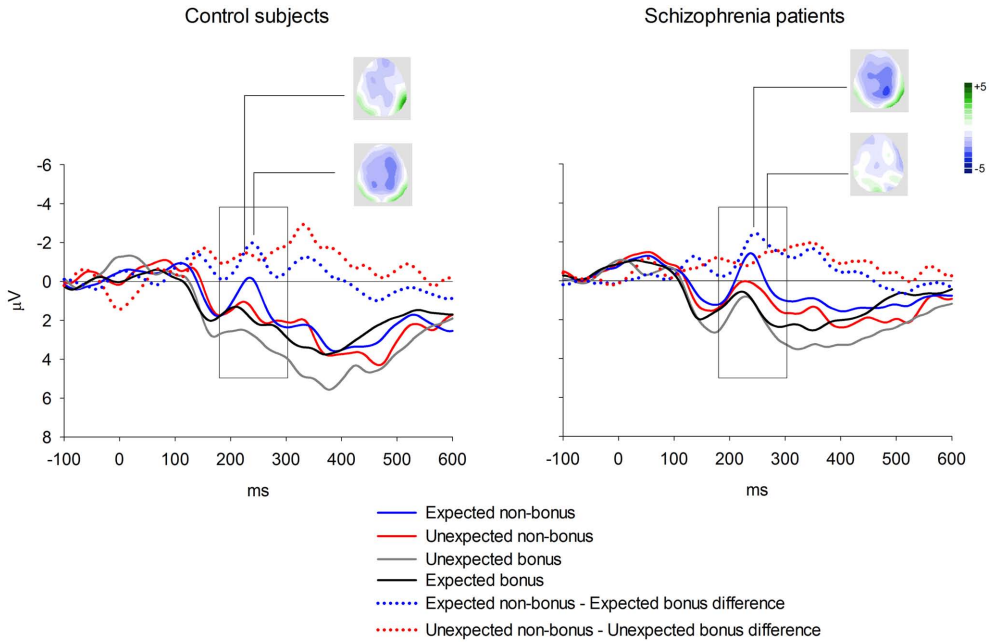


FIGURE 1 | Group averages for feedback ERN elicited by second stimulus in the passive gambling task. Data shown in waveforms are from Cz. The scoring window for the difference waves is indicated by the rectangle.

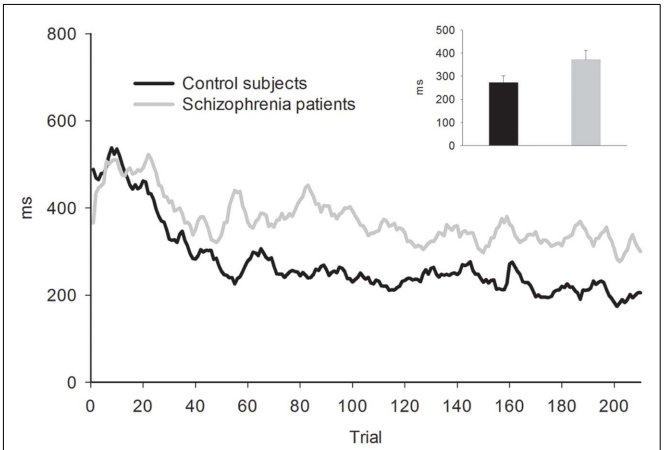


FIGURE 2 | Five-trial running average of absolute RT deviation from target RT for the time estimation task. Bar graph shows mean of absolute RT deviation. Error bars indicate SE.

Prominent negativities are observed approximately 250 ms following both "zero" and neutral FB compared to "plus" FB. As in the passive gambling task, there was no main effect of group, $F(1, 53) = 1.47$, $p = 0.23$, $\eta_p^2 = 0.03$ or any interactions involving group differences for the feedback ERN (all p values > 0.53). There was also no difference between the "zero" and neutral FB conditions in difference wave amplitude, $F(1, 53) = 2.32$, $p = 0.13$, $\eta_p^2 = 0.04$, suggesting that feedback ERN activity in both groups shows the expected pattern of dichotomous classification of outcomes as described by Holroyd et al. (2006). The difference wave amplitudes were largest at the FCz site although the amplitude difference among channels was not significant, $F(4, 212) = 2.87$, $p = 0.06$.

Flanker task

Behavior. The flanker type manipulation had the expected effects on response accuracy, $F(2, 66) = 70.22$, $p = 0.00$, $\eta_p^2 = 0.68$, and RT, $F(2, 66) = 153.23$, $p = 0.00$, $\eta_p^2 = 0.82$, with better accuracy and faster RT in the congruent compared to incongruent flanker conditions [$F(1, 33) = 68.55$ and 233.42, respectively, $p < 0.001$; See **Figure 4**]. Mean RT was slower for schizophrenia patients than control subjects, $F(1, 33) = 10.83$, $p = 0.002$, $\eta_p^2 = 0.25$.

Event-related brain potentials. Consistent with previous studies (e.g., Kopp and Rist, 1999; Morris et al., 2006), the response ERN as measured by error-correct difference wave amplitude was diminished in schizophrenia patients compared to control subjects, $F(1, 33) = 6.31$, $p = 0.02$, $\eta_p^2 = 0.16$ (**Figure 5**). Difference score amplitude was largest at FCz for control subjects and at CPz for patients [Group × electrode site interaction, $F(4, 132) = 3.57$, $p = 0.05$]. Because the participants who completed the EFT were a subset of those who completed the time estimation and passive gambling tasks, we repeated the analyses of the data from the feedback ERN tasks including only the participants who completed the flanker task to make sure that the dissociation between response ERN and feedback ERN was not due to the composition of the subject groups. The differences among these samples did not appear to account for the selective deficit in response ERN compared to feedback ERN. The results of the analysis of the ERP data from the time estimation and passive gambling tasks did not differ when individuals who did not complete the flanker task were removed from the analysis (all p values > 0.05). The fact that group differences were

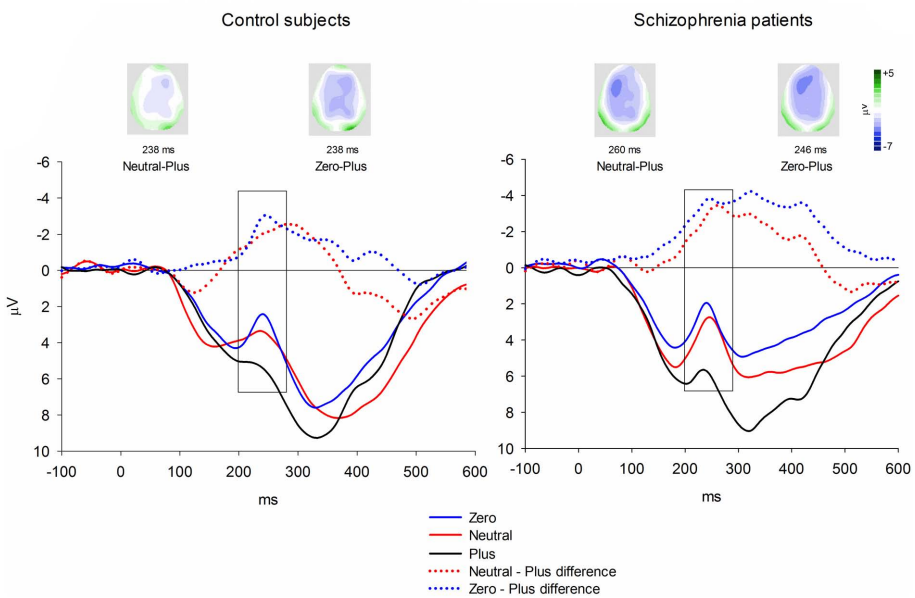


FIGURE 3 | Feedback-locked ERN group averages for the time estimation task. Data shown in waveforms are from FCz. Data in maps are shown at latency of maximal difference wave amplitude. The scoring window for the difference waves is indicated by the rectangle.

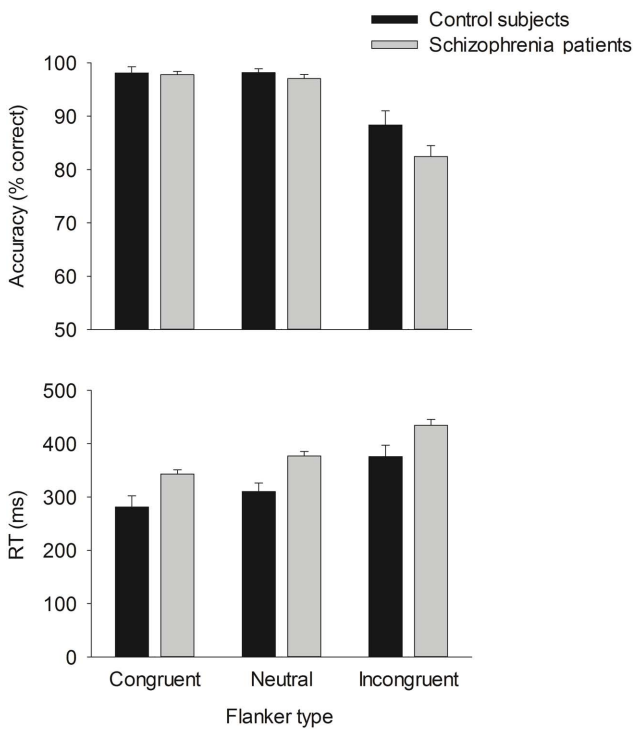


FIGURE 4 | Mean accuracy (top) and RT (bottom) for the flanker task. Error bars indicate SE.

observed in the response ERN with a smaller sample size provides some assurance that the failure to find group differences in the feedback ERN was not due to lack of power since the sample size was larger for the feedback ERN tasks. It is, however, possible that there are differences in effect size between response and feedback ERN and that small group differences in feedback ERN could have gone undetected due to lack of power.

Computational modeling

Taken together, the above results are strongly suggestive of an impaired response ERN together with a spared feedback ERN in schizophrenia. But what deficit can cause this pattern of observations? Alternative hypotheses (as reviewed above) hold that schizophrenia is associated with impaired dopamine-dependent RPE signals on the one hand (Corlett et al., 2007; Fletcher and Frith, 2009) and with impaired representations of predictive value on the other (Gold et al., 2008). These hypotheses are challenging to distinguish because of their complex interrelationship: predictive values can be derived from RPEs and RPEs reflect changes in predictive value. Computational simulations based on principles of reinforcement learning can illuminate this issue because neurally based models of decision making depend on formal representations of predictive value and RPEs (Suri, 2002; Cohen, 2008; Dayan and Daw, 2008; Cohen and Frank, 2009; Niv, 2009). Further, the RL–ERN theory specifically indicates how these parameters give rise to the response ERN and feedback ERN. For this reason, we adapted a computational model of the response ERN and feedback ERN in the PLT and of the response ERN in the EFT (Holroyd and Coles, 2002) to determine how these ERP components would be affected by changes to these quantities as might occur in schizophrenia. Our modeling efforts focused on these two tasks because they were originally simulated with the RL–ERN model (Holroyd and Coles, 2002). Further, data from our previous study using the PLT (Morris et al., 2008) allowed for comparison of the response ERN and the feedback ERN in a single task, whereas the

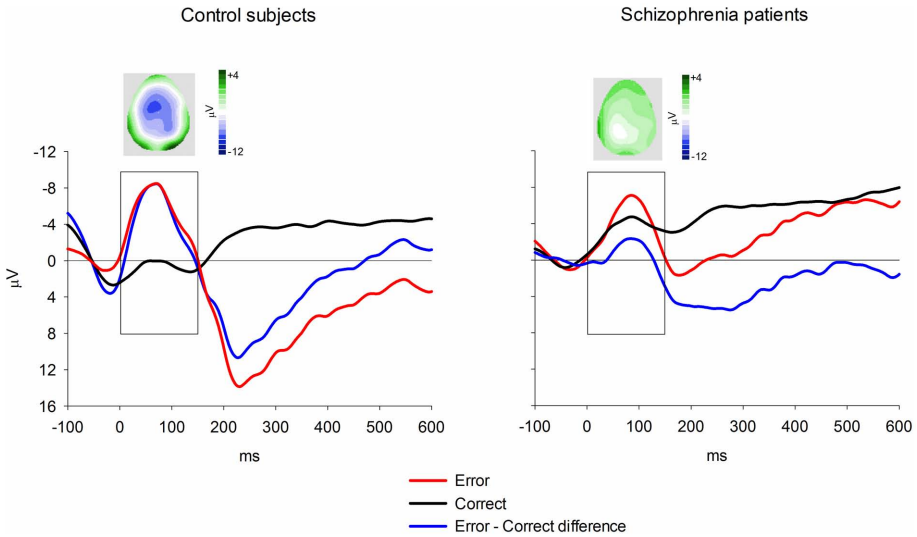


FIGURE 5 | Response-locked ERN group averages for the flanker task. Data shown in waveforms are from FCz. Topographical maps depict distribution of difference wave at latency of peak negativity. The scoring window for the difference waves is indicated by the rectangle.

EFT data from the current study allowed for examination of the response ERN in the absence of feedback-based learning (Holroyd and Coles, 2002, 2008; See also Holroyd et al., 2005).

Methods

The RL–ERN model belongs to a class of neurobiologically motivated computational models that are based on the theory of reinforcement learning (Sutton and Barto, 1998) and that simulate behavior and/or the activity of the midbrain dopamine system on trial-and-error type learning problems (Suri, 2002; Cohen, 2008; Dayan and Daw, 2008; Cohen and Frank, 2009; Niv, 2009). We simulated the ERN and performance of control subjects on the PLT and the EFT using a variant of the standard RL–ERN model, details of which are given in Holroyd and Coles (2008). The original RL–ERN model utilized multiple motor controllers that competed for control over behavior (Holroyd and Coles, 2002). As this aspect of the model was not central to the hypothesis under investigation here, we adopted a reduced model that selected actions based on state-action Values encoded by the critic module (Holroyd and Coles, 2008). The model includes units that represent important task states, namely external stimuli, stimulus–response conjunctions, and feedback stimuli, which activate when the corresponding event occurs on a given trial. Further, connection weights associated with each unit represent the internal "Value" of that state. For the PLT simulation, the Values associated with positive and negative feedback were fixed at 1 and −1, respectively, and for the EFT simulation, the Values associated with the correct and incorrect stimulus–response conjunctions were fixed at 1 and −1, respectively. For both simulations, the weights for the remaining stimuli were internalized with random values between −0.5 and 0.5. Critically, the strength of the phasic dopamine signal was related to the magnitude of the reward prediction error (i.e., the "temporal difference error"), defined as the change in Value associated with state transitions (Sutton, 1988). The RPE was used to modify the Value weights according to the temporal difference learning rule (Sutton, 1988; Sutton and Barto, 1998). Consistent with previous simulations (Holroyd and Coles, 2002, 2008; Nieuwenhuis et al., 2002) and with empirical practice (Holroyd and Coles, 2002; Holroyd and Krigolson, 2007), the simulated ERN was determined by subtracting the RPE on correct trials from the RPE on error trials. For all simulations response selection was "epsilon-greedy" (Sutton and Barto, 1998), meaning that the model chose the response with the highest Value on a fraction of randomly selected trials (70%) and chose a response at random on the remaining trials (30%). The learning rate parameter for the simulated control subjects was equal to 0.5.

Results

Probabilistic learning task. The PLT is a trial-and-error learning task where participants are required to press one of two buttons on each trial in response to presentation of an arbitrary visual image and are provided feedback indicating that they received or were penalized a small amount of money. Key to the task is that the imperative stimuli are probabilistically related to the appropriate response. Here we simulated the data of Morris et al. (2008), wherein the optimal response was associated with reward on either 100, 80, or 50% of encounters with the associated stimulus (See also Nieuwenhuis et al., 2002). The demographic and clinical characteristics of the participants were highly similar to those of the participants in the current study. The empirical accuracies for the control participants and for participants with schizophrenia are presented in **Table 3**. Note that the accuracies for the two groups are comparable, as are the accuracies for all of the simulations.

For the purposes of the computational modeling, we have reanalyzed the data of Morris et al. (2008) using a difference wave approach. **Figure 6** illustrates the ERN difference wave data for the control participants and for participants with schizophrenia. These results replicate the common finding that the response ERN

Table 3 | Empirical and simulated accuracy rates for control subjects and schizophrenia patients on the probabilistic learning task.

	Feedback condition		
	100%	**80%**	**50%**
Empirical control	83	71	50
Empirical patient	78	71	50
Simulated control	84	68	50
±RP	84	69	50
+RP	85	69	50
−RP	85	69	50
Learning rate	81	67	49
Max. value	84	68	49

Note that correct trials are defined as those in which the optimal (i.e., most frequently rewarded) response was emitted, as opposed to trials in which positive feedback stimuli were delivered. Empirical data taken from Morris et al. (2008).

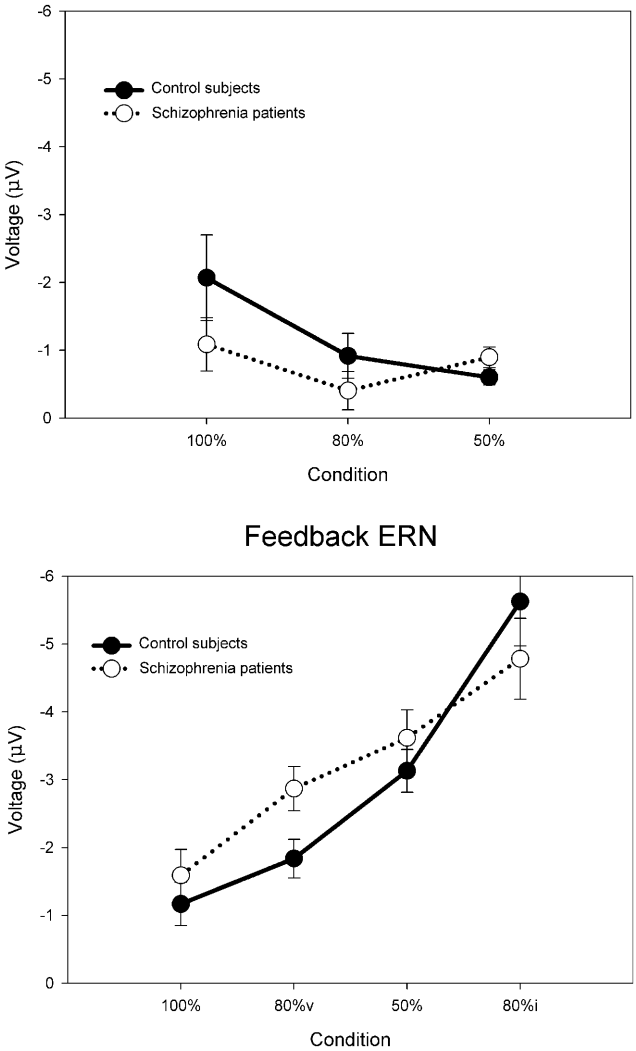


FIGURE 6 | Empirical control and schizophrenia ERN difference wave data (mean amplitude and SE) for the probabilistic learning task, for the response ERN (top) and feedback ERN (bottom). Note that "80%v" and "80%i" correspond to trials in the 80% condition with either valid or invalid feedback, respectively. Data recorded at FCz and re-analyzed from Morris et al. (2008).

and feedback ERN amplitude are inversely related such that they are larger and smaller, respectively, with increasing certainty of the outcome (Holroyd and Coles, 2002; Nieuwenhuis et al., 2002, 2005). Although a Group × Feedback type ANOVA did not show any main effects or interactions involving group, these data suggest that this inverse relationship was stronger for the control participants than for the participants with schizophrenia, who exhibited relatively less variation in ERN amplitude across conditions. To examine this more closely, we applied separate linear regressions to each individual's response ERN and feedback ERN data, with conditions entered in the order of smallest to largest ERN (response ERN: 50%, 80%, 100%; feedback ERN: 100%, 80%v, 50%, 80%i). A two-way ANOVA on the regression intercepts, with levels for group (controls, schizophrenia patients) and ERN type (response, feedback), indicated a main effect for group, $F(1, 51) = 5.8$, $p = 0.02$, such that the regression intercepts were more negative for the participants with schizophrenia than for the control participants. This observation was supported by a comparable ANOVA on the regression slopes that indicated a strong trend for a main effect of group, $F(1, 51) = 3.7$, $p = 0.06$, such that the regression slopes were smaller for the people with schizophrenia than for the control subjects. These findings confirm the visual impression in **Figure 6** that the ERN amplitudes of the participants with schizophrenia were relatively insensitive to condition, leading relatively negative intercepts and shallower slopes for the patient group. Because the RL–ERN theory is mainly concerned with the inverse relationship between the amplitudes of the response ERN and feedback ERN across conditions (Holroyd and Coles, 2002), our modeling efforts focused on accounting for the blunting of this relationship in the patient data.

Each simulation consisted of 27 runs of four blocks of 300 trials of the PLT, with each run corresponding to a simulated "participant." For the control simulation, the default parameters reproduced the inverse relationship between the response and feedback ERN amplitudes (solid lines in **Figure 7**) typically observed in this task (Holroyd and Coles, 2002; Nieuwenhuis et al., 2002, 2005). To evaluate the RPE vs. predictive Value theories of schizophrenia, we followed a hypothesis-driven approach in which we explored how plausible changes to specific model parameters would affect behavior and the ERN (see, e.g., Nieuwenhuis et al., 2002). Each hypothesis was explored by utilizing the parameters of the control simulation and scaling the size of a parameter that implemented the hypothesis above and below its default value. This was achieved by alternatively increasing and decreasing the size of the parameter of interest in increments of 1% from 0 to 200% relative to the parameter value for the control simulation. For example, we first explored the hypothesis that schizophrenia is associated with faulty phasic dopamine signals (e.g., Frank, 2008; Fletcher and Frith, 2009) that have been posited to be either larger or smaller in schizophrenia relative to the normal population (Bilder et al.,

2004). To do so, we reran the model through multiple iterations that systematically increased the size of the RPE from 0 to twice its default value in the control simulation. The iteration with the parameter value that produced the best fit to the patient electrophysiological data was then taken as the model that best accounted for the hypothesis, e.g., if the patient data were best fit by a model with a RPE signal reduced by 25% of its default (control) value. Then, this procedure was repeated for the subsequent hypothesis, for example, by systematically manipulating the maximum possible size of the predictive Values from very small (0) to very large (200% normal), and so on. Multiple sets of simulations were run in this way to investigate related hypothesis (disrupted positive and negative RPEs, disrupted negative RPEs only, disrupted positive RPEs only, normal RPEs with disrupted learning rate, etc.). Critically, each hypothesis was explored by changing only a single parameter while the remaining parameters remained fixed to those of the control simulation, so multiple parameters were never simultaneously varied within a single simulation; this practice addresses the degrees of freedom problem that is sometimes leveled at the computational modeling approach (O'Reilly and Farah, 1999).

