University of Groningen

Event-related potentials in response to cheating and cooperation in a social dilemma game

Bell, Raoul; Sasse, Julia; Möller, Malte; Czernochowski, Daniela; Mayr, Susanne; Buchner, Axel

Published in:
Psychophysiology

DOI:
[10.1111/psyp.12561](https://doi.org/10.1111/psyp.12561)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2016

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Bell, R., Sasse, J., Möller, M., Czernochowski, D., Mayr, S., & Buchner, A. (2016). Event-related potentials in response to cheating and cooperation in a social dilemma game. *Psychophysiology*, 53(2), 216-228.
<https://doi.org/10.1111/psyp.12561>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Event-related potentials in response to cheating and cooperation in a social dilemma game

RAOUL BELL,^a JULIA SASSE,^b MALTE MÖLLER,^c DANIELA CZERNOCHOWSKI,^d SUSANNE MAYR,^c AND AXEL BUCHNER^a

^aDepartment of Experimental Psychology, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

^bDepartment of Social Psychology, University of Groningen, Groningen, The Netherlands

^cDepartment of Psychology, University of Passau, Passau, Germany

^dDepartment of Cognitive Science, Technical University of Kaiserslautern, Kaiserslautern, Germany

Abstract

A sequential prisoner's dilemma game was combined with psychophysiological measures to examine the cognitive underpinnings of reciprocal exchange. Participants played four rounds of the game with partners who either cooperated or cheated. In a control condition, the partners' faces were shown, but no interaction took place. The partners' behaviors were consistent in the first three rounds of the game, but in the last round some of the partners unexpectedly changed strategies. In the first round of the game, the feedback about a partner's decision elicited a feedback P300, which was more pronounced for cooperation and cheating in comparison to the control condition, but did not vary as a function of feedback valence. In the last round, both the feedback negativity and the feedback P300 were sensitive to expectancy violations. There was no consistent evidence for a negativity bias, that is, enhanced allocation of attention to feedback about another person's cheating in comparison to feedback about another person's cooperation. Instead, participants focused on both positive and negative information, and flexibly adjusted their processing biases to the diagnosticity of the information. This conclusion was corroborated by the ERP correlates of memory retrieval. Successful retrieval of a partner's reputation was associated with an anterior positivity between 400 and 600 ms after face onset. This anterior positivity was more pronounced for both cooperator and cheater faces in comparison to control faces. The results suggest that it is not the negativity of social information, but rather its motivational and behavioral relevance that determines its processing.

Descriptors: Face learning, Face recollection, P300, Feedback negativity, Social neuroscience

The present study examines the processing of social information. Humans are constantly interacting with other human beings. To avoid being exploited by other people, it is important to know whom to trust and whom not to trust. Given that the best predictor of future behavior is past behavior, it is important for people to pay attention to, and keep track of, others' behaviors. A common theory is that the processing of social information is characterized by a negativity bias. According to this view, others' cheating and nonreciprocation should receive more attention than others' cooperative and altruistic actions. There are two broad approaches for explaining negativity effects in social information processing.

According to the first account (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001), human brains are hardwired to attend to negative information because failing to attend to negative infor-

mation has more severe consequences to an individual's fitness than ignoring positive information. Cheating in particular is seen as a threat to the individual and to society because it tears apart the fabric of social norms and cooperation (Gächter & Thöni, 2005). According to social contract theory (Cosmides & Tooby, 1989, 1992), social cooperation cannot evolve unless specialized neurocognitive mechanisms help cooperators to detect and identify individuals who fail to reciprocate. This theory implies that cheating and nonreciprocation receive enhanced processing in specialized brain modules. Given that delayed reciprocal exchange often relies on memory, it has been postulated that memory for cheaters should be enhanced (Buchner, Bell, Mehl, & Musch, 2009; Mealey, Daood, & Krage, 1996).

According to a second account, it is not the negativity of a stimulus, but rather its motivational and behavioral relevance that determines the cognitive processing of the stimulus. Negative information is prioritized only when it is more diagnostic than positive information (Skowronski & Carlston, 1987). From a functional point of view, reciprocal cooperation requires discrimination between cooperators and cheaters, but this task may be accomplished by attending to both cooperators and cheaters. Several models of reciprocal cooperation (e.g., partner choice models) imply

We thank Ulrike Körner, Marie Luisa Schaper, Lara Schlünder, Sabrina Röttger, and Daniel Steinmann for their help with data collection and/or data analysis. The research reported in this article was supported by a grant from the Deutsche Forschungsgemeinschaft (BE 4311/1-1).