The sum of squared errors (SSE) between the simulated and empirical schizophrenia ERN data was used to evaluate the fit of each simulation (**Figure 8**). In **Figure 8**, note that the horizontal dashed line at SSE = 0.14 indicates the error value associated with the control simulation compared to the empirical control data, whereas the horizontal dotted line at SSE = 0.36 indicates the error value associated with the control simulation compared to the empirical schizophrenia data; values below the dotted line indicate improvements in model fit to the schizophrenia data and values below the dashed line indicate that the fit of the schizophrenia model to the schizophrenia data is even better than that of the control model to the control data. Further, note that parameter values of 100% correspond to those of the control simulation, thus the SSEs for each set of simulations cross the dotted line at parameter values of 100%. In other words, because each of the schizophrenia simulations is the same as the control simulation except for the value of one parameter, when that one parameter is in fact the same as that of the control simulation (100%), then the schizophrenia simulation is identical to the control simulation. In that case, the SSE of the schizophrenia model is the same as that of the control model when fit to the empirical schizophrenia data, i.e., SSE = 0.36. Finally, note that for the purpose of comparison the empirical and simulated data were normalized between values of 0 and −1, where the smallest value combined across control and schizophrenic groups was equal to zero and the largest negative value was set equal to −1.

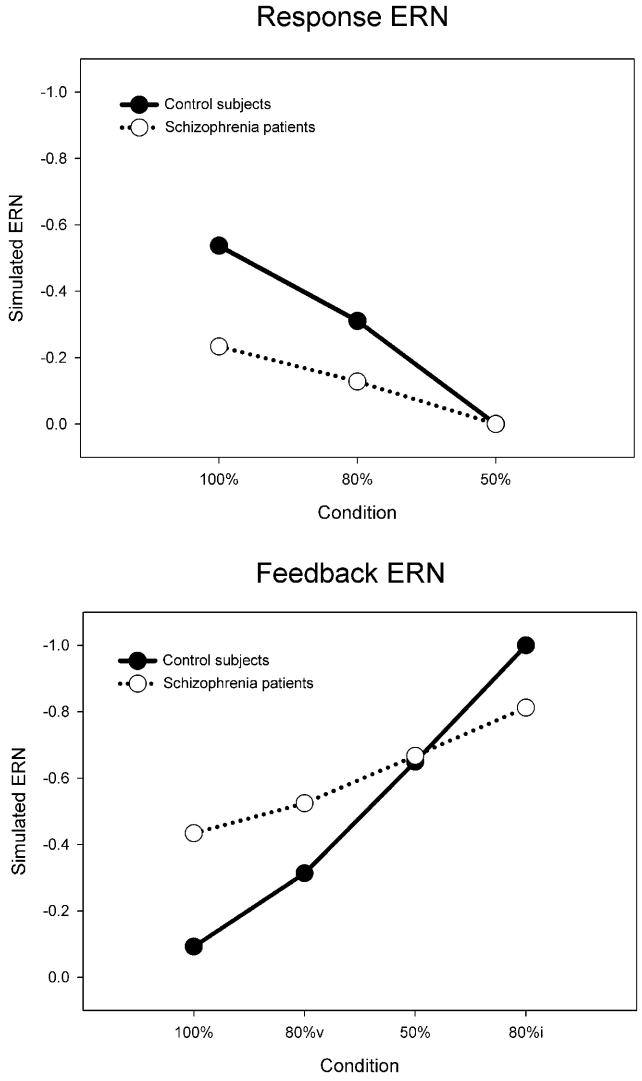


FIGURE 7 | Simulated control (default model) and schizophrenia (maximum value model) ERN difference wave data for the probabilistic learning task, for the response ERN (top) and feedback ERN (bottom). Note that "80%v" and "80%i" correspond to 80% condition trials with either valid or invalid feedback, respectively.

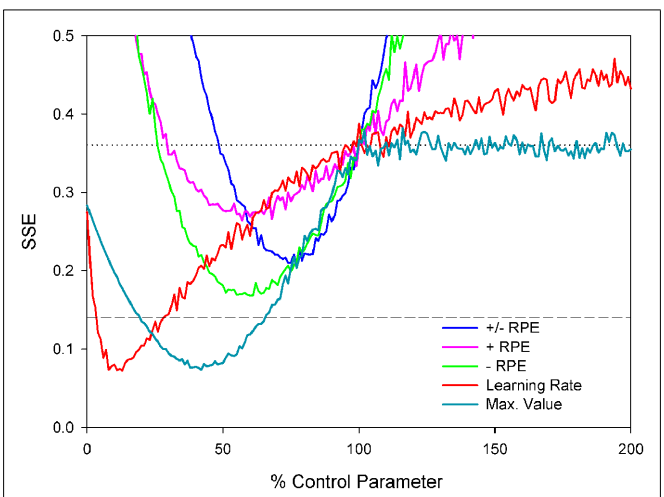


FIGURE 8 | Sum of squared errors (SSE) for each of the probabilistic learning task simulations. Note that the horizontal dashed line at SSE = 0.14 indicates the error value associated with the control simulation compared to the empirical control data and the horizontal gray dotted line at SSE = 0.36 indicates the SSE of the control simulation data compared to the empirical schizophrenia data; values below this line indicate better fits.

We first examined the basic hypothesis that schizophrenia would be associated with an abnormal ERN resulting from a disturbed dopamine/RPE signal. As described above, we scaled the size of RPE signal from 0 to 200% of the control simulation values (**Figure 8**, blue line). The best fit corresponded to RPE sizes that were 75% of normal (SSE = 0.21), producing simulated ERN amplitudes shown in **Figure 9** (blue lines, which for the feedback ERN plot is hidden by the pink and green lines). These ERN values are relatively inconsistent with the empirical schizophrenia data (**Figure 9**, black dotted lines) as they reflect less than a 50% reduction in the SSE of the control simulation relative to the schizophrenia data (**Figure 8**, gray dotted line), as well as a fit that was worse than the control simulation to the control data (**Figure 8**, gray dashed line). We also investigated whether schizophrenia might impact either the positive or negative RPE signal independently of the other by varying each while holding the other constant. Both sets of simulations were associated with optimal parameter values of 60%, but the reduced negative RPEs (green lines in **Figures 8** and **9**) produced better fits (SSE = 0.17) than the reduced positive RPE (pink lines in **Figures 8** and **9**) did (SSE = 0.26). We also explored whether changes to the positive and negative RPE signals might be inversely related such that an increase to one was associated with a commensurate decrease to the other. The best fit was associated with an SSE of 0.26 (data not shown).

Although reducing the negative RPEs to 60% of normal yielded a better fit to the empirical schizophrenia data than did the other parameter changes, inspection of **Figure 9** suggests that this result is also sub-optimal. The problem stems from the fact that for the empirical data the response ERN in the 100 and 80% conditions is smaller for the patients than for the control participants, but the feedback ERN in the 50% condition is as large or larger for the patients than for the controls (**Figure 6**). Importantly, because the feedback stimuli in the 50% condition are unpredictable, the feedback ERN in this condition reflects the "true" feedback ERN amplitude free from any learning-related changes. Taken together, these results suggest that whereas the feedback ERN for the patients is relatively normal (as inferred from the 50% condition), the response ERN is reduced (as inferred from the other conditions). By contrast, scaling the RPE produces a main effect on ERN amplitude: Increases and decreases in RPE size result, respectively, in increases and decreases in both the response ERN and feedback ERN amplitudes. Thus these results reflect the optimal solution to the competing constraints of minimizing response ERN amplitude while maximizing feedback ERN amplitude.

We next explored the related possibility that even if the phasic dopamine signal were intact, the neural targets of the dopamine system might nevertheless be insensitive to the signal. To investigate this possibility, we systematically varied the size of the model learning rate parameter from 0 to 200% of normal. Note that in terms of the impact on behavior, scaling the learning rate is formally equivalent to scaling the RPE as the change in Value is proportional to the learning rate times the RPE (see, e.g., Cockburn and Holroyd, 2010). However, ERN amplitude is hypothesized to be related to the size of the RPE rather than to the learning rate (Holroyd and Coles, 2002). Thus this manipulation dissociates indirect changes to ERN amplitude as a consequence of learning vs. direct changes to ERN amplitude as a consequence of the RPE signal itself. **Figure 8** (red line) illustrates that the optimal solution (SSE = 0.07) was associated with a learning rate that was only 13% of the control value. Unlike the previous simulations, this simulation substantially reduced response ERN amplitudes while maintaining high feedback ERN amplitudes (**Figure 9**, red lines). In this case the reduced learning rate impaired the ability of the model to develop strong predictions of trial outcomes, leading to reduced response ERNs and large feedback ERNs that were relatively insensitive to outcome probability.

Finally, we explored the alternative hypothesis that schizophrenia is characterized by a deficit in the representation of the predictive value of response options (Gold et al., 2008; Krawitz

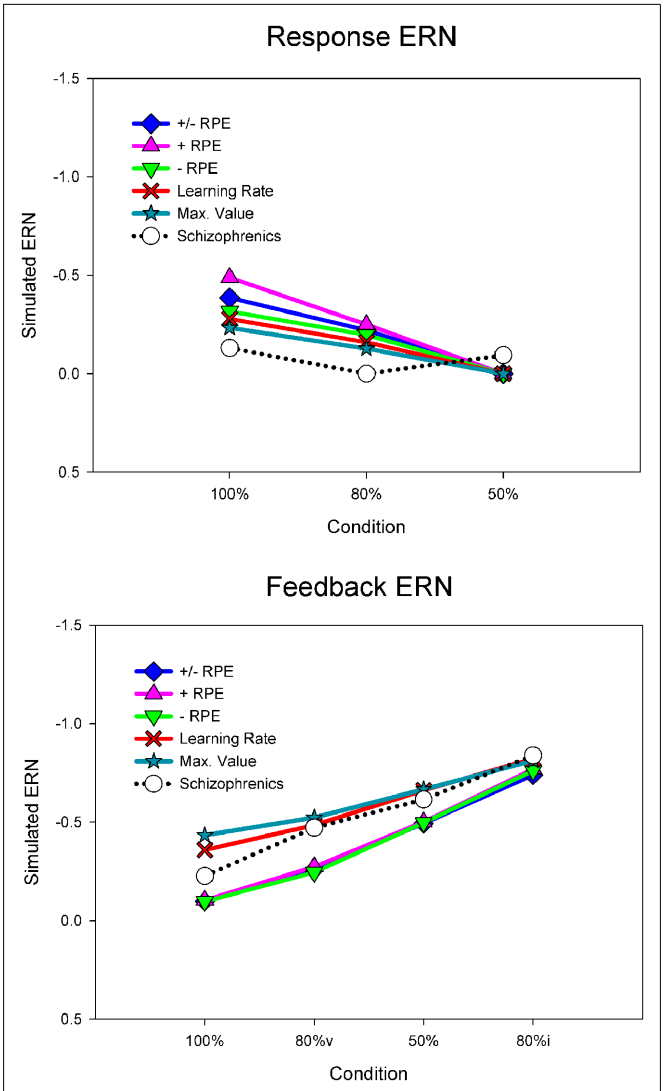


FIGURE 9 | Error-related negativity difference wave data for the probabilistic learning task, for the response ERN (top) and feedback ERN (bottom), for the simulated and empirical schizophrenia data. Note that "80%v" and "80%i" correspond to 80% condition trials with either valid or invalid feedback, respectively. Empirical data are from FCz.

et al., 2011). According to this idea, both the dopamine/RPE signal and the impact of this signal on behavior (i.e., the learning rate) are normal in schizophrenia. Rather, this position holds that schizophrenia compromises the ability to represent the value of different response options for a given imperative stimulus. To explore this possibility, the Values associated with stimulus–response conjunctions in the model, as putatively implemented as a neural target of the DA system in the basal ganglia (Samejima et al., 2005; Lau and Glimcher, 2008), were capped at a maximum. Thus, for these simulations the RPE (dopamine/ERN signal) and the learning rate were unchanged from the control simulation but the stimulus–response conjunction Values were prevented from exceeding the maximum. **Figure 8** (cyan line) illustrates the simulation SSEs when the maximum stimulus–response conjunction Value was systematically varied from 0 to 200% of normal. As can be seen, the simulations were insensitive to increases in the maximum value beyond 100% of normal because the control simulation levels never exceed this range anyway. By contrast, capping the maximum stimulus–response conjunction value at 38% of normal yielded an optimal solution with SSE = 0.08, nearly equivalent to the best SSE associated with the learning rate simulations (**Figure 8**, red line), resulting in similar fashion with relatively small response ERNs and large feedback ERNs that are insensitive to condition (**Figure 9**, cyan lines). A plot of these ERN amplitudes with the simulated control ERN amplitudes (**Figure 7**) reproduces in a qualitative manner the key relationships observed in the empirical data shown in **Figure 6**, namely more negative intercepts and shallower slopes for the simulated patient data relative to the simulated control data (when the response ERN and feedback ERN amplitudes are ordered from smallest to largest).

Flanker task. The above simulations suggest two possibilities: First, that schizophrenia may be associated with a reduced RPE learning rate, and second, that schizophrenia may be characterized by weak representation of the predictive value of stimulus–response conjunctions. Here we simulated ERN amplitudes on the EFT to decide between these two possibilities. For consistency with the PLT simulation and to ensure a high signal-to-noise ratio, we simulated the data of 27 control participants and 27 participants with schizophrenia engaged in the EFT (as opposed to the actual numbers in the empirical experiment, which were 20 and 14, respectively). Imperative stimuli consisted of 54 encounters with each of the six flanker stimuli and no feedback was provided. Interference effects on accuracy and response time induced by the presence of the incompatible flanker stimuli (Eriksen and Eriksen, 1974) were not simulated as these were not central to the hypothesis under investigation (cf. Holroyd et al., 2005). We conducted three simulations: For the first simulation the parameter values were identical to those of the control simulation in the PLT, for the second simulation the learning rate was reduced to its optimal value for the schizophrenia simulation of the PLT (i.e., 13% of that of the control simulation), and for the third simulation the response value was reduced to its optimal value for the schizophrenia simulation of the PLT (i.e., 38% of the control simulation). All other parameter values were held equivalent across simulations. In other words, we explored whether the two optimal models of the PLT schizophrenia data could, without any changes, also account for the EFT schizophrenia data. Note that the same parameter values were used as in the corresponding PLT simulations, but in contrast to the PLT simulations no parameter searches were conducted. This procedure allowed for an unbiased examination of whether the results of the PLT simulations would generalize to the EFT data.

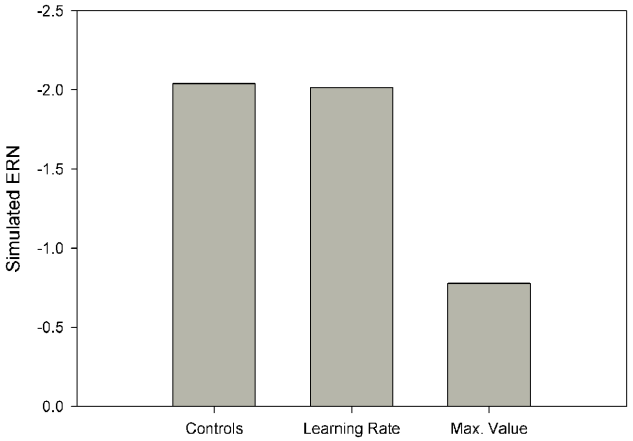


FIGURE 10 | Simulated response ERNs for the Eriksen flanker task.

Figure 10 illustrates the simulation results. As is evident by inspection, reducing the learning rate to 13% of normal did not reduce the simulated response ERN whereas capping the maximum value of stimulus–response conjunctions at 38% of normal induced a commensurate reduction in response ERN amplitude. These results follow for the simple reason that the EFT is not a learning task and so manipulating the learning rate does not affect ERN amplitude. By contrast, capping the stimulus–response conjunction Values at a low level leads to an immediate deficit in response ERN production irrespective of whether the tasks involves feedback or not. These results indicate that a single parameter change to the maximum size of the stimulus–response conjunction Value accounts for the empirical schizophrenia data for both the PLT and the EFT better than competing models involving changes to the RPE signal or the impact of this signal on behavior.

DISCUSSION

Recent empirical and theoretical advances in our understanding about the role of phasic dopamine in schizophrenia have informed contrasting models of the illness which focus on disruption of the RPE signal (Corlett et al., 2007; Frank, 2008; Fletcher and Frith, 2009) and impairment in the representation of response value (Gold et al., 2008). In order to examine these hypotheses, we administered two tasks that elicit the feedback ERN and one that elicits the response ERN, re-analyzed data from a previous study involving both types of ERN, and used a computational instantiation of the RL–ERN theory to aid the interpretation of the empirical results. Taken together, these data show that the feedback ERN is intact in schizophrenia patients despite abnormalities in the response ERN. It should be noted that the tasks used by our group to elicit the feedback ERN include a range

of response–feedback contingencies but regardless of whether the outcomes were determined by handedness of response, accuracy of time estimation or were independent of responding, schizophrenia patients generated feedback negativities that were similar in amplitude to those of control subjects. The tasks used to elicit the feedback ERN included sufficient trials to obtain a reliable feedback ERN component and the reduction in response ERN was detected despite a smaller sample of patients completing this task.

We used computational modeling to examine the hypothesis that the locus of patients' difficulties is in the evaluation of response options rather than feedback processing. Our simulations did indeed suggest that the functioning of the system is selectively disrupted due to weakened representation of response values, presumably encoded in the basal ganglia and orbitofrontal cortex (Gold et al., 2008). The rationale for this conclusion is as follows: Our empirical findings indicate that in schizophrenia the feedback ERN is relatively normal whereas the response ERN is reduced relative to controls. Because manipulations of the dopamine/RPE parameter alter the size of both the response ERN and the feedback ERN concomitantly, these simulations cannot satisfy both of these constraints in the schizophrenia data simultaneously. Further, although reducing the learning rate yields both a reduced response ERN together with a normal feedback ERN, this result occurs only for tasks that actually involve feedback-based learning. By contrast, impairment in the representation of response value leads to smaller response ERNs irrespective of whether the task involves feedback or not.

Thus, according to the model, patients are impaired at associating predictions of future outcomes with particular response options. This result is consistent with the conclusion of Gold et al. (2008) that decision making in patients is compromised by deficits in their ability to represent fully the value of different response options and stimuli predictive of outcomes and with recent fMRI data indicating that schizophrenia patients have difficulty predicting response–outcome associations (Krawitz et al., 2011). The results of this previous study and our present findings converge in that patients' difficulty in predicting response–outcome associations occurs in the presence of spared outcome monitoring. The Krawitz et al. (2011) study, however, did not rule out the possibility that patients' impaired predictions resulted from impaired learning because of a faulty evaluation mechanism. By contrast, our empirical and simulated data indicate that outcome processing is normal in schizophrenia, and thus the impairment lies with the predictive mechanism. This conclusion is also bolstered by a recent report from Gold et al. (in press) that combined behavioral and computational modeling to show that patients were able to use prediction errors to guide learning, but failed to prefer stimuli previously associated with gains over those associated with successful loss avoidance. The gain seeking vs. loss avoidance stimuli were presented at the same probabilistic levels and thus were learned by the same frequency of positive and negative prediction errors. Despite that, patients failed to prefer the gain seeking stimulus suggesting that the deficit appears to be specific to weighing the expected value of alternatives at the time of decision, rather than in processing outcomes *per se*. This formulation would suggest that patients are likely to display alterations in a host of decision making contexts where the relative prospective value of different stimuli and response alternatives must be weighed. Indeed, there is evidence that this is the case as seen in studies of delay discounting (Heerey et al., 2007, 2011) and in the demonstration of reduced transitivity of preferences (Strauss et al., 2011).

Furthermore, impaired response selection in schizophrenia has been associated with increased response times (Luck et al., 2009). Although we did not simulate response times here, previous reinforcement learning models have related response times to the strength of response values (e.g., Suri et al., 2001; Frank et al., 2007). We suggest that the decreased response values implicated by our simulations may also give rise to the increased response times observed in schizophrenia as the system takes longer to decide the appropriate course of action, a direction for future research.

The results of neuroimaging studies of RPEs in schizophrenia are mixed with regard to group differences observed following +RPEs and −RPEs. Koch et al. (2010) found that chronic schizophrenia patients showed relative hypoactivation of frontal areas following positive PEs on a PLT. Waltz et al. (2009), using a primary reinforcer, found diminished hemodynamic response in reward-related circuits following positive RPEs. In two studies (Murray et al., 2008; Koch et al., 2010), schizophrenia patients exhibited significant hypoactivation compared to control subjects following negative RPEs but in other studies (Waltz et al., 2009; Simon et al., 2010) activation following negative RPEs did not differ between patients and healthy comparison subjects. The conflicting results between our findings of normal feedback ERN in schizophrenia and these reports of diminished reward-related activation in schizophrenia are possibly a result of differences among experimental tasks and the inherent difficulty in comparing findings obtained using hemodynamic and electrophysiological methods with their differences in temporal scale and localization properties (Logothetis, 2003). Our empirical and modeling data suggest that future work examining the intersection of stimulus and response would be a fruitful path forward for resolving some remaining questions about the source of the reward processing deficit in schizophrenia.

On the passive gambling task, difference wave peak amplitude was larger following unexpected compared to expected outcomes. The effect of expectedness was not significant, however. This may be because passive guessing tasks which do not require a response produce expectancy effects on feedback ERN amplitude with smaller effect sizes (Holroyd et al., 2009). Other differences in study procedures compared to the Potts et al. (2006) study, including the amount of the bonus and the long duration of our testing session may also have contributed to the reduced expectancy effects in our data.

Consistent with patients' intact feedback ERN, the behavioral data from the time estimation paradigm suggest that patients did effectively make use of negative feedback to improve their subsequent responses. Patients' neural and behavioral sensitivity to negative feedback was surprising in light of substantial evidence of poor incorporation of feedback on a wide range of behavioral tasks, including the Wisconsin Card Sorting Task (e.g., Bryson et al., 2001; Nieuwenstein et al., 2001; Prentice et al., 2008) and the probabilistic weather prediction task (e.g., Weickert et al., 2002; Horan et al., 2008). In contrast to this apparently intact use of negative feedback, patients exhibited impaired ability to sustain accurate responding following positive feedback during

this task. Similar difficulty in adjusting behavior following positive outcomes has been observed in studies using probabilistic selection (Waltz et al., 2007), reversal learning (Weiler et al., 2009), and cued reinforcement (Murray et al., 2008) tasks. This pattern is consistent with the clinical impression that many schizophrenia patients are successful at learning to avoid punishment but do not seek out rewarding activities. Much of the existing ERP literature has focused on the brain's response to negative outcomes, so the ongoing study of neural and behavioral sensitivity to positive feedback holds promise for understanding schizophrenia patients' responsiveness to environmental contingencies.

All of the patient participants in this study were taking APM, so it is important to consider the possible effects of these medications on the DA system and on the results of this study. The results of prior work examining the impact of APM on the ERN are mixed. Acute administration of haloperidol (Zirnheld et al., 2004; de Bruijn et al., 2006) and olanzapine (de Bruijn et al., 2006) to healthy participants diminishes ERN amplitude; however, Kopp and Rist (1999) found that medication dose was unrelated to ERN amplitude in schizophrenia patients and Bates et al. (2004) found that schizophrenia patients' ERNs increased in amplitude following hospital admission and clinical stabilization (presumably involving optimization of APM). Similarly, ACC activity during a response competition task normalized in medication-naïve schizophrenia patients after treatment with atypical APM (Snitz et al., 2005). Functional MRI studies suggest that medication effects on reward-related activity vary depending on the class of APM and the phase of reward processing. Ventral striatal activity in schizophrenia patients taking atypical APMs, but not those taking typical APMs, does not differ from controls during reward anticipation (Juckel et al., 2006; Schlagenhauf et al., 2008). Patients taking typical APMs showed less ventral striatal activity than patients taking atypical APMs during reward anticipation but not following reward receipt (Kirsch et al., 2007). In the current study, it is unlikely that APM caused the feedback ERN to be normalized in schizophrenia since the response ERN was reduced in these patients, although it is arguable that the dissociation that we have documented suggests that APM could have a differential effect on the feedback and response ERN. Alternatively, it is possible that APM does indeed normalize the functioning of the DA system and the mechanisms that generate the error-related ERPs but, as described above, the input into that system (the stimulus–response values) is weak, causing the system to be responsive to external stimuli (feedback) but not internal stimuli.

Our empirical data and modeling results suggest that schizophrenia patients have a deficit in representing the value of actions to be taken in the context of stimuli that offer some type of outcome. Our findings are consistent with Feinberg's (1978) prescient observation that some symptoms of schizophrenia may arise from disturbed efference copy, a duplicate motor command signal used for the purpose of action monitoring (Angel, 1976). This concept has since been refined into computationally specific theories of motor control and their disruption in schizophrenia (e.g., Frith and Done, 1988; Frith et al., 2000). Evidence for this impairment consisted at first of behavioral data indicating abnormal error correction (Malenka et al., 1982, 1986; Frith and Done, 1989) and later of electrophysiological data indicating a reduced response ERN (Kopp and Rist, 1999; Alain et al., 2002; Mathalon et al., 2002; Bates et al., 2004; Morris et al., 2006, 2008) in this population, both of which appear to depend on a neural mechanism for evaluating internally generated motor signals (e.g., Rabbitt, 1966; Gehring et al., 1993; Allain et al., 2004). Our data do not distinguish between a reduction of ERN amplitude in schizophrenia due directly to a weakened representation of response value or indirectly to impaired efference copy giving rise to abnormal valuation. Either way, our results demonstrate that this monitoring impairment is in fact limited to internal sources of performance information and does not extend to external sources of information. Further, our computational modeling of these findings suggests that the impairment may reflect a specific inability to attribute values to behavior – that is, whether a response is "good" or "bad" – and to utilize that information to guide action selection for a specific end (Gold et al., 2008).

ACKNOWLEDGMENTS

This work was supported by the Department of Veterans Affairs Rehabilitation Research and Development Service (grant number D6064W to Sarah E. Morris); and the National Institute of Mental Health (grant numbers 1R01 MH080066, 1R24 MH72647 to James M. Gold).

REFERENCES

Alain, C., McNeely, H. E., He, Y., Christensen, B. L., and West, R. (2002). Neurophysiological evidence of error-monitoring deficits in patients with schizophrenia. *Cereb. Cortex* 12, 840–846.

Allain, S., Hasbroucq, T., Burle, B., Grapperon, J., and Vidal, F. (2004). Response monitoring without sensory feedback. *Clin. Neurophysiol.* 115, 2014–2020.

Andreasen, N. (1982). *The Scale for the Assessment of Negative Symptoms (SANS).* Iowa City: University of Iowa.

Angel, R. W. (1976). Efference copy in the control of movement. *Neurology* 26, 1164–1168.

Baker, T. E., and Holroyd, C. B. (2011). Dissociated roles of the anterior cingulate cortex in reward and conflict processing as revealed by the feedback error-related negativity and N200. *Biol. Psychol.* 87, 25–34.

Bates, A. T., Kiehl, K. A., Laurens, K. R., and Liddle, P. F. (2002). Error-related negativity and correct response negativity in schizophrenia. *Clin. Neurophysiol.* 113, 1454–1463.

Bates, A. T., Liddle, P. F., Kiehl, K. A., and Ngan, E. T. C. (2004). State dependent changes in error monitoring in schizophrenia. *J. Psychiatr. Res.* 38, 347–356.

Bilder, R. M., Volavka, J., Lachman, H. M., and Grace, A. A. (2004). The catechol-O-methyltransferase polymorphism: relations to the tonic-phasic dopamine hypothesis and neuropsychiatric phenotypes. *Neuropsychopharmacology* 29, 1943–1961.

Bryson, G., Whelahan, H., and Bell, M. (2001). Memory and executive function impairments in deficit syndrome schizophrenia. *Psychiatry Res.* 102, 29–37.

Cockburn, J., and Holroyd, C. B. (2010). Focus on the positive: computational simulations implicate asymmetrical reward prediction error signals in childhood attention-deficit/ hyperactivity disorder. *Brain Res.* 1365, 18–34.

Cohen, M. X. (2008). Neurocomputational mechanisms of reinforcement-guided learning in humans: a review. *Cogn. Affect. Behav. Neurosci.* 8, 113–125.

Cohen, M. X., Elger, C. E., and Ranganath, C. (2007). Reward expectation modulates feedback-related negativity and EEG spectra. *Neuroimage* 35, 968–978.

Cohen, M. X., and Frank, M. J. (2009). Neurocomputational models of basal ganglia function in learning, memory and choice. *Behav. Brain Res.* 199, 141–156.

Corlett, P. R., Honey, G. D., Aitken, M. R. F., Dickinson, A., Shanks, D. R., Absalom, A. R., Lee, M., Pomarol-Clotet, E., Murray, G. K., McKenna,

P. J., Robbins, T. W., Bullmore, E. T., and Fletcher, P. C. (2006). Frontal responses during learning predict vulnerability to the psychotogenic effects of ketamine: linking cognition, brain activity, and psychosis. *Arch. Gen. Psychiatry* 63, 611–621.

Corlett, P. R., Honey, G. D., and Fletcher, P. C. (2007). From prediction error to psychosis: ketamine as a pharmacological model of delusions. *J. Psychopharmacol.* 21, 238–252.

Davis, K. L. K., Rene, S., Ko, G., and Davidson, M. (1991). Dopamine in schizophrenia: a review and reconceptualization. *Am. J. Psychiatry* 148, 1474–1486.

Dayan, P., and Daw, N. D. (2008). Decision theory, reinforcement learning, and the brain. *Cogn. Affect. Behav. Neurosci.* 8, 429–453.

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

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

Eriksen, B. A., and Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Percept. Psychophys.* 16, 143–149.

Falkenstein, M. (1990). "Effects of errors in choice reaction tasks on the ERP under focused and divided attention," in *Psychophysiological Brain Research*, eds C. K. Brunia, A. Gaillard, and A. Kok (Tilburg: Tilburg University Press), 192–195.

Feinberg, I. (1978). Efference copy and corollary discharge: implications for thinking and its disorders. *Schizophr. Bull.* 4, 636–640.

First, M., Spitzer, R., Gibbon, M., and Williams, J. (1994). *Structured Clinical Interview for Axis I DSM-IV.* New York: Biometrics Research Department, New State Psychiatric Institute.

Fletcher, P. C., and Frith, C. D. (2009). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nat. Rev. Neurosci.* 10, 48–58.

Frank, M. J. (2008). Schizophrenia: a computational reinforcement learning perspective. *Schizophr. Bull.* 34, 1008–1011.

Frank, M. J., Samanta, J., Moustafa, A. A., and Sherman, S. J. (2007). Hold your horses: impulsivity, deep brain stimulation, and medication in Parkinsonism. *Science* 318, 1309–1312.

Frith, C. D., Blakemore, S., and Wolpert, D. M. (2000). Explaining the symptoms of schizophrenia: abnormalities in the awareness of action. *Brain Res. Rev.* 31, 357–363.

Frith, C. D., and Done, D. J. (1988). Towards a neuropsychology of schizophrenia. *Br. J. Psychiatry* 153, 437–443.

Frith, C. D., and Done, D. J. (1989). Experiences of alien control in schizophrenia reflect a disorder in the central monitoring of action. *Psychol. Med.* 19, 359–363.

Gehring, W. J., Goss, B., Coles, M. G., Meyer, D. E., and Donchin, E. (1993). A neural system for error detection and compensation. *Psychol. Sci.* 4, 385–390.

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

Gold, J. M., Waltz, J. A., Matveeva, T. M., Kasanova, Z., Strauss, G. P., Herbener, E. S., Collins, A. G. E., and Frank, M. J. (in press). Negative symptoms in schizophrenia result from a failure to represent the expected value of rewards: behavioral and computational modeling evidence. *Arch. Gen. Psychiatry*.

Gold, J. M., Waltz, J. A., Prentice, K. J., Morris, S. E., and Heerey, E. A. (2008). Reward processing in schizophrenia: a deficit in the representation of value. *Schizophr. Bull.* 34, 835–847.

Gratton, G., Coles, M. G. H., and Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalogr. Clin. Neurophysiol.* 55, 468–484.

Heerey, E. A., Bell-Warren, K. R., and Gold, J. M. (2008). Decision-making impairments in the context of intact reward sensitivity in schizophrenia. *Biol. Psychiatry* 64, 62–69.

Heerey, E. A., and Gold, J. M. (2007). Patients with schizophrenia demonstrate dissociation between affective experience and motivated behavior. *J. Abnorm. Psychol.* 116, 268–278.

Heerey, E. A., Matveeva, T. M., and Gold, J. M. (2011). Imagining the future: degraded representations of future rewards and events in schizophrenia. *J. Abnorm. Psychol.* 120, 483–489.

Heerey, E. A., Robinson, B. M., McMahon, R. P., and Gold, J. M. (2007). Delay discounting in schizophrenia. *Cogn. Neuropsychiatry* 12, 213–221.

Holroyd, C. B., and Coles, M. G. (2008). Dorsal anterior cingulate cortex integrates reinforcement history to guide voluntary behavior. *Cortex* 44, 548–559.

Holroyd, C. B., and Coles, M. G. H. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychol. Rev.* 109, 679–709.

Holroyd, C. B., Hajcak, G., and Larsen, J. T. (2006). The good, the bad and the neutral: electrophysiological responses to feedback stimuli. *Brain Res.* 1105, 93–101.

Holroyd, C. B., and Krigolson, O. E. (2007). Reward prediction error signals associated with a modified time estimation task. *Psychophysiology* 44, 913–917.

Holroyd, C. B., Krigolson, O. E., Baker, R., Lee, S., and Gibson, J. (2009). When is an error not a prediction error? An electrophysiological investigation. *Cogn. Affect. Behav. Neurosci.* 9, 59–70.

Holroyd, C. B., Krigolson, O. E., and Lee, S. (2011). Reward positivity elicited by predictive cues. *Neuroreport* 22, 249–252.

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

Holroyd, C. B., Yeung, N., Coles, M. G., and Cohen, J. D. (2005). A mechanism for error detection in speeded response time tasks. *J. Exp. Psychol. Gen.* 134, 163–191.

Horan, W. P., Green, M. F., Knowlton, B. J., Wynn, J. K., Mintz, J., and Nuechterlein, K. H. (2008). Impaired implicit learning in schizophrenia. *Neuropsychology* 22, 606–617.

Juckel, G., Schlagenhauf, F., Koslowski, M., Filonov, D., Wustenberg, T., Villringer, A., Knutson, B., Kienast, T., Gallinat, J., Wrase, J., and Heinz, A. (2006). Dysfunction of ventral striatal reward prediction in schizophrenic patients treated with typical, not atypical, neuroleptics. *Psychopharmacology (Berl.)* 187, 222–228.

Kapur, S. (2003). Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. *Am. J. Psychiatry* 160, 13–23.

Kirsch, P., Ronshausen, S., Mier, D., and Gallhofer, B. (2007). The influence of antipsychotic treatment on brain reward system reactivity in schizophrenia patients. *Pharmacopsychiatry* 40, 196–198.

Koch, K., Schachtzabel, C., Wagner, G., Schikora, J., Schultz, C., Reichenbach, J. R., Sauer, H., and Schlosser, R. G. (2010). Altered activation in association with reward-related trial-and-error learning in patients with schizophrenia. *Neuroimage* 50, 223–232.

Kopp, B., and Rist, F. (1999). An event-related brain potential substrate of disturbed response monitoring in paranoid schizophrenic patients. *J. Abnorm. Psychol.* 108, 337–346.

Krawitz, A., Braver, T. S., Barch, D. M., and Brown, J. W. (2011). Impaired error-likelihood prediction in medial prefrontal cortex in schizophrenia. *Neuroimage* 54, 1506–1517.

Lau, B., and Glimcher, P. W. (2008). Value representations in the primate striatum during matching behavior. *Neuron* 58, 451–463.

Logothetis, N. K. (2003). The underpinnings of the BOLD functional magnetic resonance imaging signal. *J. Neurosci.* 23, 3963–3971.

Luck, S. J. (2005). *An Introduction to the Event-Related Potential Technique.* Cambridge, MA: MIT Press.

Luck, S. J., Kappenman, E. S., Fuller, R. L., Robinson, B., Summerfelt, A., and Gold, J. M. (2009). Impaired response selection in schizophrenia: evidence from the P3 wave and the lateralized readiness potential. *Psychophysiology* 46, 776–786.

Malenka, R. C., Angel, R. W., Hampton, B., and Berger, P. A. (1982). Impaired central error-correcting behavior in schizophrenia. *Arch. Gen. Psychiatry* 39, 101–107.

Malenka, R. C., Angel, R. W., Thiemann, S., Weitz, C. J., and Berger, P. A. (1986). Central error-correcting behavior in schizophrenia and depression. *Biol. Psychol.* 21, 263–273.

Marco-Pallares, J., Cucurell, D., Münte, T. F., Strien, N., and Rodriguez-Fornells, A. (2011). On the number of trials needed for a stable feedback-related negativity. *Psychophysiology* 48, 852–860.

Mathalon, D. H., Fedor, M., Faustman, W. O., Gray, M., Askari, N., and Ford, J. M. (2002). Response-monitoring dysfunction in schizophrenia: an event-related brain potential study. *J. Abnorm. Psychol.* 111, 22–41.

Matthysse, S. (1973). Antipsychotic drug actions: a clue to the neuropathology of schizophrenia? *Fed. Proc.* 32, 200–205.

Miller, G. A., Gratton, G., and Yee, C. M. (1988). Generalized implementation of an eye movement correction procedure. *Psychophysiology* 25, 241–243.

Miltner, W., Braun, C. H., and Coles, M. G. H. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task:

evidence for a generic neural system for error detection. *J. Cogn. Neurosci.* 9, 788–798.

Montague, P. R., and Sejnowski, T. J. (1994). The predictive brain: temporal coincidence and temporal order in synaptic learning mechanisms. *Learn. Mem.* 1, 1–33.

Morris, S. E., Heerey, E. A., Gold, J. M., and Holroyd, C. B. (2008). Learning-related changes in brain activity following errors and performance feedback in schizophrenia. *Schizophr. Res.* 99, 274–285.

Morris, S. E., Yee, C. M., and Nuechterlein, K. H. (2006). Electrophysiological analysis of error monitoring in schizophrenia. *J. Abnorm. Psychol.* 115, 239–250.

Murray, G. K., Corlett, P. R., Clark, L., Pessiglione, M., Blackwell, A. D., Honey, G., Jones, P. B., Bullmore, E. T., Robbins, T. W., and Fletcher, P. C. (2008). Substantia nigra/ventral tegmental reward prediction error disruption in psychosis. *Mol. Psychiatry* 13, 267–276.

Nieuwenhuis, S., Nielen, M. M., Mol, N., Hajcak, G., and Veltman, D. J. (2005). Performance monitoring in obsessive-compulsive disorder. *Psychiatry Res.* 134, 111–122.

Nieuwenhuis, S., Ridderinkhof, K. R., Talsma, D., Coles, M. G. H., Holroyd, C. B., Kok, A., and van der Molen, M. W. (2002). A computational account of altered error processing in older age: dopamine and the error-related negativity. *Cogn. Affect. Behav. Neurosci.* 2, 19–36.

Nieuwenstein, M. R., Aleman, A., and de Haan, E. H. (2001). Relationship between symptom dimensions and neurocognitive functioning in schizophrenia: a meta-analysis of WCST and CPT studies. *J. Psychiatr. Res.* 35, 119–125.

Niv, Y. (2009). Reinforcement learning in the brain. *J. Math. Psychol.* 53, 139–154.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., and Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* 304, 452–454.

O'Reilly, R. C., and Farah, M. J. (1999). Simulation and explanation in neuropsychology and beyond. *Cogn. Neuropsychol.* 16, 49–72.

Overall, J. E., and Gorham, D. R. (1962). The brief psychiatric rating scale. *Psychol. Rep.* 10, 799–812.

Potts, G. F., Martin, L. E., Burton, P., and Montague, P. (2006). When things are better or worse than expected: the medial frontal cortex and the allocation of processing resources. *J. Cogn. Neurosci.* 18, 1112–1119.

Prentice, K. J., Gold, J. M., and Buchanan, R. W. (2008). The Wisconsin card sorting impairment in schizophrenia is evident in the first four trials. *Schizophr. Res.* 106, 81–87.

Rabbitt, P. M. (1966). Error correction time without external error signals. *Nature* 212, 438.

Ruchsow, M., Grothe, J., Spitzer, M., and Kiefer, M. (2002). Human anterior cingulate cortex is activated by negative feedback: evidence from event-related potentials in a guessing task. *Neurosci. Lett.* 325, 203–206.

Samejima, K., Ueda, Y., Doya, K., and Kimura, M. (2005). Representation of action-specific reward values in the striatum. *Science* 310, 1337–1340.

Schlagenhauf, F., Juckel, G., Koslowski, M., Kahnt, T., Knutson, B., Dembler, T., Kienast, T., Gallinat, J., Wrase, J., and Heinz, A. (2008). Reward system activation in schizophrenic patients switched from typical neuroleptics to olanzapine. *Psychopharmacology (Berl.)* 196, 673–684.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *J. Neurophysiol.* 80, 1–27.

Schultz, W. (2002). Getting formal with dopamine and reward. *Neuron* 36, 241–263.

Schultz, W., Romo, R., Ljungberg, T., Mirenowicz, J., Hollerman, J. R., and Dickinson, A. (1995). "Reward-related signals carried by dopamine neurons," in *Models of Information Processing in the Basal Ganglia*, eds J. Houk, J. Davis, and D. Beiser (Cambridge, MA: MIT Press), 233–248.

Simon, J. J., Biller, A., Walther, S., Roesch-Ely, D., Stippich, C., Weisbrod, M., and Kaiser, S. (2010). Neural correlates of reward processing in schizophrenia – relationship to apathy and depression. *Schizophr. Res.* 118, 154–161.

Smith, A., Li, M., Becker, S., and Kapur, S. (2006). Dopamine, prediction error and associative learning: a model-based account. *Network* 17, 61–84.

Snitz, B. E., MacDonald, A. III, Cohen, J. D., Cho, R. Y., Becker, T., and Carter, C. S. (2005). Lateral and medial hypofrontality in first-episode schizophrenia: functional activity in a medication-naive state and effects of short-term atypical antipsychotic treatment. *Am. J. Psychiatry* 162, 2322–2329.

Strauss, G. P., Robinson, B. M., Waltz, J. A., Frank, M. J., Kasanova, Z., Herbener, E. S., and Gold, J. M. (2011). Patients with schizophrenia demonstrate inconsistent preference judgments for affective and nonaffective stimuli. *Schizophr. Bull.* 37, 1295–1304.

Suri, R. E. (2002). TD models of reward predictive responses in dopamine neurons. *Neural Netw.* 15, 523–533.

Suri, R. E., Bargas, J., and Arbib, M. A. (2001). Modeling functions of striatal dopamine modulation in learning and planning. *Neuroscience* 103, 65–85.