Address correspondence to: Raoul Bell, Department of Experimental Psychology, Heinrich-Heine-University Düsseldorf, 40225 Düsseldorf, Germany. E-mail: Raoul.Bell@hhu.de

that the information about an interactant's identity is stored together with a preference value (reflecting the outcome of previous interactions), but do not imply that cheating receives extra attention (e.g., Hruschka & Henrich, 2006). According to this latter account, cheating only receives special attention when it is more diagnostic of another person's disposition or more relevant to the participants' goals (Barclay, 2008; Bell, Buchner, & Musch, 2010; Kroneisen, Woehe, & Rausch, 2015).

Behavioral studies have produced a set of inconclusive results. There are a number of studies demonstrating a cheater detection effect in the Wason selection task (Cosmides & Tooby, 1989, 1992), but others have attributed these findings to relevance (Carlisle & Shafir, 2005; Sperber, Cara, & Girotto, 1995). Furthermore, there is evidence of enhanced memory for the association between a cheater's face and norm-violating behavior (Buchner et al., 2009; Chiappe et al., 2004), but the enhancement might be due to the increased informational value of cheating when cheating violates the participants' expectations (Barclay, 2008; Bell, Buchner, & Musch, 2010; Kroneisen et al., 2015; Suzuki & Suga, 2010). The cheating advantage may disappear when cheating and cooperation are equally expected (Bell, Buchner, & Musch, 2010). Given this state of affairs, the goal of the present study is to use psychophysiological data to provide additional information about the processing of cheating and cooperation in social dilemmas in an attempt to disambiguate the available evidence.

So far, a few studies have used psychophysiological measures to understand decision making in social dilemma games (Boksem & De Cremer, 2010; Polezzi et al., 2008; Wang et al., 2013). This literature is guided by the assumption that the feedback about another person's cheating or cooperation elicits ERPs from the same brain systems that monitor outcomes in other performance domains (see also Sanfey, 2007; Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003). Wang and colleagues (2013) have postulated that two ERP components should be particularly sensitive to the valence and motivational relevance of outcomes in social dilemma games: the feedback negativity and the feedback P300. The feedback negativity is a negative deflection in the ERP at anterior and central electrodes between 250 and 300 ms after feedback onset. It is believed to be generated in the anterior cingulate cortex (ACC) as a response to the conflict between a desired and an actual outcome (Holroyd & Coles, 2002). Its amplitude is enhanced when participants fail to achieve a desired feedback (Hajcak, Moser, Holroyd, & Simons, 2006; Wang et al., 2013; Wu & Zhou, 2009). The second ERP component reflecting evaluations of feedback is the feedback P300, a broadly distributed, positive deflection in the ERP between 300 and 600 ms after feedback onset with a maximum over medial central and medial posterior electrodes. Its generation probably involves different brain structures including the ACC and the parietal cortex (Linden, 2005). The P300 is often thought to reflect the amount of attentional resources devoted to information about rewards and losses (Wu & Zhou, 2009). This component may be determined by the motivational relevance of a feedback (Wang et al., 2013; Wu & Zhou, 2009), as well as by the degree to which the feedback is unexpected (Hajcak, Moser, Holroyd, & Simons, 2007). If a feedback violates expectations, additional resources have to be recruited to update the mental models that help to respond appropriately to the social and physical environment (Polich, 2007).

Wang and colleagues (2013) used psychophysiological measures to examine cognitive processing in the chicken game, which is used as a game-theoretical model of social competition. In this two-person social dilemma game, one receives the highest outcome

when choosing to aggress while the other player chooses the cooperative option. Mutual cooperation only leads to moderate gains. Choosing to cooperate while the other player aggresses leads to a moderate loss. When both players choose to aggress, both have the worst outcome. This game represents a social dilemma because there is an incentive to aggress when the other player cooperates (unilateral aggression leads to the highest outcome), but both players benefit more from mutual cooperation than from mutual aggression. The P300 was larger when participants chose to aggress, and when the other partner cooperated, suggesting that the monetary gains associated with an opponent's cooperation elicited more attention. To examine the feedback negativity at medial anterior electrodes, difference waves were created by subtracting the trials in which partners cooperated from trials in which partners aggressed. Results showed that these difference waves had a higher amplitude when participants chose to cooperate, rather than aggress, possibly reflecting the participant's disappointment when the opponent did not reciprocate cooperation. Wang et al. concluded from their findings that the P300 and the feedback negativity do not only reflect the valence or magnitude of outcomes, but also fairness considerations (e.g., the social pain elicited by unanticipated aggression).