Sutton, R. S. (1988). Learning to predict by the method of temporal differences. *Mach. Learn.* 3, 9–44.

Sutton, R. S., and Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.

Ursu, S., Kring, A. M., Gard, M. G., Minzenberg, M. J., Yoon, J. H., Ragland, J. D., Solomon, M., and Carter, C. S. (2011). Prefrontal cortical deficits and impaired cognition-emotion interactions in schizophrenia. *Am. J. Psychiatry* 168, 276–285.

Waltz, J. A., Frank, M. J., Robinson, B. M., and Gold, J. M. (2007). Selective reinforcement learning deficits in schizophrenia support predictions from computational models of striatal-cortical dysfunction. *Biol. Psychiatry* 62, 756–764.

Waltz, J. A., Schweitzer, J. B., Gold, J. M., Kurup, P. K., Ross, T. J., Salmeron, B. J., Rose, E. J., McClure, S. M., and Stein, E. A. (2009). Patients with schizophrenia have a reduced neural response to both unpredictable and predictable primary reinforcers. *Neuropsychopharmacology* 34, 1567–1577.

Weickert, T. W., Terrazas, A., Bigelow, L. B., Malley, J. D., Hyde, T., Egan, M. F., Weinberger, D. R., and Goldberg, T. E. (2002). Habit and skill learning in schizophrenia: evidence of normal striatal processing with abnormal input. *Learn. Mem.* 9, 430–442.

Weiler, J. A., Bellebaum, C., Brune, M., Juckel, G., and Daum, I. (2009). Impairment of probabilistic reward-based learning in schizophrenia. *Neuropsychology* 23, 571–580.

Ziauddeen, H., and Murray, G. K. (2010). The relevance of reward pathways for schizophrenia. *Curr. Opin. Psychiatry* 23, 91–96.

Zirnheld, P. J., Carroll, C. A., Kieffaber, P. D., O'Donnell, B. F., Shekhar, A., and Hetrick, W. P. (2004). Haloperidol impairs learning and error-related negativity in humans. *J. Cogn. Neurosci.* 16, 1098–1112.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 28 July 2011; accepted: 10 October 2011; published online: 28 October 2011.

Citation: Morris SE, Holroyd CB, Mann-Wrobel MC and Gold JM (2011) Dissociation of response and feedback negativity in schizophrenia: electrophysiological and computational evidence for a deficit in the representation of value. Front. Hum. Neurosci. ***5****:123. doi: 10.3389/fnhum.2011.00123*

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 01 May 2012
doi: 10.3389/fnhum.2012.00092

Influence of cue exposure on inhibitory control and brain activation in patients with alcohol dependence

Verena Mainz*, Barbara Drüke, Maren Boecker, Ramona Kessel, Siegfried Gauggel and Thomas Forkmann

Institute of Medical Psychology and Medical Sociology, University Hospital of RWTH Aachen University, Aachen, Germany

Edited by:
Patrizia Thoma, Ruhr-University Bochum, Germany

Reviewed by:
Pierre Maurage, University of Louvain, Belgium
Xavier Noel, Fonds de la Recherche Scientifique, Belgium

****Correspondence:***
Verena Mainz, Institute of Medical Psychology and Medical Sociology, University Hospital of RWTH Aachen University, Pauwelsstr. 19, 52074 Aachen, Germany.
e-mail: vmainz@ukaachen.de

Alcohol dependence is a serious condition characterized by persistent desires to drink and unsuccessful efforts to control alcohol consumption despite the knowledge of dysfunction through the usage. The study at hand examined the influence of an alcohol exposure on inhibitory processes. Research provides evidence that trying to resist the temptation to drink exerts self-control, a limited resource which is used during all acts of inhibition. In line with this, studies demonstrate an impaired ability to regulate an already initiated response in alcohol-dependent and healthy subjects when confronted with alcohol-related stimuli. The related neuronal correlates in alcohol-dependent patients remain to be elucidated. The inhibition performance of 11 male alcohol-dependent patients during an alcohol exposure was compared with the task performance during a control condition. Behavioral data and neural brain activation during task performance were acquired by means of functional magnetic resonance imaging. The alcohol cue exposure led to subjectively stronger urges to drink which was accompanied by differential neural activation in amygdala and hippocampus. Moreover, the results revealed typical neural activation during inhibition performance across both conditions. Anyhow, we could not detect any behavioral deficits and only subtle neural differences between induction conditions during the performance of the inhibition task within the inferior frontal cortex. The results suggest that although the sample reports a subjectively stronger urge to drink after the alcohol cue exposure this effect was not strong enough to significantly impair task performance. Coherently, we discover only subtle differential brain activation between conditions during the inhibition task. In opposition to findings in literature our data do not reveal that an exposure to alcohol-related cues and thereby elicited cue reactivity results in impaired inhibition abilities.

Keywords: cue exposure, inhibition, inferior frontal cortex, alcohol dependence

INTRODUCTION

Alcohol dependence is a condition characterizing the concerned person by a "persisting substance use despite clear evidence of overtly harmful consequences" according to the International Classification of Diseases, 10th edition (World Health Organization, 1992). Recent research proposes alcohol dependence could result from an imbalance between increased automatic (e.g., cognitive biases/reactions toward emotionally laden stimuli) and decreased controlled processes (e.g., executive control required for response inhibition; Wiers et al., 2007; Noel et al., 2010). According to a model on the development of alcohol dependence repeated alcohol consumption during adolescence leads to a sensitized appetitive system triggering automatic (drinking) behavior and an underdeveloped regulatory executive system (Wiers et al., 2007). An imbalance of these systems impacts on an outlasting global loss of willpower by affecting reactive mechanisms on direct incentives and reflective mechanisms which moderate impulsive behavior (Noel et al., 2010). Hence alcohol may trigger automatic attentional, memory, and associated emotional systems (bottom-up) which modulate (top-down) goal driven attentional resources needed to reflectively regulate ongoing voluntary behavior. This could explain why patients with alcohol dependence keep up consumption or relapse despite their knowledge about the severe consequences.

Alcohol craving is another central criterion for the diagnosis of alcohol dependence and is described as "a strong desire or sense of compulsion to take alcohol" (World Health Organization, 1992). Craving is thought to contribute to relapse in alcohol-dependent patients through loss of self-control, an ability strongly related to willpower (Rankin et al., 1983; Modell et al., 1992; Anton et al., 1995; Littleton, 1995). Longitudinal studies investigating therapeutic processes and outcomes in alcohol-dependent patients support this relationship (O'Malley et al., 1992; Volpicelli et al., 1992; Paille et al., 1995).

One mechanism that is thought to explain the existence of craving symptoms is based on classical conditioning. The theory postulates that exposure to cues that have been regularly associated with alcohol consumption can elicit conditioned urges to drink alcohol (i.e., craving; Anton, 1999). Alcohol-dependent patients who are trying not to drink must, therefore, expend great effort to overcome such conditioned responses when they are confronted with alcohol-related cues (Brown, 1998; Everitt and Robbins, 2005). In other words, they need much self-control to resist the temptation to drink. The ability to self-control seems, however,

to be a limited resource, which must be used during all acts of inhibition (Muraven et al., 1998; Muraven and Baumeister, 2000; Muraven and Shmueli, 2006). Thus, according to the "resource depletion model" exerting self-control during alcohol craving can reduce drinkers' ability to exert self-control in other realms. Also, according to Tiffany and co-workers' "cognitive processing model of craving" mental processes that are effortful are triggered by situations in which craving is induced and hence may interfere with other cognitive tasks (Tiffany, 1999; Tiffany and Conklin, 2000). The researchers assume that, contrary to classically conditioned responses, some craving induced physiological changes reflect reactions to cognitive demands of certain situations, i.e., representing, if anything, secondarily conditioned effects. Thus, craving effects are not restricted to conditioned responses, a view which offers to investigate data applying broader cognitive psychophysiological models. The model states that performing complex behavior will be guided by automatic and non-automatic processes. Craving is related to the activation of non-automated processes (Tiffany, 1990). It is elicited in situations when attempting to overcome impediments to automated consumption or to avoid the execution of an automatic drug use sequence. The cognitive substrate associated with craving is reflected in behavior, self-report, and autonomic responses visible in alcoholics who try to stay abstinent and those who do not give up consumption. Above from this, being viewed as a non-automated process craving is capacity limiting, hence hindering successful operation of other cognitively demanding processes (Tiffany and Conklin, 2000). As a consequence the ability to resist the urge to drink alcohol and automated drug use is diminished.

There is an ever-growing body of evidence underpinning this limited resource model and the cognitive model of craving by demonstrating that inhibitory performance is impaired when self-control has to be shown. In studies conducted by Muraven and Shmueli (2006) and by Gauggel et al. (2010) craving-related cue reactivity was elicited via a cue exposure paradigm within laboratory settings. In both studies, cue reactivity was induced by letting the participants take a smell of their favorite alcoholic beverage in contrast to exposing the subjects to the smell of a glass of water. Muraven and Shmueli (2006) investigated a sample of 160 social drinkers whereas Gauggel et al. (2010) studied 20 detoxified patients with alcohol dependence. Both studies used the well-established stop-signal paradigm (SSP; Logan, 1994), a task requiring the ability to cancel an already initiated motor response. Results from both studies support the resource depletion model by demonstrating that cue exposure leads to impairment in subsequent self-control tasks such as the SSP. Importantly, the effect size was much larger in the study by Gauggel et al. (2010) than in Muraven and Shmueli's (2006) study, suggesting that detoxified alcohol-dependent patients have even greater inhibitory deficits than social drinkers after exposure to alcohol.

The models and results discussed above underline the importance of exploring inhibitory processing during cue exposure among alcohol-dependent patients. There is ample evidence on impairment of various domains of functioning associated with alcohol dependence. Cognitive deficits and impairments in emotional realms have been widely discussed to be intermingled in both development of alcohol dependence and probability of relapse. Nonetheless differences between investigations and findings maintain the discussion concerning underlying processes and changes leading to alcohol dependence and relapse. Studies differ with respect to characteristics displayed by the investigated patients (e.g., age, onset and duration of illness, comorbidities) or progress related factors (e.g., treatment duration, number of detoxifications). Thus, the study at hand was initiated to further investigate the imbalance between attention consuming reactions to salient stimuli and the functioning of regulatory executive systems in patients with alcohol dependence who display a prolonged consumption history.

Moreover, the neuronal correlates of inhibitory processes in alcohol-dependent patients to whom craving is experimentally induced through cue exposure have not yet been intensively investigated. Therefore, the present study will address neuronal mechanisms and correlates involved in inhibitory processes after alcohol cue exposure.

With regard to brain mechanism associated with drug use and craving there is ample evidence supporting the role of the limbic system (Rodriguez de and Navarro, 1998; Miller and Goldsmith, 2001; Heinz et al., 2010). Even so, brain regions intermingled in cue reactivity and reported as being most relevant may slightly differ across individual studies due to differing imaging techniques, stimuli, and populations. For example, confronting subjects with alcohol-related stimuli like pictures, words, or odors associated with beverages has been related to activation in limbic areas such as the anterior cingulate cortex (ACC), amygdala, hippocampus, and thalamus (George et al., 2001; Schneider et al., 2001; Tapert et al., 2003, 2004; Vollstädt-Klein et al., 2011). It has been proposed that different limbic circuits are important in several specific aspects of alcohol-related reward-signal processing (Rodriguez de and Navarro, 1998). Amygdala and hippocampus are thought to be involved in the remembrance and encoding of significant affective stimuli, the appraisal of the acute emotional state as well as the initiation of responses associated with drug exposure whereas aspects of perceptual and attentional nature are predominantly undertaken and coordinated by prefrontal and cingulate cortices (Rodriguez de and Navarro, 1998; Vollstädt-Klein et al., 2011).

The neural source of inhibitory motor control has been widely studied in the past years by revealing the neural correlates during the performance of the SSP. The inferior frontal cortex (IFC), the pre-supplementary motor area (pre-SMA), and the basal ganglia are discussed as key areas for the "inhibitory control network" in animal, lesion, and fMRI studies (Rubia et al., 1999, 2001, 2003; Aron et al., 2003; Rieger et al., 2003; Gauggel et al., 2004; Aron and Poldrack, 2006; Boecker et al., 2011). Aron et al. (2003) found lesion volume affecting the right IFC (rIFC) to be highly correlated with inhibition performance while Chambers et al. (2006) found that a temporary deactivation of the pars opercularis in the rIFC via transcranial magnetic stimulation (TMS) impairs the ability to inhibit an already initiated action. The role of pre-SMA as a part of the dorsomedial frontal cortex and its relation to the rIFC is still unclear. There is emerging evidence from animal and human studies describing the pre-SMA as "negative motor area" (Aron et al., 2007; Aron, 2011) which generates control signals for specific actions rather than controlling whether or not a movement is made (Scangos and Stuphorn, 2010) while the rIFC is thought to

be responsible for the implementation of inhibitory control (Aron et al., 2007). Evidence for an involvement of the basal ganglia in inhibitory processes comes from studies with patients suffering from Parkinson's disease (Gauggel et al., 2004; van den Wildenberg et al., 2006) and lesion studies with rats where inhibitory processes were requested (Eagle et al., 2008).

Altogether, the present study aims at further extending our knowledge about the impact of cue exposure on self-control and response inhibition by investigating the performance of detoxified patients with alcohol dependence in the SSP during alcohol cue exposure as compared to a neutral exposure condition. Cue exposure was implemented by presenting the smell of the participants' favorite alcoholic beverage. The smell of orange juice was applied as a neutral olfactory control stimulation in order to improve potential limitations of prior investigations (Muraven and Shmueli, 2006; Gauggel et al., 2010).

We expected (1) that participants would report greater subjective craving in the alcohol cue exposure compared to the alcohol–neutral exposure. We further hypothesized that (2) alcohol cue exposure would result in less available self-control for the execution of the SSP which would be indicated in longer reaction time needed to inhibit an initiated response compared to the control condition. (3) During the alcohol cue exposure we expected more metabolic activity in limbic areas (especially thalamus, amygdala, hippocampus, ACC) compared to the alcohol–neutral condition. Finally, we hypothesized that (4) differences in the IFC, which is deemed important for response inhibition as measured with the SSP, would be found between the alcohol cue exposure condition and the alcohol–neutral condition.

MATERIALS AND METHODS

PARTICIPANTS

Fourteen male alcohol-dependent patients, who fulfilled ICD-10 criteria for alcohol dependence (F10.21), were recruited from a collaborating psychiatric hospital in Aachen (Germany) to participate in the present study. All participants were undergoing inpatient treatment for alcohol dependence and were abstinent for at least 1 week. Hence, none of the patients was medicated to reduce withdrawal symptoms. After excluding 3 participants due to technical problems within the scanner environment (e.g., malfunctioning interlinkage between scanner and button or headphone devices), 11 participants (mean age = 44, SD = 10, range 25–54 years) remained in the analyses. Means and SDs of the participants' characteristics are presented in **Table 1**.

All participants gave informed consent and were paid for their participation. The present study was approved by the Ethical Committee of the Medical Faculty of the University Hospital Aachen (EK 096/08) and was conducted according to the Declaration of Helsinki (World Medical Association, 1999). Any patient who met a standard exclusion criterion for MRI investigation (e.g., metallic implants that obscure or interfere with MRI) could not participate in the present study.

DESIGN

The design of the present study was a within subject design, in which we compared an alcohol condition (alcohol cue exposure) with a control condition (alcohol–neutral exposure) on the dependent variables [reported subjective craving, stop-signal reaction time (SSRT) and error rate of the SSP]. The order of conditions was randomized, so that half of the participants received the alcohol exposure before the alcohol–neutral exposure and vice versa.

Table 1 | Participant characteristics.

	N	*M*	SD	Range
DEMOGRAPHIC DATA				
Age (years)	11	44	11	25–54
School (years)	11	10.6	1.9	9–13
PATIENT SHEET				
Beginning of problematic drinking behavior (age)	11	29	11	17–54
No. of ambulant treatments	3	2	0	
No. of inpatient treatments	5	7	9	
FDDA ALCOHOL				
Regular consumption of alcohol since (age)	10	26	8	17–45
Drinks max./day*	11	35	14	25–71
FDDA OTHER DRUGS				
Regular consumption of tobacco since (age)	9	18	3	14–26
Regular consumption of cannabis since (age)	2	18	3	16–20

**To be able to compare different beverages the statements on this were converted to drinks: e.g., 1 l beer = 5 drinks; 1 l wine = 9 drinks.*

PROCEDURE

Pretest

One week prior to the experiment volunteers that met the inclusion criteria were informed about the study contents and procedures. Moreover, participants filled out a volunteers sheet including demographic questions (e.g., age, education), a patient form to assess individual drinking habits and treatment history, the Edinburgh Handedness Inventory (Oldfield, 1971) to ensure only right handed patients would participate and a questionnaire assessing individual drug history ("Fragebogen zur differenzierten Drogenanamnese," FDDA; Grüsser et al., 2004). Here, we were especially interested in participants' average alcohol consumption and their age of first and regular consumption (see **Table 1**). The FDDA documents the consumption onset and duration of other drug use as well (see **Table 1**). Except tobacco and cannabis, which nine respectively two patients consumed regularly no other drug was frequently used in our sample. In addition, participants were asked about their favorite alcohol and about situations in which they usually and never drink alcohol. This information served as guideline for the induction of conditions during the experiment. Therefore, detailed descriptions of the respective situations (including, e.g., sounds, smells, other persons, surroundings) and accompanying subjective feelings, thoughts, and body sensations were documented.

Training phase

Prior to performing the SSP in the fMRI scanner environment, participants performed two training sessions of 5 min each to become familiar with the task (**Figure 1**). The SSP consists of circles or triangles (onset stimuli) to which participants have

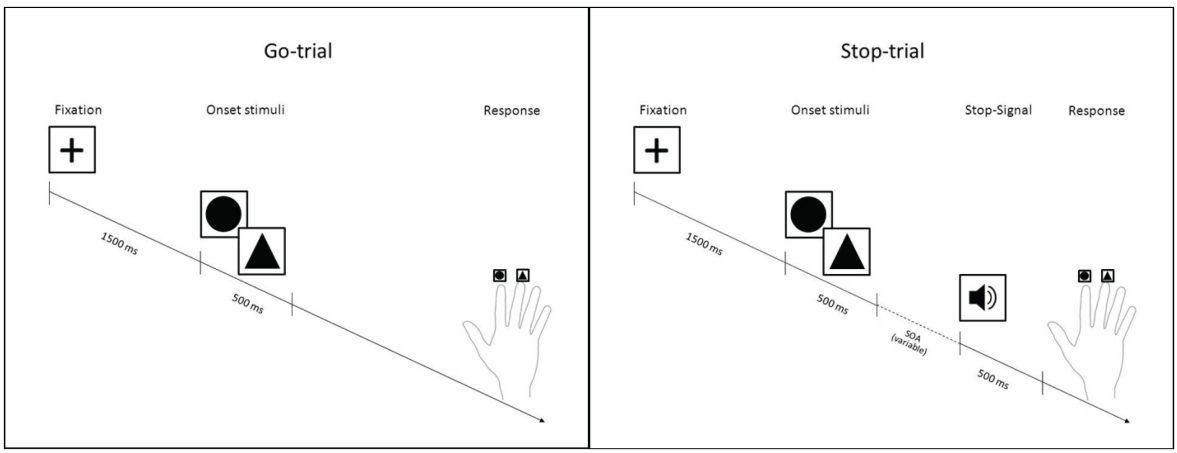


FIGURE 1 | Stop-signal paradigm.

to respond. The paradigm contains three trial types (Go-trials, Stop-trials, and Null-events) each lasting for 3000–3500 ms. A fixation cross is presented in the middle of a gray screen for 1500 ms before each trial. In case of a Null-event, the fixation cross is presented instead of an onset stimulus. In the Go-trials participants perform a simple discrimination task on the two different onset stimuli. They are asked to respond to triangles by pressing a response button with the index finger and to respond to circles by pressing another response button with the middle finger of the right hand. In case of a Stop-trial a Stop-Signal (1000-Hz tone) is presented after the onset stimulus for 500 ms and participants are instructed not to press any button, thus to inhibit their initiated response to the stimulus. Importantly, the delay from onset stimulus to the presentation of the Stop-Signal varies (stimulus onset asynchrony, SOA) according to the staircase-tracking algorithm (Kaernbach, 1991). The SOA is adapted to the participants' responses, in a way that an inhibition rate of 50% is attained. At the beginning, the SOA is set to 250 ms. If a response can be inhibited successfully the SOA is enhanced by 50 ms in the next Stop-trial. If the response can not be inhibited, the SOA is decreased by 50 ms in the next Stop-trial. The SSRT, which is the time participants needed to inhibit an initiated response, is calculated by means of the difference between the mean reaction time on correctly answered Go-trials and the mean Stop-SOA. This measure indicates participants' rate of inhibition controlling for their speed of responses to Go-trials. This difference is important, because people who react slower can inhibit a response more easily than people who react faster on the same SOAs.

Training session 1 aimed at making participants familiar with the discrimination task and therefore consisted of 33 Go-trials and 3 Null-events only. In order to practice the inhibition task, training session 2 consisted of 90 trials (56 Go-trials, 24 Stop-trials, and 10 Null-events). Participants were instructed to react as fast and accurately as possible to the stimuli and not to wait for the Stop-Signal, but to try to inhibit their response whenever the Stop-Signal appeared. They were also informed that they could not always be successful, because the Stop-Signals were adapted according to an algorithm which leads to a success rate of 50%.

At the end of the training phase participants were asked to remember the circumstances under which they would always or never drink alcohol discussed the week before in order to make the individual situations accessible during the standardized induction of the two conditions in the scanner environment. The experimenter repeated the scripts for both conditions in detail in order to facilitate elicitation of accompanying thoughts, feelings, and bodily sensations.

Exposure of alcohol and control conditions

In both alcohol and control condition all participants listened to the same standardized auditory scripts while lying in the scanner environment.