In the present study, we extend the literature in two ways. First, we examine feedback processing in a sequential prisoner's dilemma game, which is a standard paradigm for studying reciprocal cooperation (Axelrod & Hamilton, 1981; Trivers, 1971). In this game, the highest payoff is obtained when a player cheats against a cooperative partner. Mutual cooperation leads to moderate gains. When both players choose to defect, both have a moderate loss. The worst outcome results when choosing to cooperate while the partner chooses to cheat. The game represents a social dilemma because there is an incentive to cheat (unilateral cheating leads to the highest monetary outcome), but mutual cooperation leads to a better outcome than mutual defection. When the game is played only once, the only rational strategy is to defect. Higher levels of cooperation can be sustained in repeated games that allow for retaliation against uncooperative players. In the last round, there is often a drop in the cooperation rate because there is no opportunity to retaliate cheating, which has been called the endgame effect (Gächter & Thöni, 2005). Since we were most interested in the participants' responses to the other player's cheating or cooperative decision, we used a sequential prisoner's dilemma game (Clark & Sefton, 2001) in which the stimulus person chose to cooperate or to cheat after the participant had invested into the game. We were most interested in comparing the ERP responses to a partner's cheating with the ERP responses to a partner's cooperative choice. We planned to analyze the response to the feedback about the partner's decision in Round 1, in which participants had no knowledge about the partner's previous behavior and, therefore, no specific expectations about the partner's future behavior. These results were contrasted with those of Round 4, in which participants had built up strong expectations about their partner's behavior based on their experiences in the previous rounds. Given that some of the partners changed their strategies in Round 4, it was possible to examine how the ERP responses to cheating and cooperation would be affected by expectancy violations. According to the negativity hypothesis, information about another person's cheating should receive more attention, and should elicit an enhanced P300 response in comparison to information about another person's cooperation. Alternatively, according to the diagnosticity hypothesis, participants should show no consistent negativity bias, but should instead attend both to cheating and to cooperation,

whichever reveals more information about another person's character. If this hypothesis is correct, then we should find an enhanced P300 in response to cheating and cooperation in comparison to an irrelevant feedback, and an enhanced P300 in response to unexpected information in Round 4 of the game. Based on the view that the feedback negativity codes outcome negativity, it seems possible to postulate that the feedback negativity should be more pronounced in response to the feedback about another person's cheating in comparison to the feedback about another person's cooperation in the sequential prisoner's dilemma game. However, based on the assumption that feedback negativity reflects the salience of prediction errors rather than negativity per se, it seems possible to postulate that the unexpected behavior in the last round of the game (when the expectations about the other person's behavior are strongest) should result in a larger feedback negativity amplitude in comparison to the expected behaviors.

Considering the important role of memory in reciprocal cooperation (Dugatkin, 2002; Milinski & Wedekind, 1998), we also assessed participants' memory for the social interaction partners in a surprise memory test. We were most interested in source memory, that is, memory for the association between the persons' faces and their behavior in the social dilemma game. As discussed above, early behavioral studies have obtained evidence for a memory advantage for faces of cheaters (Buchner et al., 2009; Mealey et al., 1996), but more recent studies suggest that both cheaters and cooperators are preferentially remembered in comparison to control persons (Bell, Buchner, & Musch, 2010). In the study of Bell et al., participants played a sequential prisoner's dilemma game with partners whose faces were shown. Some of the partners cooperated (invested as much money as the participant did) and some of the partners cheated (invested less). In a control condition, participants were told that their investments were not registered because the Internet connection broke off. While it seems conceivable that an Internet disconnection may have been initially experienced as frustrating, it is not diagnostic about a partner's character, and thus less valuable to remember. In line with this interpretation, this control condition was poorly remembered in comparison to cooperation and cheating in the previous study, but cooperators and cheaters were equally well remembered. We expected to replicate these findings.

Furthermore, we postulated that the retrieval of another person's cheating or cooperation in a social dilemma game should be associated with similar ERP responses as the retrieval of other social context information (MacKenzie & Donaldson, 2007, 2009; Paller, Gonsalves, Grabowecky, Bozic, & Yamada, 2000; Paller et al., 2003). When face-specific brain regions are involved, the ERP correlates of context retrieval are somewhat different from those observed in studies using verbal stimulus materials (MacKenzie & Donaldson, 2009; Yovel & Paller, 2004). Paller et al.'s (2000) participants saw faces with or without rich social context information (biographical information such as "I'm Jimmy and I was a roadie for the Grateful Dead"). At test, participants had to discriminate between faces shown with or without context information. While the successful recognition of previously presented faces was associated with a posterior positivity, the retrieval of person-specific context information was associated with a more positive potential at anterior electrodes between 400 and 600 ms after face onset (see also Paller et al., 2003). Since then, other studies have confirmed that this anterior component is associated with the retrieval of contextual information such as a person's name in response to face stimuli (e.g., MacKenzie & Donaldson, 2007). Enhanced retrieval of another person's cheating or cooperation should thus be

reflected in a more pronounced anterior positivity in response to cheater faces and cooperator faces in comparison to control faces. Again, the negativity hypothesis would predict a more pronounced effect for cheater faces in comparison to cooperator faces while the diagnosticity hypothesis predicts enhanced retrieval of socially relevant contexts (cheating and cooperative behavior) in comparison to the control condition.