Within the alcohol condition (A), cue reactivity was induced by instructing participants to remember and imagine the situation in which they usually drink alcohol before listening to the standardized script within the scanner environment. In the alcohol condition the script referred to the imagination of the typical place in which the participant would be drinking alcohol including the typical smell and sounds of this place as well as to the imagination of the bodily sensations in this situation and the rising urge to drink alcohol. In addition to the auditory instructions, the smell of participants' favorite alcohol was presented by placing a saturated cloth in the conditioner of the fMRI scanner. This smell of their favorite alcohol was presented every 5 min for 30 s throughout the whole experiment.

In the control condition (N), participants were instructed to imagine the situation, in which they never drink alcohol before listening to a standardized auditory script containing aspects of the place in which this situation would occur as the typical sound and smell. Moreover, it was referred to the bodily sensations in a comfortable situation and the patients were instructed to concentrate on the associated feelings. This time, the smell of oranges was presented to the participants, which again was offered every 5 min for 30 s throughout the whole measurement.

Course of MRI investigation

The subsequent fMRI investigation comprised the accomplishment of two SSPs of about 20 min each intermediated by an acquisition of the brain structure for about 10 min, and a 20-min break. The functioning of the buttons required for the SSP was tested prior to the task execution. Both SSPs comprised 340 trials [210 Go-trials (70%), 90 Stop-trials (30%), and 40 Null-events].

Imminently before the task was performed either the A or N were induced as described above. To assess the individual cue reactivity, two items of the Alcohol Craving Questionnaire (ACQ-Now; Singleton, 1996) were answered while lying in the scanner, once, immediately after the respective induction and another time after the SSP. The statements ("I want to drink so bad I can almost taste it." and "I would feel less restless if I drank alcohol.") had to be rated on a seven-point Likert scale (−3 = strongly disagree to 3 = strongly agree). Higher scores indicate stronger substance craving.

Finally, at the day of the MRI investigation, patients filled out the German version of the Brief Symptom Inventory (BSI, Derogatis and Melisaratos, 1983; Derogatis, 1993; Franke, 2000) in order to acquire the occurrence of comorbidities and symptopathology other than alcohol dependence in the sample. The BSI is a self-report assessment of a patient's symptoms on nine primary dimensions and their intensity at a specific point in time. The BSI provides t-distributed ($M = 50$, SD = 10) norm values for the nine dimensions and moreover a global severity index (GSI) can be calculated which allows to quantify the patient's over all severity-of-illness. **Table 2** provides the information gained from this assessment in our patient sample.

Table 2 | Comorbidities assessed with the Brief Symptom Inventory.

BSI	DEP	ANX	GSI
1	71	74	74
2	59	48	55
3	71	64	67
4	56	64	51
5	59	64	60
6	80	79	72
7	80	80	80
8	43	54	52
9	43	54	37
10	59	48	49
11	80	80	80
Mean*	63.7	64.5	61.5
SD	13.7	12.4	14.1
SEM	4.1	3.7	4.2
Median	59	64	60
Min–Max	43–80	48–80	37–80

**The table displays t-distributed values (M = 50, SD = 10), values >60 indicate high psychological stress. BSI, Brief Symptom Inventory; DEP, depression; ANX, anxiety; GSI, global severity index; SD, standard deviation; SEM, standard error of mean.*

MR TECHNICAL PARAMETERS

Structural and functional MR measurements were acquired at the University hospital of the RWTH Aachen using a 3-T Magneton TRIO TIM MR scanner (Siemens, Erlangen, Germany) with a standard CP head coil. For functional imaging, "Akzent_bold" gradient-echo echoplanar T2*-weighted images (EPI) were acquired [time repetition (TR) = 2400 ms, time echo (TE) = 30 ms, flip angle (FA) = 90°, field of view (FoV) = 220 mm, voxel size (VS) = 2.5 mm × 2.5 mm × 2.5 mm, basis resolution (BR) = 88 mm × 88 mm, slice thickness (ST) = 2.5 mm, 42 axial slices, interleaved slice acquisition].

Anatomical images were acquired using a T1-weighted 3D magnetization-prepared, rapid acquisition gradient echo (MP-RAGE) pulse sequence (TR = 2300 ms, TE = 2.98 ms, time inversion = 900 ms, FA = 9°, VS = 1 mm × 1 mm × 1.2 mm, BR = 256 mm × 256 mm, ST = 1.2 mm, 160 sagittal slices).

The SSP as well as the ACQ statements were presented by means of the software "presentation" on a head mounted display inside the fMRI scanner. Possible visual defects were adjusted with appropriate lenses. The volume of the Stop-Signal was adjusted for every participant, so that each participant was able to hear the tone in a comfortable manner inside the scanner.

STATISTICAL ANALYSES

Behavioral data

Numerical data were analyzed using the software package PASW Statistics 18 (SPSS Inc, 2009) applying analyses of variance (ANOVA) with a within subject factor "condition" (A, N) and parameters of the SSP (RTs, errors, SSRTs in correct Stop- and Go-trials) and the two ACQ rating scores (pre and post SSP) as dependent measures. Moreover, in order to investigate whether the ratings in the ACQ stayed stable across time in the induction conditions, which would indicate that induction effects hold throughout SSP, paired T-tests were calculated between the ACQ ratings pre and post SSP for both conditions.

fMRI data

Functional data were analyzed with SPM5 (Wellcome Trust Centre for Neuroimaging, London, UK; http://www.fil.ion.ucl.ac.uk/spm) implemented in MATLAB 7 (The Mathworks Inc., Natick, MA, USA). Data were realigned, normalized into standard stereotactic space (Talairach coordinates), and smoothed with a Gaussian Kernel of 8 mm (full width half maximum). For each participant and each condition (A and N) the following events were modeled with a canonical hemodynamic response function (HRF): correctly responded Go-trials, successfully inhibited Stop-trials, incorrect Go-trials, and Stop-trials that were responded to.

For the Go-trials the onsets of the events were set to the time of the presentation of the respective onset stimulus. For the Stop-trials the event onsets were modeled at the time of the Stop-Signal due to the high variability of the SOA, hence ensuring a good coverage of activation related to individual response inhibition. Finally, for each subject contrast images were calculated and submitted to a second-level random effects analysis with a within subject factor "condition" referring to A and N assuming measures

to be dependent, and a factor "trial type" referring to the contrasts of the respective Go- and Stop-trials assuming that these are independent measures.

As there is a lot of evidence on the relevant brain regions recruited for Go- and Stop-trials during the performance of the SSP (Rubia et al., 2001; Aron and Poldrack, 2006; Boecker et al., 2011) as well as regions intermingled in craving-associated processes (Miller and Goldsmith, 2001) we conducted specific region of interest (ROI) analyses to test whether the same brain areas would be involved during task performance in the participating subjects in the present investigation as reflected in similar activation patterns.

As reported in Aron and Poldrack (2006) the IFC (BA 47), the sub-thalamic nucleus (STN), the globus pallidus (GP; lateral, medial), and the motor cortex (MC; BA 4, 6) were chosen in the present study as relevant regions for inhibitory processes within the SSP as assessed with the Stop-trials. Based on the assumptions of the same authors the prefrontal cortex (PFC; BA 8, 9, 10, 11, 44, 45, 46, 47), the striatum (S), the GP (lateral, medial), the thalamus (TH), and the MC (BA 4, 6) were selected as areas relevant to response selection, i.e., regions that should be recruited during the Go-trials (Aron and Poldrack, 2006). In order to test whether these typical brain regions would be recruited during the SSP task performance we firstly analyzed the data looking at the total of Stop- and Go-trials across conditions (Stop_A&N; Go_A&N). Hereafter, we calculated the differential contrasts for successful Stop- and Go-trials between the conditions (Stop_A > N; Stop_N > A; Go_A > N, Go_N > A). In all analyses a conservative threshold of $p < 0.05$, corrected for multiple comparisons (FWE) and an extend threshold of >1 voxel were applied.

Moreover, to survey our third hypothesis the left and right limbic lobe as reported in Miller and Goldsmith (2001) served as ROI for the subsequent analyses. Here we chose to inspect the differential activation between the conditions (A > N; N > A) applying conjunction analyses (with $p < 0.05$, uncorrected on voxel level) including all Stop- and Go-trials in order to investigate activation associated with cue reactivity across the complete duration of task performance.

Altogether, all second level contrasts were calculated within these specific ROIs (described above) as defined by the aal-coordinates (Maldjian et al., 2003). Anatomical labeling provided in the tables was performed with help of the aal-coordinates provided by the WFU-Pickatlas (Maldjian et al., 2003).

RESULTS

BEHAVIORAL DATA

Exposure of alcohol and control conditions

The ANOVA showed a significant effect for the within subject factor "condition" [$F(4) = 4.3$; $p < 0.05$; $\eta_p^2 = 0.7$]. **Figure 2** shows that patients indicate a significantly stronger urge to drink alcohol in the A condition as compared to the N condition in both items of the ACQ (all $p < 0.05$) before and after the SSP. **Table 3** summarizes means, SDs, and p-values. Importantly, the results from a subsequent paired T-test showed that the answers before and after task performance within the conditions did not differ indicating that the specific exposure conditions remained stable over time (values of the paired T-tests are shown in **Table 3**).

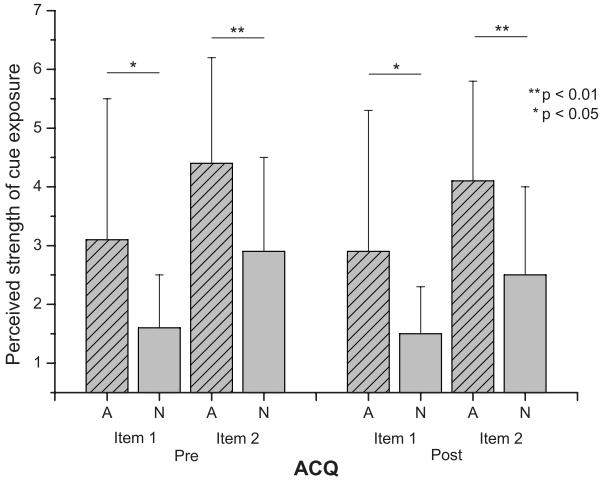


FIGURE 2 | Exposure of alcohol and control conditions.

Table 3 | Exposure of alcohol and control condition.

		A	N	Statistics[1]		
		***M* (SD)**	***M* (SD)**	***F*(1,4)**	***p***	η_p^2
Pre SSP	Item 1	3.1 (2.4)	1.6 (0.9)	8.7	<0.05	0.5
	Item 2	4.4 (1.8)	2.9 (1.6)	21.7	<0.01	0.7
Post SSP	Item 1	2.9 (2.4)	1.5 (0.8)	6.0	<0.05	0.4
	Item 2	4.1 (1.7)	2.5 (1.5)	11.1	<0.01	0.5
		pre SSP	**post SSP**	**Statistics[2]**		
		***M* (SD)**	***M* (SD)**	***T*(10)**	***p***	
Item 1	A	3.1 (2.4)	2.9 (2.4)	0.8	0.44	
	N	1.6 (0.9)	1.5 (0.8)	1.5	0.17	
Item 2	A	4.4 (1.8)	4.1 (1.7)	0.7	0.52	
	N	2.9 (1.6)	2.5 (1.5)	1.5	0.18	

A, alcohol; N, neutral; [1]ANOVA [F(4) = 4.3; p < 0.05; $\eta_p^2 = 0.7$]; [2]Paired T-tests.

Stop-signal paradigm

The ANOVA showed no significant effect for the within subject factor "condition" [$F(9) = 1.3$; $p = 0.52$; $\eta_p^2 = 0.9$]. **Table 4** summarizes means, SDs, and p-values. There were neither significant error- nor significant RT differences in the correct Go-trials and the correctly responded Stop-trials (i.e., trials in which participants pressed the correct button despite the Stop-Signal) between the induction conditions (A and N). Importantly, the SSRT as well did not differ between conditions.

As the occurrence of other symptoms like anxiousness and depressive mood is frequent in patients with alcohol dependence we re-examined our results by including the BSI scores of the depression and anxiety dimensions and the GSI score as covariates in our analysis. This analysis served to look for effects of cue exposure conditions on inhibition performance while keeping the influences of other psychopathology constant. When introducing the BSI scores as covariates *F*-statistics reveal no significant effect of the factor condition on the SSRT [$F(1,7) = 5.4$; $p = 0.053$; $\eta_P^2 = 0.4$]. Hence including the covariates into the model did not change the over all pattern of results. Moreover, this analysis revealed that neither the BSI Depression score [$F(1,7) = 0.00$; $p = 0.9$; $\eta_P^2 = 0.00$], nor the BSI Anxiety score [$F(1,7) = 0.33$; $p = 0.6$; $\eta_P^2 = 0.05$] nor the BSI GSI [$F(1,7) = 0.11$; $p = 0.7$; $\eta_P^2 = 0.02$] had any significant effect on the inhibition performance (SSRT).

Table 4 | Stop-signal paradigm – ANOVA.

***N* = 11**	**A**		**N**		**Statistics[1]**		
	M	**SD**	***M***	**SD**	***F*(1,10)**	***p***	η_P^2
Go-RT	594	97	591	123	0.02	0.88	0.00
Stop-RT*	529	88	519	103	0.31	0.59	0.03
SSRT[2]	231	42	237	48	0.15	0.71	0.02
% Go correct	95	3.1	95	3.9	0.03	0.87	0.00
% Stop correct*	52	2.0	52	2.7	0.12	0.73	0.01
					T(10)	*p*	
Stop-SOA	363	97.4	355	149	0.28	0.78	

*A, alcohol; N, neutral; SOA, stimulus onset asynchrony. *Stop-RT and % Stop correct refer to trials in which subjects correctly reacted despite the Stop-Signal; [1]ANOVA [F(9) = 1.3; p = 0.52; $\eta_P^2 = 0.9$]; paired T-test (two-sided); [2]Difference between mean Go-RT and mean Stop-SOA.*

fMRI DATA

Exposure of alcohol and control conditions

For A > N the conjunction analysis revealed significant differential activation in the left hippocampus (peak voxel at: $x = -30$, $y = -14$, $z = -11$) and the left amygdala (peak voxel at: $x = -26$, $y = -7$, $z = -18$) whereas the inverse contrast (N > A) showed the maximum activation in the posterior cingulate cortex (peak voxel at: $x = -4$, $y = -24$, $z = 27$; $p < 0.05$, uncorrected on voxel level, extend threshold >1 voxel). **Table 5** summarizes all significant coordinates of the conjunction analyses.

Stop-signal paradigm: Stop_A&N, Go_A&N

Across both conditions the ROI analyses revealed the maximum activation within the IFG (BA 13; peak voxel at: $x = 30$, $y = 20$, $z = 6$) during the correct inhibited Stop-trials while the maximum activation was located in the precentral gyrus (BA 4) during the correct responded Go-trials (peak voxel at: $x = -46$, $y = -13$, $z = 52$; $p < 0.05$, FWE corrected, extend threshold >1 voxel). **Table 6** summarizes all significant suprathreshold maxima. **Figure 3** displays the distribution of activations for the respective contrasts.

Stop-signal paradigm: Go_A > N, Go_N > A, Stop_A > N, Stop_N > A

The differential contrasts between the conditions revealed neither for the correct responded Go-trials nor for the successfully inhibited Stop-trials any significant suprathreshold activation ($p < 0.05$, FWE corrected, extend threshold >1 voxel).

However, in order to account for the small behavioral effects of the induction as reported by the patients and to unravel even very small differential activation between the conditions in the Stop-trials, we also compared data between conditions for the Stop-trials on a more lenient threshold ($p < 0.05$ uncorrected with an extent threshold of >1 voxels). Applying

Table 5 | Exposure of alcohol and control condition – conjunction analysis.

k*	***p*****	***Z*****	**Talairach coordinates[1]**			**Region label and (BA)**	**Hemisphere**
			x	***y***	***z***		
A > N[2]							
36	0.003	2.77	−30	−14	−11	Hippocampus	L
54	0.005	2.59	−26	−7	−18	Amygdala	L
N > A[2]							
40	0.003	2.79	−4	−24	27	Cingulate gyrus (BA 23)	L
38	0.005	2.56	8	23	41	Cingulate gyrus (BA 32)	R
67	0.007	2.46	−4	−1	28	Cingulate gyrus (BA 24)	L
4	0.012	2.25	−20	34	15	Anterior cingulate (BA 32)	L
19	0.013	2.21	24	−3	−28	Uncus (BA 36)	R
25	0.015	2.17	18	13	31	Cingulate gyrus (BA 24)	R
15	0.02	2.06	6	−35	39	Cingulate gyrus (BA 31)	R
7	0.034	1.83	−12	−39	39	Cingulate gyrus (BA 31)	L
2	0.049	1.66	38	−24	−22	Parahippocampal gyrus (BA 36)	R

*[1]Talairach coordinates of the voxel of maximal statistical significance; [2]Conjunction across all stop- and go-trials; A, alcohol; N, alcohol–neutral condition; L, left; R, right; *Number of voxels; **p < 0.05, uncorrected; ***Z-score for the voxel of maximum significance.*

Table 6 | Activation in the stop-signal paradigm across conditions.

k*	p**	Z***	Talairach coordinates[1]			Region label and (BA)	Hemisphere
			x	y	z		
Stop_A&N[2]							
1279	0.037	4.26	30	20	6	Inferior frontal gyrus (BA 13)	R
Go_A&N[2]							
503	0.002	5.04	−46	−13	52	Precentral gyrus (BA 4)	L
88	0.010	4.69	−61	−13	41	Precentral gyrus (BA 6)	L

*[1]Talairach coordinates of the voxel of maximal statistical significance; [2]Stop-and Go-trials across both conditions; A, Alcohol, N, alcohol–neutral condition; L, left; R, right; *Number of voxels; **FWE p < 0.05; ***Z-score for the voxel of maximum significance.*

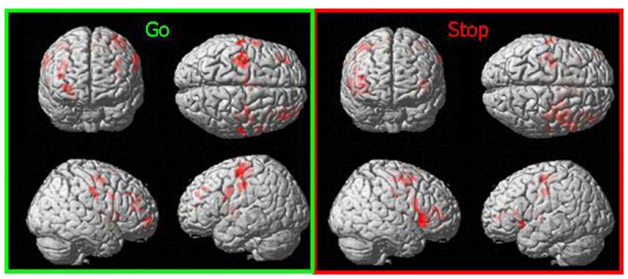


FIGURE 3 | Activation during correct responded Go-trials and successfully inhibited Stop-trials across conditions; SPM5; ROI; $p < 0.05$, corrected for multiple comparisons (FWE).

this threshold revealed most significant differential activation within the IFG (BA 9) for the contrast A > N and within the GP for the inverse contrast (N > A). **Table 7** summarizes all remaining suprathreshold activations and **Figure 4** displays the distribution of activations for the respective *post hoc* contrasts.

DISCUSSION

The resource depletion model (Muraven and Baumeister, 2000) and the cognitive processing model of craving (Tiffany and Conklin, 2000) both postulate that effortful mental processes interfere with other cognitive tasks (e.g., response inhibition) when triggered by situations in which craving is induced and self-control is demanded. Based on these assumptions we examined the effect of an alcohol cue exposure on the ability of detoxified alcohol-dependent patients to inhibit their ongoing responses in a SSP and the associated neural activation.

The finding that alcohol-dependent patients report a stronger urge to drink alcohol when confronted with alcohol-related cues is in line with results reported in previous studies (Schneider et al., 2001; Muraven and Shmueli, 2006; Gauggel et al., 2010). Moreover, looking at the elicited neural activation during the alcohol exposure indicates that typical brain regions were triggered. Amygdala and hippocampus are both key structures within the limbic system, which in turn has been reported to orchestrate stress responses and reward-related aspects within drug abuse and craving processes by incorporating the anatomical requirements for successful accomplishment of emotional as well as motivational tasks (Rodriguez de and Navarro, 1998; Miller and Goldsmith, 2001; Heinz et al., 2010).

Cue reactivity is a learned response that connects a substance and the typical surroundings and context under which consumption takes place. On the neural level amygdala and hippocampus are recruited to correctly remember such situations and the emotions that are associated with the circumstances of drug consumption and experience (Schneider et al., 2001; Goldstein and Volkow, 2002; Weiss, 2005; Heinz et al., 2010). Our results are in accordance with the above assumptions underpinning that the participants were vividly reminded of the respective situation and that alcohol cue exposure involved conditioned emotional responses as mediated by the amygdala and hippocampus.

The finding that the neutral control condition was predominantly associated with elicited activation within the posterior cingulate cortex and not with activation of the amygdala or hippocampus further strengthened the discovery that solely the alcohol cue exposure elicited conditioned emotionally laden cue reactivity responses. While Goldstein and Volkow (2002) found the ACC to be associated with higher order motivational functions and attention processes, such as context dependent tracking, modulating, and updating certain values as a function of the expectation and ability to control and suppress behavior, the general heterogeneity of the functional associations of the cingulate cortex is long known (Vogt et al., 1992). For example, anatomically the anterior cingulate is reciprocally connected to the amygdala whereas the posterior cingulate cortex is not. Researchers describe the role of the ACC in intoxication and craving, and its deactivation during withdrawal but rarely embed the posterior part, a region involved in functions considered to be untypical for the limbic system, in the discussion (Goldstein and Volkow, 2002). More generally, the ACC has been described as a region serving executive functions whenever behavioral and neuroendocrine responses need to be controlled while it is suggested that the posterior part is predominantly involved in assessing context and memory rather than initiating behavioral processes (Vogt et al., 1992). Although the parts of the cingulate cortex certainly are not independent, one might speculate, that the activation found in our investigation during the alcohol–neutral exposure indeed reflects recruitment of environmental or context factors, namely such information that represent situations in which the patients

Table 7 | Differential activation between conditions during correct inhibited Stop-trials.

k*	p**	Z***	Talairach coordinates[1]			Region label and (BA)	Hemisphere
			x	y	z		
Stop_A > N[2]							
327	0.000	3.36	57	5	22	Inferior frontal gyrus (BA 9)	R
224	0.002	2.82	−28	27	−11	Inferior frontal gyrus (BA 47)	L
66	0.004	2.63	14	−18	67	Precentral gyrus (BA 6)	R
13	0.015	2.17	6	−31	72	Paracentral lobe (BA6)	R
20	0.018	2.09	44	−11	56	Precentral gyrus (BA 4)	R
7	0.019	2.07	18	−32	62	Postcentral gyrus (BA 4)	R
8	0.023	2.00	18	−24	66	Precentral gyrus (BA 4)	R
4	0.029	1.90	−48	42	−11	Inferior frontal gyrus (BA 47)	L
2	0.029	1.89	−20	−5	8	Inferior frontal gyrus, GP	L
15	0.030	1.88	28	25	−15	Inferior frontal gyrus (BA 47)	R
14	0.030	1.88	−46	27	−11	Inferior frontal gyrus (BA 47)	L
2	0.033	1.84	−30	38	−9	Middle frontal gyrus (BA 47)	L
2	0.036	1.80	48	−5	9	Precentral gyrus (BA 6)	R
2	0.038	1.77	48	48	−2	Inferior frontal gyrus (BA 10)	R
2	0.045	1.70	61	20	21	Inferior frontal gyrus (BA 45)	R
Stop_N > A[2]							
66	0.001	3.03	18	−8	−3	Lentiform nucleus, GP	R
8	0.001	3.03	18	11	64	Superior frontal gyrus (BA 6)	R
12	0.004	2.67	−28	1	61	Middle frontal gyrus (BA 6)	L
67	0.005	2.59	−18	5	66	Superior frontal gyrus (BA 6)	L
38	0.006	2.53	30	−18	67	Precentral gyrus (BA 6)	R
114	0.011	2.28	−40	9	35	Middle frontal gyrus (BA 9)	L
31	0.012	2.26	20	24	56	Superior frontal gyrus (BA 6)	R
7	0.013	2.23	24	−4	68	Superior frontal gyrus (BA 6)	R
30	0.016	2.15	24	−27	3	Thalamus	R
7	0.016	2.13	57	35	−3	Inferior frontal gyrus (BA 47)	R
21	0.018	2.11	−6	−13	50	Medial frontal gyrus (BA 6)	L
20	0.019	2.08	−6	−9	12	Thalamus, Anterior nucleus	L
18	0.019	2.08	−8	19	62	Superior frontal gyrus (BA 6)	L
4	0.020	2.06	18	−1	9	Lentiform nucleus, GP	R
4	0.027	1.92	−34	−25	49	Precentral gyrus (BA 4)	L
3	0.032	1.85	−57	31	−5	Inferior frontal gyrus (BA 47)	L
3	0.034	1.83	24	−11	4	Lentiform nucleus, GP	R
3	0.034	1.83	48	35	−3	Middle frontal gyrus (BA 47)	R

*[1]Talairach coordinates of the voxel of maximal statistical significance; [2]Correctly inhibited Stop-trials between conditions; A, alcohol; N, alcohol–neutral condition; L, left; R, right; *Number of voxels; **$p < 0.05$, uncorrected; ***Z-score for the voxel of maximum significance.*

would never drink alcohol, but that this condition did not evoke emotionally laden cue reactivity responses similar to the alcohol exposure.

In the study at hand we hypothesized according to the resource depletion model and according to assumptions made by the cognitive processing model of craving (Tiffany, 1999; Muraven and Baumeister, 2000; Tiffany and Conklin, 2000), that the conducted alcohol cue exposure should have resulted in diminished performances in the SSP in patients with alcohol dependence. On the neural level we therefore expected that albeit typical brain regions should be activated during task performance across conditions, a finding strengthening the premise that the paradigm effectively operated, differential neural activation should become evident during the Stop-trials predominantly within the IFC between the two exposure contexts.

On the behavioral level, we could not replicate the findings by Muraven and Shmueli (2006) or Gauggel et al. (2010) who found that, when confronted with alcohol-smell, the performance in the SSP in social drinkers and alcohol-dependent patients is impaired. The participating patient sample in our study did not show any differences between the performances during the alcohol cue exposure and the control condition. RTs, error rates, and most importantly SSRTs did neither differ in the Go- nor in the Stop-trials between conditions. Even when controlling for the influence

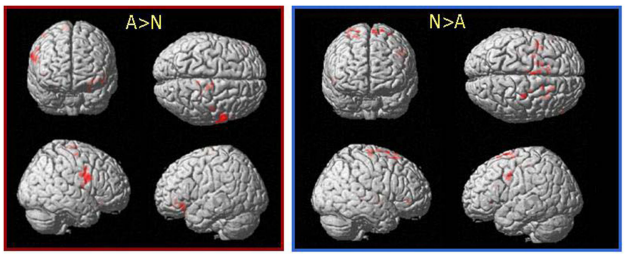


FIGURE 4 | Differential activation between conditions during successfully inhibited Stop-trials; SPM5; ROI; $p < 0.05$, uncorrected on voxel level.

of typical comorbidities in alcohol dependence, like depression and anxiety, the over all pattern of results did not change. Moreover, these findings were accompanied by only subtle differential neural activation in the postulated regions between the two exposure conditions, although across conditions the characteristic neural activation during task performance was elicited (Aron et al., 2003; Aron and Poldrack, 2006; Boecker et al., 2011). Even though the study was designed on the basis of successful experiments which suggested the postulated direction of results and although we believe we could improve the induction of conditions there might be several reasons for the discrepant findings.

First, it could be that the alcohol-dependent subjects could neither perform the task during the alcohol cue exposure nor the neutral condition. When comparing our behavioral results in both conditions with findings by Aron et al. (2003) our participants show longer RTs compared to a healthy sample of a similar age and similar RTs as patients with lesions of the right frontal lobe. Hence, it could be that the patients in our study display a floor effect, i.e., generally slow RTs under both conditions, which impeded an additional slowing of responses through our alcohol cue exposure. However, this argument remains speculative as no healthy control group was included in the study and other research, including our own investigations, found similar RTs in healthy and alcohol-dependent samples (Gauggel et al., 2010; Boecker et al., 2011). Hence, as shown in our previous investigation (Gauggel et al., 2010) we expected the differences between conditions to be strong enough to become evident in a within subject design – a presumption we could not substantiate with our investigation.

A second explanation why the RTs and SSRTs in the Stop-trials between alcohol and neutral cue exposure did not differ could be that the urge to drink alcohol has no influence on performance at all. Although this is a weak assumption as most research, including our own (Gauggel et al., 2010), found stable evidence for a close connection between craving-related processes and impaired (inhibition) performance (Noel et al., 2001, 2007; Fillmore, 2003; Kamarajan et al., 2005), there are studies reporting divergent findings (Bradizza et al., 1995; Townshend and Duka, 2007). For example Bradizza et al. (1995), trying to test Tiffany's predictions that urges to drink alcohol would interfere with performance on cognitive demanding tasks, could not provide support for Tiffany's assumptions. Moreover, Townshend and Duka (2007) found an avoidance of alcohol-related stimuli in alcohol-dependent inpatients in comparison to social drinkers. However, according to incentive salience theories (Robinson and Berridge, 1993) attentional orientation toward alcohol-related cues is an important conditioned response mediating drug-seeking in alcoholic subjects. Moreover, alcohol-relevant cues can increase attention toward alcohol-related stimuli interfering with the processing of other ongoing tasks (Cox et al., 1999). The authors interpreted these findings as evidence that patients become increasingly aware of their inability to control their drinking behavior during therapy leading to an attention withdrawal when confronted with alcohol-related stimuli rather than an attention bias toward the drug-related cues (Townshend and Duka, 2007). This assumption was further supported by the results found in the assessment of craving with a questionnaire providing scores on four factors of alcohol craving. Here, the patients rated their perceived "loss of control over drinking" higher and "mild desires and intentions to drink" lower compared to social drinkers (Townshend and Duka, 2007). Outgoing from the above assumptions one might speculate that our patients actively detracted their attention from the wish to consume alcohol hence leaving the performance unaffected between the conditions. This would go in line with the over all relatively low ratings on the urge to drink alcohol in the ACQ. The SSP is moreover a task which is quite attention consuming and might have facilitated a successful distraction from other processes.

Finally it is possible that the craving intensity or the urge to drink alcohol as elicited with our induction procedure was simply not strong enough to have a significant impact on task performance and thereby lead to different RT and SSRTs between the conditions. After all we do observe differences between the exposure conditions as reflected in a higher self-reported desire to drink in the items of the ACQ after alcohol cue exposure as well as in differential neural activation between the two conditions in brain areas relevant to craving-related processes. Moreover, when applying a more lenient threshold on the neural data for the SSP we find differential activation in the hypothesized brain network between conditions.

During alcohol cue exposure the Stop-trials revealed maximum differential activation within the rIFC which is known to play the key role in the ability to perform tasks where inhibition of ongoing responses is required. Numerous studies on healthy subjects and patients support this assumption (Rubia et al., 2001, 2003; Aron et al., 2003, 2004; Rieger et al., 2003; Aron and Poldrack, 2006; Chambers et al., 2006; Boecker et al., 2011).

The alcohol–neutral exposure on the contrary elicited maximum differential activation within the GP during the Stop-trials. This region is also known to be intermingled in inhibitory processes (Aron and Poldrack, 2006) a fact which was expected as we assumed the patients to try to inhibit their responses under both conditions. Looking at the neuroanatomical connections between regions involved in inhibitory processes initially the STN receives excitatory input from the frontal lobes (IFC; BA4, 6, 8; Mink, 1996; Aron and Poldrack, 2006). The GP as part of the basal ganglia receives excitatory output from its connection to the STN and further sends output projections to thalamus and brainstem (Mink,

1996). Interestingly, applying the more lenient threshold we also find differential activation of the thalamus during the alcohol–neutral condition and not during the alcohol cue exposure. It seems that looking at the neural network of motor response inhibition STN, GP, and thalamus are triggered subsequently after the IFC (Mink, 1996; Aron and Poldrack, 2006). This would support the view that in our patients, when confronted with alcohol-related cues, neuronal the inhibitory processes are affected at a very early level of processing compared to the control condition. One might speculate that the patients put more effort into successful response inhibition under alcohol cue exposure from the very beginning resulting in enhanced activation of the IFC to compensate possible impairments in subsequent mechanisms.

Functionally, it has been shown that a stimulation of the STN improves the SSRT in patients and that lesions of this region led to impaired performance (slower SSRTs; van den Wildenberg et al., 2006; Eagle et al., 2008). Moreover in a study by Aron and Poldrack (2006) the GP was activated during successfully inhibited Stop-trials but not during Stop-trials that were responded to. Both results argue for our finding that the neutral exposure condition during which we expected the patients to be more successful in inhibiting ongoing responses elicited a maximum of activation in the GP, i.e., in parts of the basal ganglia intermingled in successful inhibitory motor responses.

A review on the role of the basal ganglia in motor responses states that the basal ganglia broadly inhibit competing motor mechanisms thereby allowing actions to proceed without interference (Mink, 1996). When intended movement is generated, as for example through the presentation of an onset stimulus in the SSP, motor areas in the cerebral cortex send a signal to the STN leading to an excitation of the GP and a subsequent inhibition of motor pattern generators for competing motions. Moreover, it is described that depending on the movement and involved mechanisms, the number of concurring mechanisms may increase leading to progressive slowing of the actions (Mink, 1996). Hence when during the inhibition of an already initiated response resources are required through processes triggered by an alcohol cue exposure (adding even more competing mechanisms to the desired movement) this should have an impact on the RTs and SSRTs in the SSP and be neuronally reflected by aberrant activation of areas related to the basal ganglia as the GP. As we could not detect any behavioral deficits in our alcohol-dependent sample the subtle neural differences between induction conditions during the performance of the SSP are only coherent.

Altogether the above assumptions remain speculative. The study at hand suffers from some general limitations which could have led to the subtle results in comparison to other studies as the small sample size of patients with alcohol dependence, the absence of a healthy control group and general difficulties in the assessment of subjectively reported craving.

The chosen sample – as in most other investigations – certainly represents a specific population of patients with alcohol dependence showing a characteristic state of personality, severity, and duration of illness. Moreover, the state of detoxification might come along with social desirable responses on questions concerning the triggered urge to drink.

Another general limitation of the study is the absence of a healthy control group which could have helped to explain the diverging RT findings between the investigations of the SSP discussed above. Note however that the study by Gauggel et al. (2010) showed differences in RT between exposure conditions in the SSP in patients with alcohol dependence. This led to our assumption that the exposure conditions would affect task performance and that the effect of an alcohol cue exposure would be strong enough to cause this difference in patients with alcohol dependence rendering a control group unnecessary. A different stimulus selection and study surrounding might have caused that we could not replicate the findings of the preceding study. During the SSP the first investigation presented word stimuli whereas the study at hand asked the participants to respond to symbols instead. Additionally, the specific experimental setting in our investigation (i.e., lying in the scanner environment in comparison to sitting in front of a computer screen) could have contributed to an enhanced arousal or elicited alertness hence interfering with effective exposure procedures and task performances.

Finally, concerning common difficulties in the assessment of subjectively reported craving, it would be interesting to replicate the experiment including more physiological measures (e.g., measurements of heart beat, saliva, and skin conductance responses). This would strengthen findings gained from the subjective reports specifying the urge to drink that was elicited through a cue exposure.

All in all this study investigated the ability of alcohol-dependent patients to inhibit already initiated responses when confronted with alcohol-related cues. Moreover the neuronal correlates during task performance under cue exposure were examined. The subjectively stronger urge to drink was accompanied by activation of limbic brain regions during the alcohol cue exposure compared to the control condition. Moreover, during performance of the SSP typical brain regions were recruited across exposure conditions. The results hint to the direction that alcohol-dependent patients participating in the investigation at hand are able to compensate impairments in inhibitory control induced by an alcohol cue exposure.

REFERENCES

Anton, R. F. (1999). What is craving? Models and implications for treatment. *Alcohol Res. Health* 23, 165–173.

Anton, R. F., Moak, D. H., and Latham, P. (1995). The Obsessive Compulsive Drinking Scale: a self-rated instrument for the quantification of thoughts about alcohol and drinking behavior. *Alcohol. Clin. Exp. Res.* 19, 92–99.

Aron, A. R. (2011). From reactive to proactive and selective control: developing a richer model for stopping inappropriate responses. *Biol. Psychiatry* 69, 55–68.

Aron, A. R., Behrens, T. E., Smith, S., Frank, M. J., and Poldrack, R. A. (2007). Triangulating a cognitive control network using diffusion-weighted magnetic resonance imaging (MRI) and functional MRI. *J. Neurosci.* 27, 3743–3752.

Aron, A. R., Fletcher, P. C., Bullmore, E. T., Sahakian, B. J., and Robbins, T. W. (2003). Stop-signal inhibition disrupted by damage to right inferior frontal gyrus in humans. *Nat. Neurosci.* 6, 115–116.

Aron, A. R., and Poldrack, R. A. (2006). Cortical and subcortical contributions to stop signal response inhibition: role of the subthalamic nucleus. *J. Neurosci.* 26, 2424–2433.

Aron, A. R., Robbins, T. W., and Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends Cogn. Sci. (Regul. Ed.)* 8, 170–177.

Boecker, M., Drueke, B., Vorhold, V., Knops, A., Philippen, B., and Gauggel, S. (2011). When response inhibition is followed by response reengagement: an event-related fMRI study. *Hum. Brain Mapp.* 32, 94–106.

Bradizza, C. M., Lisman, S. A., and Payne, D. G. (1995). A test of Tiffany's cognitive model of drug urges and drug-use behavior. *Alcohol. Clin. Exp. Res.* 19, 1043–1047.

Brown, J. M. (1998). "Self-regulation and the addictive behaviors," in *Treating Addictive Behaviors*, eds W. R. Miller and N. Heather (New York: Plenum Press), 61–73.

Chambers, C. D., Bellgrove, M. A., Stokes, M. G., Henderson, T. R., Garavan, H., Robertson, I. H., Morris, A. P., and Mattingley, J. B. (2006). Executive "brake failure" following deactivation of human frontal lobe. *J. Cogn. Neurosci.* 18, 444–455.

Cox, W. M., Yeates, G. N., and Regan, C. M. (1999). Effects of alcohol cues on cognitive processing in heavy and light drinkers. *Drug Alcohol Depend.* 55, 85–89.

Derogatis, L. R. (1993). *The Brief Symptom Inventory (BSI): Administration, Scoring and Procedures Manual*, 3rd Edn. Minneapolis, MN: National Computer System.

Derogatis, L. R., and Melisaratos, N. (1983). The Brief Symptom Inventory: an introductory report. *Psychol. Med.* 13, 595–605.

Eagle, D. M., Baunez, C., Hutcheson, D. M., Lehmann, O., Shah, A. P., and Robbins, T. W. (2008). Stop-signal reaction-time task performance: role of prefrontal cortex and subthalamic nucleus. *Cereb. Cortex* 18, 178–188.

Everitt, B. J., and Robbins, T. W. (2005). Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nat. Neurosci.* 8, 1481–1489.

Fillmore, M. T. (2003). Drug abuse as a problem of impaired control: current approaches and findings. *Behav. Cogn. Neurosci. Rev.* 2, 179–197.

Franke, G. H. (2000). *Brief Symptom Inventory von L. R. Derogatis (Kurzform der SCL-90-R), Deutsche Version.* Göttingen: Beltz Test GmbH.

Gauggel, S., Heusinger, A., Forkmann, T., Boecker, M., Lindenmeyer, J., Cox, W. M., and Staedtgen, M. (2010). Effects of alcohol cue exposure on response inhibition in detoxified alcohol-dependent patients. *Alcohol. Clin. Exp. Res.* 34, 1584–1589.

Gauggel, S., Rieger, M., and Feghoff, T. A. (2004). Inhibition of ongoing responses in patients with Parkinson's disease. *J. Neurol. Neurosurg. Psychiatr.* 75, 539–544.

George, M. S., Anton, R. F., Bloomer, C., Teneback, C., Drobes, D. J., Lorberbaum, J. P., Nahas, Z., and Vincent, D. J. (2001). Activation of prefrontal cortex and anterior thalamus in alcoholic subjects on exposure to alcohol-specific cues. *Arch. Gen. Psychiatry* 58, 345–352.

Goldstein, R. Z., and Volkow, N. D. (2002). Drug addiction and its underlying neurobiological basis: neuroimaging evidence for the involvement of the frontal cortex. *Am. J. Psychiatry* 159, 1642–1652.

Grüsser, S. M., Wöfling, K., Düffert, S., Mörsen, C., and Flor, H. (2004). Psychometric properties and initial validation of the questionnaire on differentiated assessment of addiction. *Psychother. Psychosom. Med. Psychol.* 54, 405–412.

Heinz, A., Beck, A., Mir, J., Grüsser, S. M., Grace, A. A., and Wrase, J. (2010). "Alcohol craving and relapse prediction. Imaging studies," in *Advances in the Neuroscience of Addiction*, Chap. 4. eds C. M. Kuhn and G. F. Kroob (Boca Raton, FL: CRC Press), 137–162.

Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Percept. Psychophys.* 49, 227–229.

Kamarajan, C., Porjesz, B., Jones, K. A., Choi, K., Chorlian, D. B., Padmanabhapillai, A., Rangaswamy, M., Stimus, A. T., and Begleiter, H. (2005). Alcoholism is a disinhibitory disorder: neurophysiological evidence from a Go/No-Go task. *Biol. Psychol.* 69, 353–373.

Littleton, J. (1995). Acamprosate in alcohol dependence: how does it work? *Addiction* 90, 1179–1188.

Logan, G. D. (1994). "On the ability to inhibit thought and action: a users' guide to the stop-signal paradigm," in *Inhibitory Processes in Attention, Memory and Language*, eds D. Dagenbach and T. H. Carr (San Diego: Academic Press), 189–239.

Maldjian, J. A., Laurienti, P. J., Kraft, R. A., and Burdette, J. H. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *Neuroimage* 19, 1233–1239.

Miller, N. S., and Goldsmith, R. J. (2001). Craving for alcohol and drugs in animals and humans: biology and behavior. *J. Addict. Dis.* 20, 87–104.

Mink, J. W. (1996). The basal ganglia: focused selection and inhibition of competing motor programs. *Prog. Neurobiol.* 50, 381–425.

Modell, J. G., Glaser, F. B., Cyr, L., and Mountz, J. M. (1992). Obsessive and compulsive characteristics of craving for alcohol in alcohol abuse and dependence. *Alcohol. Clin. Exp. Res.* 16, 272–274.

Muraven, M., and Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: does self-control resemble a muscle? *Psychol. Bull.* 126, 247–259.

Muraven, M., and Shmueli, D. (2006). The self-control costs of fighting the temptation to drink. *Psychol. Addict. Behav.* 20, 154–160.

Muraven, M., Tice, D. M., and Baumeister, R. F. (1998). Self-control as limited resource: regulatory depletion patterns. *J. Pers. Soc. Psychol.* 74, 774–789.

Noel, X., Bechara, A., Brevers, D., Verbanck, P., and Campanella, S. (2010). Alcoholism and the loss of willpower: a neurocognitive perspective. *J. Psychophysiol.* 24, 240–248.

Noel, X., Paternot, J., Van der Linden, M., Sferrazza, R., Verhas, M., Hanak, C., Kornreich, C., Martin, P., De Mol, J., Pelc, I., and Verbanck, P. (2001). Correlation between inhibition, working memory and delimited frontal area blood flow measure by 99mTc-Bicisate SPECT in alcohol-dependent patients. *Alcohol Alcohol.* 36, 556–563.

Noel, X., Van der Linden, M., d'Acremont, M., Bechara, A., Dan, B., Hanak, C., and Verbanck, P. (2007). Alcohol cues increase cognitive impulsivity in individuals with alcoholism. *Psychopharmacology (Berl.)* 192, 291–298.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113.

O'Malley, S. S., Jaffe, A. J., Chang, G., Schottenfeld, R. S., Meyer, R. E., and Rounsaville, B. (1992). Naltrexone and coping skills therapy for alcohol dependence. A controlled study. *Arch. Gen. Psychiatry* 49, 881–887.

Paille, F. M., Guelfi, J. D., Perkins, A. C., Royer, R. J., Steru, L., and Parot, P. (1995). Double-blind randomized multicentre trial of acamprosate in maintaining abstinence from alcohol. *Alcohol Alcohol.* 30, 239–247.

Rankin, H., Hodgson, R., and Stockwell, T. (1983). Cue exposure and response prevention with alcoholics: a controlled trial. *Behav. Res. Ther.* 21, 435–446.

Rieger, M., Gauggel, S., and Burmeister, K. (2003). Inhibition of ongoing responses following frontal, nonfrontal, and basal ganglia lesions. *Neuropsychology* 17, 272–282.

Robinson, T. E., and Berridge, K. C. (1993). The neural basis of drug craving: an incentive-sensitization theory of addiction. *Brain Res. Brain Res. Rev.* 18, 247–291.

Rodriguez de, F. F., and Navarro, M. (1998). Role of the limbic system in dependence on drugs. *Ann. Med.* 30, 397–405.

Rubia, K., Overmeyer, S., Taylor, E., Brammer, M., Williams, S. C., Simmons, A., and Bullmore, E. T. (1999). Hypofrontality in attention deficit hyperactivity disorder during higher-order motor control: a study with functional MRI. *Am. J. Psychiatry* 156, 891–896.

Rubia, K., Russell, T., Overmeyer, S., Brammer, M. J., Bullmore, E. T., Sharma, T., and Simmons, A., Williams, S. C., Giampietro, V., Andrew, C. M., and Taylor, E. (2001). Mapping motor inhibition: conjunctive brain activations across different versions of go/no-go and stop tasks. *Neuroimage* 13, 250–261.

Rubia, K., Smith, A. B., Brammer, M. J., and Taylor, E. (2003). Right inferior prefrontal cortex mediates response inhibition while mesial prefrontal cortex is responsible for error detection. *Neuroimage* 20, 351–358.

Scangos, K. W., and Stuphorn, V. (2010). Medial frontal cortex motivates but does not control movement initiation in the countermanding task. *J. Neurosci.* 30, 1968–1982.

Schneider, F., Habel, U., Wagner, M., Franke, P., Salloum, J. B., Shah, N. J., Toni, I., Sulzbach, C., Honig, K., Maier, W., Gaebel, W., and Zilles, K. (2001). Subcortical correlates of craving in recently abstinent alcoholic patients. *Am. J. Psychiatry* 158, 1075–1083.

Singleton, E. G. (1996). Alcohol Craving Questionnaire (ACQ-NOW). *Alcohol Alcohol.* 32, abstr. 344.

SPSS Inc. (2009). *PASW Statistics 18.0.0 [Computer Software]*. Chicago, IL. Available at: www.spss.com

Tapert, S. F., Brown, G. G., Baratta, M. V., and Brown, S. A. (2004). fMRI BOLD response to alcohol stimuli in alcohol dependent young women. *Addict. Behav.* 29, 33–50.

Tapert, S. F., Cheung, E. H., Brown, G. G., Frank, L. R., Paulus, M. P.,

Schweinsburg, A. D., Meloy, M. J., and Brown, S. A. (2003). Neural response to alcohol stimuli in adolescents with alcohol use disorder. *Arch. Gen. Psychiatry* 60, 727–735.

Tiffany, S. T. (1990). A cognitive model of drug urges and drug-use behavior: role of automatic and nonautomatic processes. *Psychol. Rev.* 97, 147–168.

Tiffany, S. T. (1999). Cognitive concepts of craving. *Alcohol Res. Health* 23, 215–224.

Tiffany, S. T., and Conklin, C. A. (2000). A cognitive processing model of alcohol craving and compulsive alcohol use. *Addiction* 95(Suppl. 2), 145–153.

Townshend, J. M., and Duka, T. (2007). Avoidance of alcohol-related stimuli in alcohol-dependent inpatients. *Alcohol. Clin. Exp. Res.* 31, 1349–1357.

van den Wildenberg, W. P., van Boxtel, G. J., van der Molen, M. W., Bosch, D. A., Speelman, J. D., and Brunia, C. H. (2006). Stimulation of the subthalamic region facilitates the selection and inhibition of motor responses in Parkinson's disease. *J. Cogn. Neurosci.* 18, 626–636.

Vogt, B. A., Finch, D. M., and Olson, C. R. (1992). Functional heterogeneity in cingulate cortex: the anterior executive and posterior evaluative regions. *Cereb. Cortex* 2, 435–443.

Vollstädt-Klein, S., Loeber, S., Richter, A., Kirsch, M., Bach, P., von der Goltz, C., Hermann, D., Mann, K., and Kiefer, F. (2011). Validating incentive salience with functional magnetic resonance imaging: association between mesolimbic cue reactivity and attentional bias in alcohol-dependent patients. *Addict. Biol.* doi: 10.1111/j.1369-1600.2011.00352.x. [Epub ahead of print].

Volpicelli, J. R., Alterman, A. I., Hayashida, M., and O'Brien, C. P. (1992). Naltrexone in the treatment of alcohol dependence. *Arch. Gen. Psychiatry* 49, 876–880.

Weiss, F. (2005). Neurobiology of craving, conditioned reward and relapse. *Curr. Opin. Pharmacol.* 5, 9–19.

Wiers, R. W., Bartholow, B. D., van den Wildenberg, E., Thush, C., Engels, R. C., Sher, K. J., Grenard, J., Ames, S. L., and Stacy, A. W. (2007). Automatic and controlled processes and the development of addictive behaviors in adolescents: a review and a model. *Pharmacol. Biochem. Behav.* 86, 263–283.

World Health Organization. (1992). *The ICD-10 Classification of Mental and Behavioral Disorders: Clinical Descriptions and Diagnostic Guidelines.* Geneva: World Health Organization.

World Medical Association. (1999). Proposed revision of the Declaration of Helsinki. *Bull. Med. Ethics* 150, 18–22.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 16 January 2012; accepted: 30 March 2012; published online: 01 May 2012.

Citation: Mainz V, Drüke B, Boecker M, Kessel R, Gauggel S and Forkmann T (2012) Influence of cue exposure on inhibitory control and brain activation in patients with alcohol dependence. Front. Hum. Neurosci. ***6****:92. doi: 10.3389/fnhum.2012.00092*

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 17 April 2012
doi: 10.3389/fnhum.2012.00090

Can executive control be influenced by performance feedback? Two experimental studies with younger and older adults

Barbara Drueke*, Maren Boecker, Verena Mainz, Siegfried Gauggel and Lydia Mungard

Department of Medical Psychology and Medical Sociology, University Hospital of RWTH Aachen University, Aachen, Germany

Edited by:
Patrizia Thoma, Ruhr-University Bochum, Germany

Reviewed by:
Rebecca Elliott, University of Manchester, UK
Michael Falkenstein, Leibniz Research Centre für Working Environment and Human Factors, Germany

****Correspondence:***
Barbara Drueke, Department of Medical Psychology and Medical Sociology, University Hospital of RWTH Aachen University, Pauwelsstr. 30, 52074 Aachen, Germany.
e-mail: bdrueke@ukaachen.de

Executive control describes a wide range of cognitive processes which are critical for the goal-directed regulation of stimulus processing and action regulation. Previous studies have shown that executive control performance declines with age but yet, it is still not clear whether different internal and external factors—as performance feedback and age—influence these cognitive processes and how they might interact with each other. Therefore, we investigated feedback effects in the flanker task in young as well as in older adults in two experiments. Performance feedback significantly improved executive performance in younger adults at the expense of errors. In older adults, feedback also led to higher error rates, but had no significant effect on executive performance which might be due to stronger interference. Results indicate that executive functions can be positively influenced by performance feedback in younger adults, but not necessarily in older adults.

Keywords: age differences, feedback, cognition, executive control

INTRODUCTION

Previous studies have shown that older adults perform poorly in executive control tasks as compared to younger controls (Andrés and van der Linden, 2000; Treitz et al., 2007). These performance deficits in the elderly are usually explained by age-related changes in the brain, especially in the frontal lobes. There is evidence that the prefrontal areas of the brain which are supposed to be involved in executive control are affected more than other parts of the brain during the course of aging (West, 1996; Raz, 2000; Tisserand and Jolles, 2003; Raz and Rodrigue, 2006). This assumption is known as the *frontal hypotheses of cognitive aging* (West, 1996; Raz, 2000). But although aging can accompanied by changes in the brain and cognitive decline, there is a large inter-individual variability due to differential aging effects and compensatory mechanisms (Salthouse, 1996; Deary and Der, 2005; Reuter-Lorenz and Lustig, 2005). For example, Salthouse (1996) found that many of the effects of age on cognition are mediated by age-related variance in processing resource variables.

One factor that has also been shown to modulate cognitive performance in the elderly is feedback. Providing participant's feedback about their performance seems to influence their subsequent performance. The influence of feedback has been studied in a variety of cognitive tasks, including different types of learning, decision-making, memory and meta-memory (e.g., Kulik and Kulik, 1988; Diehl and Sterman, 1995; Thompson, 1998; West et al., 2005; Butler and Roediger, 2008). Meta-analyzes indicate that feedback can have a positive effect on performance ranging from $d = 0.12$ to $d = 1.24$ (Kluger and DeNisi, 1996; Hattie and Timperley, 2007). It is suggested that the feedback effect mainly results from expanding more effort in terms of intensity and persistence (Locke and Latham, 1990). Feedback may thus have an influence on performance by allocating attentional resources to the task which is realized by the so-called executive control system. The question arises whether there is a performance conflict when a feedback intervention is combined with performing a task that also requires a substantial amount of executive control. According to the integrated resource allocation model proposed by Kanfer and Ackerman (1989, 1996), a person's performance is a joint function of his or her relative attentional capacity, task demands, and motivation. It is suggested that motivational interventions (e.g., giving performance feedback) have context-dependent effects on performance by increasing cognitive interference and attentional allocations to the task. The model would thus predict that a task that requires the exertion of executive control would interfere more with the processing of feedback than a task that does not require executive control. Nevertheless, previous studies investigating dual-task performance indicate that feedback has a positive influence on task performance (Kramer et al., 1995, 1999; Bherer et al., 2005, 2008). The findings suggest that despite the existing performance conflict that evolves when combining an executive control task with a feedback intervention, one is still able to profit from feedback. However, it should be noted that in complex tasks, the feedback effect seems to be smaller as compared to more simple tasks (for a review, see Kluger and DeNisi, 1996).

With regard to possible aging effects, it has been shown that performance feedback in memory tasks led to increased performance in older adults (Stadtlander and Coyne, 1990; West et al., 2005, 2009). West and colleagues (2005) demonstrated that objective feedback about the number of items remembered was

sufficient to improve recall in older as well as younger adults. Moreover, feedback led to higher motivation and goal commitment with even stronger effects in older adults. Further evidence for the influence of feedback on performance was found in a time estimation task (Wild-Wall et al., 2009). Wild-Wall and colleagues found that older as well as younger adults had a higher probability to respond correctly after positive feedback as compared to negative feedback. In a recent study, Bherer and colleagues (2008) demonstrated that continuous individualized adaptive feedback led to improvement in dual-task performance in older as well as younger adults. This study indicated that not only memory can be influenced by feedback but also executive functions.

Results suggest that feedback has an impact on the performance of participants in different age groups. However, the effect might be attenuated in older adults as compared to younger adults (West and Thorn, 2001; West et al., 2001). This attenuation might be due to weakened phasic activity of the dopaminergic system in older adults which seems to be involved in feedback processing and the allocation of attentional resources (Nieuwenhuis et al., 2002; Wild-Wall et al., 2009). At first sight, this result (West and Thorn, 2001) seems to be contradictory to the above-mentioned results from West et al. (2005): on the one hand they found an attenuated feedback effect in older adults and on the other hand they found a positive performance change by older adults to a goal-condition that included objective feedback. West et al. (2005) hypothesized a reduced memory self-efficacy may lead the older adults to interpret a neutral or inconsistent feedback as negative which may result in poorer memory performance (West and Thorn, 2001; West et al., 2001).

The aim of the two present experiments was to investigate whether feedback has an influence on executive control performance and whether there are differential aging effects existing. We were interested if possible feedback effects found in younger participants can also be found in a group of elderly participants. As it is still unclear if performance feedback interacts with the degree of executive control or complexity involved in the task, we aimed to investigate in a first experiment if performance in a task involving executive control (i.e., flanker task) can be influenced by performance feedback. Therefore, we examined a group of younger participants with a typical executive control task (i. e., flanker task) and allocated them to a feedback and a no-feedback group, respectively. We hypothesized that feedback would improve task performance in young adults. Furthermore, we expected that feedback would interact with congruent and incongruent trials of the flanker task as they differ in complexity and the demand of executive control. In a second study, a large group of older adults was investigated to replicate the findings of the first study. Here again, we hypothesized feedback to have a positive influence on performance. Such a replication is of importance because aging has been associated with the deterioration of the brain especially in prefrontal areas known to be involved in executive control (e.g., West, 1996; Raz, 2000). As previous literature has shown that feedback in the elderly has an influence on cognitive tasks such as memory (West et al., 2005, 2009) or time estimation tasks (Wild-Wall et al., 2009), it can be hypothesized that feedback would influence executive control performance in the elderly as well. But as the processing of feedback itself requires the exertion of executive control, it is questionable if older adults are able to profit from feedback in an executive control task in the same way as younger adults. As older adults have been reported to use a more cautious criterion than younger adults, i.e., focusing on accuracy to the detriment of speed (Salthouse, 1979; Strayer and Kramer, 1994; Smith and Brewer, 1995), we hypothesized that performance feedback would have an influence on executive control in the elderly, but not at the expense of errors. Still, we expected an attenuated feedback effect in the elderly due to a deficit in allocating attentional resources (Tsang and Shaner, 1998; Nieuwenhuis et al., 2002; Wild-Wall et al., 2009). Since younger and older adults differ in many characteristics, a separate study was performed and analyzed. To enable a comparison of both studies, effect sizes (ES) were reported.

EXPERIMENT 1: INFLUENCE OF PERFORMANCE FEEDBACK ON FLANKER TASK PERFORMANCE IN YOUNGER ADULTS

The goal of the study was to examine context-dependent effects on performance in an executive control task, i.e., to test if performance in a task involving executive control (flanker task) can be influenced by performance feedback. To test this hypothesis we provided positive, negative as well as neutral performance feedback in a flanker task with congruent, incongruent, and neutral trials expecting feedback to interact with task complexity.

METHODS

PARTICIPANTS

A total of 46 young healthy students, 26 males and 20 females, with a mean age of 23.9 years (SD = 3.1) participated in this experiment. Participants were recruited by means of flyers distributed on the university campus. Half of the group performed the feedback version while the other half performed the no-feedback version of the paradigm. The allocation to the respective feedback group was completely randomized and there was no difference in age, sex, or handedness (all participants were right-handed) between both groups. Participants were informed about the objectives and procedure of the present study. The study protocol was approved by the local ethics committee and all subjects gave their written consent, participated voluntarily and were paid a small allowance.

MATERIALS AND DESIGN

A modified version of the flanker task was employed (e.g., Kopp et al., 1996). Participants were required to identify whether a central arrow presented on a computer screen pointed left or right by pressing the equivalent button on the keyboard with their preferred hand. Participants were asked to respond as quickly and accurately as possible. The target arrow was flanked on either side by two arrows in the same direction (congruent condition), or in the opposite direction (incongruent condition). As in the incongruent condition flanking stimuli point to the direction opposite to the target, this condition is more complex and requires more executive control than the congruent condition. In each trial, one central arrow accompanied by four flankers was presented. Targets and flankers appeared simultaneous. The two flanker conditions are depicted in **Figure 1**.

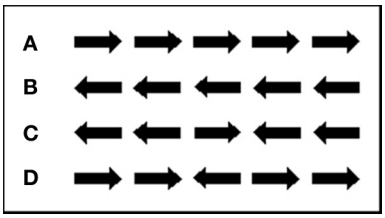


FIGURE 1 | The four target conditions of the flanker task: (A) congruent condition, target right, (B) congruent condition, target left, (C) incongruent condition, target right, (D) incongruent condition, target left.

Participants performed one baseline block followed by nine experimental blocks with 40 trials each, resulting in 360 experimental trials altogether. Half of the trials were congruent, half were incongruent, resulting in a total of 180 congruent and 180 incongruent trials. The ratio of targets pointing to the left and pointing to the right was also balanced.

Participants were randomly allocated into two groups: the feedback group and the no-feedback group. The feedback group received performance feedback which was presented on the computer screen after each block displaying the mean reaction time (RT in milliseconds) of the preceding block of trials. In addition, mean RTs of all preceding blocks were presented to inform participants about the course of their performance. The no-feedback group received no performance feedback. The words "rest period" were presented on the screen after each block. Participants were required to press a button after each block to start the next block of trials.

The stimuli were placed in the center of the screen, subtending a visual angle of 2.86° horizontally and 0.24° vertically. In each trial, a fixation cross was first presented for 900–2100 ms. The target arrow with flankers was then shown up to 2000 ms in the baseline block and for the duration of an individually computed reaction time window in the experimental blocks, respectively. After a response, the fixation cross was presented and the next trial started. An individual response window was calculated for each participant to force speeded responses and to make the task more difficult. The individual response window was determined by adding one standard deviation to the mean reaction time in the baseline block.

PROCEDURE

Participants first completed a health questionnaire after a verbal instruction of the investigator. No participant had to be excluded because of health status and there was no history of neurological or mental disorder. While participants were seated approximately 60 cm in front of a computer screen, the experiment was conducted using the presentation software package (Neurobehavioral Systems, San Francisco, CA). Participants were instructed to respond as quickly and accurately as possible.

Before the flanker task was performed, participants carried out a practice block with 10 trials which they were allowed to repeat until they were familiar with the task. During the practice block, participants received feedback whether their response was correct or incorrect. After each experimental block, one group received feedback about their mean reaction time (feedback group) while the other group received no-feedback (no-feedback group). Total duration of the flanker task was about 20–30 min depending on the individual response window and the duration of self-paced rest periods between the blocks.

STATISTICAL ANALYSIS

For data analysis, only valid trials and trials with a reaction time between 200 and 2000 ms were considered. In addition, an individual outlier analysis was performed. Trials with a reaction time two standard deviations above the condition mean were not considered. For further analysis of error percentage only response errors (i. e., pushing the wrong button) were considered. Omission errors were not included because there were two types of error coded in this variable (no response at all and no response within the reaction time window, respectively). As an additional variable the congruency effect was computed which is a measure of executive control. It is defined as the difference between reaction time or errors in congruent and incongruent trials (Nieuwenhuis et al., 2006). A small difference indicates better conflict resolution and thus better executive control. Two repeated measures ANOVAs with congruency as within-subjects factor and feedback as between-subjects factor were calculated. As dependent variables, reaction times as well as response error percentage were analyzed and Greenhouse-Geisser F-values are reported. Additionally, ES bias-corrected according to Hedges (Hedges and Olkin, 1985) and 95% confidence intervals (CI) were calculated.

RESULTS

REACTION TIMES

Results of the repeated measures ANOVA with reaction times as dependent variable revealed that feedback had a significant influence [$F_{(1, 44)} = 5.35$, $p = 0.025$] on reaction times. Participants receiving feedback showed faster responses ($M = 391$ ms, SD = 28 ms) than participants without feedback ($M = 406$ ms, SD = 22 ms; ES = 0.62, CI = 0.02–1.21). Furthermore, there was also a congruency effect [$F_{(1, 44)} = 298.96$, $p < 0.001$]. As expected, incongruent trials elicited slower responses ($M = 410$, SD = 27) than congruent trials ($M = 384$, SD = 24; ES = 1.01, CI = 0.58–1.44). The Interaction between congruency and feedback was marginally significant [$F_{(1, 44)} = 3.91$, $p = 0.054$]. The congruency effect was smaller in the feedback group ($M = 23$ ms, SD = 12) as compared to the no-feedback group ($M = 29$ ms, SD = 9; ES = 0.56, CI = −0.03–1.14). When calculating the relative congruency effect which considers percental change, results are in line showing a smaller effect for the feedback group (M = 6.1%, SD = 3.0) as compared to the no-feedback group (M = 7.5%, SD = 2.3; ES = 0.50, CI = 0.08–0.91). **Table 1** provides an overview of all variables.

ERROR PERCENTAGE

Analyzes of error percentage as dependent variable showed a significant influence of feedback on errors [$F_{(1, 44)} = 11.16$, $p <$ 0.005], but reversely to reaction times. The feedback group committed relatively more errors ($M = 4.3$ %, SD = 3.3) than the no-feedback group ($M = 1.8$ %, SD = 1.4; ES = 0.97,

Table 1 | Arithmetic mean (M) and standard deviation (SD) of ExFuNet variables for feedback (FB) vs. no-feedback group (noFB) for younger adults.

	FB $N = 23$		No-FB $N = 23$		All participants $N = 46$	
	M	SD	*M*	SD	*M*	SD
Overall RT (ms)	391	28	406	22	399	26
Congruent RT (ms)	377	26	390	20	384	24
Incongruent RT (ms)	400	30	419	20	410	27
Congruency effect (ms)	23	12	29	9	26	11
Relative effect (%)	6.1	3.0	7.5	2.3	6.8	2.8
Response Errors (%)	4.3	1.8	1.4	3.1	3.1	2.8
Congruent Errors (%)	2.1	1.9	0.8	0.7	1.4	1.4
Incongruent Errors (%)	6.5	5.3	2.9	2.5	4.7	4.5
Congruency effect (%)	4.4	4.6	2.1	2.4	3.3	3.8
Overall accuracy (%)	86.2	4.5	87.7	5.1	87.0	4.8

CI = 0.36–1.58). In addition, there was also a congruency effect [$F_{(1,\ 44)} = 37.45$, $p = < 0.001$]. More errors were made during incongruent trials ($M = 4.7$ %, SD = 4.5) than during congruent trials ($M = 1.4$ %, SD = 1.4; ES = 0.98, CI = 0.55–1.41). The interaction between congruency and feedback was also significant [$F_{(1,\ 44} = 4.85$, $p < 0.05$]. Contrary to reaction times, the congruency effect in error percentage was larger in the feedback group ($M = 4.4$ %, SD = 4.6) as compared to the no-feedback group ($M = 2.1$ %, SD = 2.4; ES = 0.62, CI = 0.02–1.21).

EXPERIMENT 2: FEEDBACK EFFECT IN OLDER ADULTS

Experiment 1 provides initial support for the hypothesis of an interaction between task complexity (congruency) and feedback. Experiment 2 was designed to replicate the findings of the first experiment for older participants to detect a possible interaction between aging, task complexity and feedback.

METHODS

PARTICIPANTS

A total of 168 healthy elderly persons, 82 males and 86 females, with a mean age of 70.5 years (SD = 7.1) participated in this experiment. Participants were recruited by a press report in the local newspapers as well as by the means of flyers. They had a mean education of 13.4 years (SD = 3.6). Of all participants, 157 were right-handed, seven were left-handed, and four were ambidexter. Participants were randomly assigned to a feedback and a no-feedback group resulting in 84 participants in each group. Both groups did not differ in age, sex, and handedness. There was a significant difference [$T_{166} = 2.58$, $p < 0.05$; ES = 0.40, CI = 0.09–0.70] in years of education as the no-feedback group had more years of education ($M = 14.1$, SD = 3.7) than the feedback group ($M = 12.7$, SD = 3.4). The difference in education years had no impact on the results obtained as there were no correlations between this variable and performance in the flanker task. All participants were informed about the objectives and procedure of the present study. The study protocol was approved by the local ethics committee and all subjects gave their written consent, participated voluntarily and were paid a small allowance.

MATERIALS AND DESIGN

See Experiment 1

PROCEDURE

See Experiment 1

STATISTICAL ANALYSIS

See Experiment 1

RESULTS

REACTION TIMES

Results reveal that feedback had no significant influence on reaction times [$F_{(1,\ 166)} < 1$, $p = 0.85$]. The slightly faster reaction times of the feedback group ($M = 515$ ms, SD = 51) did not differ from those of the no-feedback group ($M = 523$ ms, SD = 55; ES = 0.15, CI = −0.15–0.45). Congruency had a significant influence [$F_{(1,\ 166)} = 456.8$, $p < 0.001$] on reaction times. As expected, incongruent trials elicited slower responses ($M = 533$ ms, SD = 55) than congruent trials ($M = 507$ ms, SD = 52; ES = 0.48, CI = 0.27–0.70). The interaction between congruency and feedback did not reach significance [$F_{(1,\ 166)} = 1.29$, $p = 0.40$]. There was no difference between the congruency effect in the feedback group ($M = 25$ ms, SD = 15) and the no-feedback group ($M = 28$ ms; SD = 17; ES = 0.19, CI = −0.49–0.12). The same result is obtained when calculating the relative congruency effect (percental change) which also shows no difference between the feedback group ($M = 5.1$%, SD = 3.1) and the no-feedback group ($M = 5.5$%, SD = 3.2; ES = 0.14, CI = −0.07–0.35). **Table 2** provides an overview of all variables.

Furthermore, we calculated a another analysis of variance with the same factors as above and the additional factor "graduation" as the feedback and the no-feedback group differed in their education (see Methods). Results showed a significant main effect of graduation [$F_{(1,\ 166)} = 13.6$, $p < 0.01$] but no significant

Table 2 | Arithmetic mean (M) and standard deviation (SD) of ExFuNet variables for feedback (FB) vs. no-feedback group (noFB) for older adults.

	FB $N = 84$		No-FB $N = 84$		All participants $N = 168$	
	M	SD	*M*	SD	*M*	SD
Overall RT (ms)	515	51	523	55	519	53
Congruent RT (ms)	504	51	510	53	507	52
Incongruent RT (ms)	529	52	538	57	533	55
Congruency effect (ms)	25	15	28	17	26	16
Relative effect (%)	5.0	3.1	5.5	3.2	5.2	3.2
Response Errors (%)	2.5	2.1	1.8	1.8	2.2	2.0
Congruent Errors (%)	1.8	1.9	1.3	1.8	1.6	1.8
Incongruent Errors (%)	3.2	2.9	2.3	2.1	2.8	2.6
Congruency effect (%)	1.4	2.3	1.0	1.7	1.2	2.1
Overall accuracy (%)	93.1	3.9	93.1	4.0	93.1	3.9

interaction effects and the overall pattern of results remains the same.

ERROR PERCENTAGE

Analysis of error percentage showed that feedback had a significant influence on error percentage [$F_{(1, 166)} = 5.3$, $p < 0.05$]. The feedback group ($M = 2.5\%$, SD = 2.1) committed more errors than the no-feedback group ($M = 1.8\%$, SD = 1.8; ES = 0.36, CI = 0.05–0.66). Congruency had also a significant influence on errors [$F_{(1, 166)} = 56.3$, $p < 0.001$]. More errors were committed during incongruent trials ($M = 2.8\%$, SD = 2.6) as compared to congruent trials ($M = 1.6\%$, SD = 1.8; ES = 0.54, CI = 0.32–0.75). Thus, although the feedback group did not significantly profit from feedback regarding reaction times, it showed an increase in errors. The interaction between congruency and feedback was not significant [$F_{(1, 166)} = 1.8$, $p > 0.05$]. There was no difference between the congruency effect in the feedback group ($M = 1.4\%$, SD = 2.3) as compared to the no-feedback group ($M = 1.0\%$, SD = 1.7; ES = 0.20, CI = −0.11–0.50).

GENERAL DISCUSSION

The first study examined the influence of performance feedback on executive control in young adults. Results indicated that feedback had an influence on both reaction times and errors in the flanker task. The feedback group responded faster than the no-feedback group, but this reaction time improvement was at the expense of errors which points to a feedback-induced speed-accuracy trade-off (e.g., Luce, 1986). However, the speed-accuracy trade-off is not surprising as feedback was only provided about reaction times and not about errors. Therefore, participants focused on faster reaction times rather than accuracy. A second important finding was that feedback had a positive influence on executive control performance which was reflected in the smaller congruency effect in reaction times. This finding indicates that although more attentional resources are required to perform the incongruent trials of the task, there is still the possibility of improving the exertion of executive control due to the feedback intervention. As participants focused on reaction times, the better executive control performance was at the expense of errors which was reflected in a higher congruency effect in errors.

Taken together, younger adults were able to adjust their attentional resources accordingly and showed faster responses in flanker task performance as well as a smaller congruency effect. This result is in line with Bherer and colleagues (2008) who investigated the influence of feedback on dual-task performance. Results of their study showed that feedback had an influence on a dual-task despite the fact that the task itself required the exertion of executive control.

In the first experiment it can be inferred that participants receiving feedback on reaction times allocated their attention resources accordingly and focused on speed only. This resulted in a feedback-induced shift in the speed-accuracy trade-off. One could speculate that feedback caused a shift toward a more risky criterion resulting in a higher number of errors. Support for this speculation is provided by a study carried out by Brébion (2001) who demonstrated that the instruction to focus on speed, not on accuracy, led to a shift in response criterion. Because in our study feedback was provided about reaction times, participants focused on speed at the expense of errors which may have resulted in a shift of the response criterion.

According to the integrated resource allocation model (Kanfer and Ackerman, 1989, 1996), a task that requires executive control interferes more with the processing of feedback than a non-executive control task. Results of the present study showed that even in an executive control task such as the flanker task performance feedback had a significant positive influence. The question remains if the feedback effect would have been larger in case a non-executive control task was employed.

In conclusion, it was shown that performance feedback had an impact on the flanker task including the congruency effect which supports the hypothesis that executive control can be positively influenced by performance feedback. In young adults, performance feedback can thus be applied to improve executive control performance.

The second experiment examined if the feedback effects on flanker task performance found in younger adults in Study 1 can be replicated in a group of older adults. Results indicated that feedback had an influence on errors, but not on reaction times. The feedback group committed more errors as compared to the no-feedback group, but did not respond faster. Although participants were not able to increase their reaction times with feedback, the increase in error rates indicates that older adults attempted to regulate their behavior according to the task, but failed in doing so. This might be due to older adults' deficits in allocating attentional resources to the task (Tsang and Shaner, 1998; Nieuwenhuis et al., 2002; Wild-Wall et al., 2009) which requires the exertion of executive control.

Results are in line with previous accounts reporting deficits in executive control performance in older adults (Andrés and van der Linden, 2000; Treitz et al., 2007) and with the notion that especially prefrontal brain areas supposed to be involved in executive control are affected during the course of aging (West, 1996; Raz, 2000; Tisserand and Jolles, 2003; Raz and Rodrigue, 2006).

It can be speculated that older adults already reached their performance limit because of the executive control requirements of the flanker task itself, and failed in speeding up their performance. This result is mirrored by the lack of an interaction between feedback and congruency for reaction times as well as for errors. As older adults reached their resource limit in performing the flanker task, feedback had no further impact on executive control performance in older adults as measured by the congruency effect. Despite helping to improve performance, feedback seems to have distracted participants away from the task. Together with Tsang and Shaner (1998) we speculate that adults experience a decreased flexibility in resource allocation.

Taken together, results support our hypothesis that performance feedback has an influence on flanker task performance in the elderly. However, older adults did not profit from feedback and feedback had no influence on executive control performance. As the flanker task itself required the exertion of executive control, it appears to have interfered with feedback processing resulting in

performance decline. This is in accordance with the integrated resource allocation model (Kanfer and Ackerman, 1989, 1996) which predicts that motivational interventions increase cognitive interference. Results indicate that in older adults, performance feedback cannot be used to improve executive control performance as measured by the flanker task.

The aim of the present studies was to investigate if performance feedback has an impact on executive control and if feedback effects can equally be found in younger as well as older adults. Regarding younger adults, it could be shown that even in a task that requires the exertion of executive control, participants can profit from performance feedback which was shown in faster reaction times. Furthermore, feedback in younger adults had an influence on the congruency effect indicating better executive control regarding reaction times. Thus, it can be inferred that performance feedback in younger adults can be used to influence the exertion of executive control. Younger adults were able to speed up their reaction times after receiving performance feedback although the faster responses were accompanied by higher error rates. It is unlikely that this was due to the difficulty of the executive control task itself as the phenomenon of a speed-accuracy trade-off has been shown for a variety of non-executive control tasks as well (e.g., Kounios et al., 1994; Ratcliff, 2002; Ratcliff and Rouder, 2000; Rinkenauer et al., 2004).

Concerning older adults, we found an influence of feedback on error rates as well. However, the higher error rate was not accompanied by reaction time improvement as in younger adults. Thus, the feedback effect in older adults was attenuated probably due to stronger interference between the executive control task and the feedback intervention as predicted by the integrated resource allocation model (Kanfer and Ackerman, 1989, 1996). It seems most likely that older adults reached their resource limit in performing the flanker task and thus were not able to decrease their reaction times according to the feedback intervention. The fact that the feedback group shows a slight, but insignificant reaction time gain (8 ms; ES = 0.15) supports this interpretation.

Nevertheless, older adults showed a feedback-induced increase of errors which indicates that they tried to focus on improving speed at the expense of errors. Similar results were obtained in the above-mentioned study carried out by Brébion (2001) where it was found that older adults were able to shift their response criterion toward a more risky criterion when instructed to focus on speed only. It was reported that older adults still remained slower and a little more accurate than younger adults. This result could not be attributed to a more cautious strategy which is why it was concluded that older adults have a slower processing system. As previous studies have shown that older adults especially display deficits in executive control performance (Andrés and van der Linden, 2000; Treitz et al., 2007) which might be due to the deterioration of the brain in areas involved in executive control (West, 1996; Raz, 2000; Tisserand and Jolles, 2003; Raz and Rodrigue, 2006), it can be speculated that older adults have a less flexible processing system resulting in difficulties in allocating attentional resources appropriately (Tsang and Shaner, 1998; Nieuwenhuis et al., 2002; Wild-Wall et al., 2009).

It can be argued that there was no significant gain in reaction time in older adults because those already operating on their reaction time limit were not able to further speed up their reaction time. But when dividing the elderly sample into those with relatively fast and those with slow reaction times, no difference can be found regarding the influence of feedback. It can also be claimed that education might have an influence on the ability to profit from feedback as the younger participants in the first study were all students. Therefore, we analyzed a subgroup of elderly participants with a relatively high educational level (at least 12 years of school education) separately revealing the same pattern of results. Another important aspect between the two populations (younger vs. older participants) is their familiarity with playing games on a computer. The younger group might be more familiar with computerized games as many games basically use a structure where feedback is provided and fast responses are required whereas the older adults are likely to spend far less time playing computer games. As we cannot rule out that familiarity with computer games might have an influence on our results this factor should be considered in future studies investigating feedback effects. However, it cannot be ruled out that the low frequency of the feedback intervention (after each block) and the relatively neutral presentation of feedback (reaction times instead of direct negative and positive feedback) were not enough to activate a significant influence of feedback in the elderly. We also cannot exclude the possibility that differences in feedback evaluation may have had an influence on our findings (Kluger and DeNisi, 1996). Some participants may have evaluated the performance feedback as a slightly negative feedback; some might have evaluated the feedback as positive in case their reaction times improved from block to block. Against these arguments remains the fact that younger participants showed significant feedback effects and interactions.

Taken together, it was shown that performance feedback of reaction times had an influence on flanker task performance in younger as well as older adults. While in younger adults a functional feedback effect was found (i.e., faster responses); in older adults the effect was dysfunctional (i.e., no difference in reaction times between the feedback and the no-feedback group). Moreover, feedback had also an influence on the exertion of executive control as measured by the congruency effect in younger adults which indicates that in this age group performance feedback can be used to improve executive control.

It can be concluded that performance feedback not necessarily has a positive influence on executive control performance and that age should be considered when applying feedback interventions. Future studies concerning different sorts of feedback interventions with higher frequencies and stronger valence are needed to clarify the conditions under which older adults may or may not profit from feedback in tasks that require executive control.

FUNDING

This work was supported by the START program (grant number 690709) of the Medical Faculty of RWTH Aachen University, Germany.

REFERENCES

Andrés, P., and van der Linden, M. (2000). Age-related differences in supervisory attentional system functions. *J. Gerontol. Psychol. Sci.* 55, P373–P380.

Bherer, L., Kramer, A. F., Peterson, M. S., Colcombe, S., Erickson, K., and Becic, E. (2005). Training effects on dual-task performance: are there age-related differences in plasticity of attentional control? *Psychol. Aging* 20, 695–709.

Bherer, L., Kramer, A. F., Peterson, M. S., Colcombe, S., Erickson, K., and Becic, E. (2008). Transfer effects in task-set cost and dual-task cost after dual-task training in older and younger adults: further evidence for cognitive plasticity in attentional control in late adulthood. *Exp. Aging Res.* 34, 188–219.

Brébion, G. (2001). Language processing, slowing, and speed/accuracy trade-off in the elderly. *Exp. Aging Res.* 27, 137–150.

Butler, A. C., and Roediger, H. L. III. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Mem. Cogn.* 36, 604–616.

Deary, I., and Der, G. (2005). Reaction time, age and cognitive ability: longitudinal findings from age 16 to 63 years in representative population samples. *Aging Neuropsychol. Cogn.* 12, 187–215.

Diehl, E., and Sterman, J. D. (1995). Effects of feedback complexity on dynamic decision making. *Organ. Behav. Hum. Decis. Process.* 62, 198–215.

Hattie, J., and Timperley, H. (2007). The power of feedback. *Rev. Educ. Res.* 77, 81–110.

Hedges, L. V., and Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. Orlando, FL: Academic Press.

Kanfer, R., and Ackerman, P. L. (1989). Motivation and cognitive abilities: an integrative/aptitude-treatment interaction approach to skill acquisition. *J. Appl. Psychol.* 74, 657–690.

Kanfer, R., and Ackerman, P. L. (1996). "A self-regulatory skills perspective to reducing cognitive interference," in *Cognitive Interference: Theories, Methods, and Findings,* eds I. G. Sarason, B. R. Sarason, and G. R. Pierce (Hillsdale, NJ: Erlbaum), 153–171.

Kluger, A. N., and DeNisi, A. (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychol. Bull.* 119, 254–284.

Kopp, B., Rist, F., and Mattler, U. (1996). N200 in the flanker task as a neurobehavioural tool for investigating executive control. *Psychophysiology* 33, 282–294.

Kounios, J., Montgomery, E. C., and Smith, R. W. (1994). Semantic memory and the granularity of semantic relations: Evidence from speed-accuracy decomposition. *Mem. Cogn.* 22, 729–741.

Kramer, A. F., Larish, J. F., and Strayer, D. L. (1995). Training for attentional control in dual task settings: a comparison of young and old adults. *J. Exp. Psychol. Appl.* 1, 50–76.

Kramer, A. F., Larish, J. L., Weber, T. A., and Bardell, L. (1999). "Training for executive control," in *Attention and Performance XVII,* eds D. Gopher and A. Koriat (Cambridge, MA: MIT Press), 617–652.

Kulik, J. A., and Kulik, C. I. (1988). Timing of feedback and verbal learning. *Rev. Educ. Res. J.* 21, 79–97.

Locke, E. A., and Latham, G. P. (1990). *A Theory of Goal Setting and Task Performance*. Englewood Cliffs, NJ: Prentice-Hall.

Luce, R. D. (1986). *Response Times: Their Role in Inferring Elementary Mental Organization*. Oxford: Oxford University Press.

Nieuwenhuis, S., Ridderinkhof, R., and Talsma, D. (2002). A computational account of altered error processing in older age: dopamine and the error-related negativity. *Cogn. Affect. Behav. Neurosci.* 2, 19–36.

Nieuwenhuis, S., Stins, J. F., Posthuma, D., Polderman, T. J., Boomsma, D. I., and de Geus, E. J. (2006). Accounting for sequential trial effects in the flanker task: conflict adaptation or associative priming? *Mem. Cogn.* 34, 1260–1272.

Ratcliff, R. (2002). A diffusion model account of response time and accuracy in a brightness discrimination task: Fitting real data and failing to fit fake but plausible data. *Psychon. Bull. Rev.* 9, 278–291.

Ratcliff, R., and Rouder, J.-N. (2000). A diffusion model account of masking in two-choice letter identification. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 127–140.

Raz, N. (2000). "Aging of the brain and its impact on cognitive performance: Integration of structural and functional findings," in *Handbook of Aging and Cognition – II,* eds F. I. M. Craik and T. A. Salthouse (Mahwah, NJ: Erlbaum), 1–90.

Raz, N., and Rodrigue, K. R. (2006). Differential aging of the brain: patterns, cognitive correlates and modifiers. *Neurosci. Biobehav. Rev.* 30, 730–748.

Reuter-Lorenz, P. A., and Lustig, C. (2005). Brain aging: reorganizing discoveries about the aging mind. *Curr. Opin. Neurobiol.* 15, 245–251.

Rinkenauer, G., Osman, A., Müller-Gethmann, H., and Mattes, S. (2004). On the locus of speed-accuracy trade-off in reaction time inferences from the lateralized readiness potential. *J. Exp. Psychol. Gen.* 133, 261–282.

Salthouse, T. A. (1979). Adult age and the speed-accuracy trade-off. *Ergonomics* 22, 811–821.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychol. Rev.* 103, 403–428.

Smith, G. A., and Brewer, N. (1995). Slowness and age: speed-accuracy mechanisms. *Psychol. Aging* 10, 238–247.

Stadtlander, L. M., and Coyne, A. C. (1990). The effect of goal-setting and feedback on age differences in secondary memory. *Exp. Aging Res.* 16, 91–94.

Strayer, D. L., and Kramer, A. F. (1994). Aging and skill acquisition: learning-performance distinctions. *Psychol. Aging* 9, 589–605.

Thompson, W. B. (1998). Metamemory accuracy: effects of feedback and the stability of individual differences. *Am. J. Psychol.* 111, 33–42.

Tisserand, D. J., and Jolles, J. (2003). On the involvement of prefrontal networks in cognitive ageing. *Cortex* 39, 1107–1128.

Treitz, F. H., Heyder, K., and Daum, I. (2007). Differential course of executive control changes during normal aging. *Aging Neuropsychol. Cogn.* 14, 370–393.

Tsang, P. S., and Shaner, T. L. (1998). Age, attention, expertise, and time-sharing performance. *Psychol. Aging* 13, 323–347.

West, R. L. (1996). An application of prefrontal cortex function theory to cognitive aging. *Psychol. Bull.* 120, 272–292.

West, R. L., Bagwell, D. K., and Dark-Freudeman, A. (2005). Memory and goal setting: the response of older and younger adults to positive and objective feedback. *Psychol. Aging* 20, 195–201.

West, R. L., Dark-Freudeman, A., and Bagwell, D. K. (2009). Goals-feedback conditions and episodic memory: mechanisms for memory gains in older and younger adults. *Memory* 17, 233–244.

West, R. L., and Thorn, R. M. (2001). Goal-setting, self-efficacy, and memory performance in older and younger adults. *Exp. Aging Res.* 27, 41–65.

West, R. L., Welch, D., and Thorn, R. M. (2001). Effects of goal setting and feedback on memory performance and beliefs among older and younger adults. *Psychol. Aging* 16, 240–250.

Wild-Wall, N., Willemsen, R., and Falkenstein, M. (2009). Feedback-related processes during a time-production task in young and older adults. *Clin. Neurophysiol.* 120, 407–413.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 24 January 2012; accepted: 29 March 2012; published online: 17 April 2012.

Citation: Drueke B, Boecker M, Mainz V, Gauggel S and Mungard L (2012) Can executive control be influenced by performance feedback? Two experimental studies with younger and older adults. Front. Hum. Neurosci. ***6****:90. doi: 10.3389/fnhum.2012.00090*