Method

Participants

Participants were 36 right-handed students at Heinrich Heine University (20 of whom were female, M age = 23, SD = 3.3) who participated for monetary compensation. They had normal vision or wore glasses. None of the participants had a self-reported neurological or psychiatric illness. Participants were instructed to attend for testing well rested, and to refrain from consuming alcohol or drugs the evening before participation.

Stimuli

We sampled 150 female and 150 male faces of young white adults from the FACES database (Ebner, Riediger, & Lindenberger, 2010), the NimStim face stimulus set (Tottenham et al., 2009), the Karolinska Directed Emotional Faces set (Goeleven, De Raedt, Leyman, & Verschuere, 2008; Lundqvist, Flykt, & Öhman, 1998), the Center of Vital Longevity database (Minear & Park, 2004), and the Radboud Faces database (Langner et al., 2010). The color pictures (250×375 pixels) showed the faces in frontal view with a neutral expression on a uniform black background. For each participant, the faces were randomly assigned to conditions with the restriction that facial gender was counterbalanced between conditions.

Procedure

Participants signed an informed consent before extensive instructions for the social dilemma game were presented both visually and auditorily. The game was similar to that used in our previous studies on memory for reputations (Bell, Buchner, & Musch, 2010; Giang, Bell, & Buchner, 2012), but adjusted to allow for the investigation of ERP effects. An example for a single trial is shown in Figure 1. First, participants were informed about their account balance. Then, a fixation cross was presented, after which the face of the interaction partner was shown for 1,000 ms. Participants were prompted to invest a certain amount of money into a joint business venture with the interaction partner shown on the screen. Participants were always required to invest some amount of money, but they could choose to invest 15 or 30 cent by pressing the upper or lower button on a response box, using the index fingers of their right or left hand. Hand position was counterbalanced among participants. The participant's investment was shown in yellow for 1 s. The participant was instructed to wait a randomly determined interval varying from 700 to 1,000 ms in 50-ms steps. During that time, three red dots (.) were presented beneath the picture. Then, the partner's choice was presented for 1 s. If the partner cooperated, the word "Cooperation" appeared in red beneath the partner's face. In the instructions, participants were informed that cooperative partners reciprocated, and invested exactly the same amount of money as the participant. If the partner cheated, the word "Cheating" appeared in red beneath the partner's face. Participants

Figure 1. Schematic illustration of a single trial of the cooperation game. The picture was taken from the Center of Vital Longevity database (Minear & Park, 2004).

knew that cheaters invested nothing (0 cents). The investments were summed up, and a bonus was added, which was always one third of the investments. The total sum was split up between both partners, and both received half of the total sum, regardless of what they had invested. Thus, an interaction with a cooperator resulted in a gain, and an interaction with a cheater resulted in a loss of the same magnitude. The gain or loss was presented for 1.5 s in yellow beneath the partner's face. Then, the updated account balance was shown for 2 s before the face of the next partner was presented. There was also a control condition, in which the word "Break-Off" appeared in red beneath the partner's face. The participant knew that this meant that it was not possible to establish a stable Internet connection to this partner, as a result of which their own investment was not correctly registered by the computer system, and the transaction was aborted. In consequence, the transaction resulted in neither a loss nor a gain for the participant and his or her partner.

There were three rounds of the game before the surprise memory test. In each round, each participant saw the same 50 cooperator faces, 50 cheater faces, and 50 control faces in a different, randomly determined order. The behavior of the partners was always consistent with their behavior in the first round of the game. Then, a standard source memory test followed in which each participant saw the 50 cooperator faces, 50 cheater faces, and 50 control faces intermixed with 150 new faces. The order of the faces was randomly determined. Each trial started with the presentation of a white fixation cross, which was presented for an interval varying randomly from 700 to 1,000 ms in 50-ms steps. Next, a face was presented for 1 s, after which the participant was prompted to respond with "old" or "new" by pressing the right or left button of the response box. If a face was classified as old, the participant was prompted to make a source judgment by indicating whether the face was previously associated with cooperation, cheating, or break-off by clicking on one of three buttons of the response box. The response-button mapping varied randomly between participants.

Then, the participant was informed that the last round of the game was about to follow. In this round, each participant encountered only the 50 cooperator faces and the 50 cheater faces (known from the previous rounds). Some of the partners unexpectedly changed strategies. Of the 50 cooperators, 30 cooperated as expected, but 20 unexpectedly cheated. Of the 50 cheaters, 30 cheated as expected, but 20 unexpectedly cooperated. The order of the faces was randomly determined. Participants knew that they played for real money, and that they would be paid out at the end of the experiment.

EEG Recordings and Data Analysis

Brain activity was recorded with ActiCap active Ag-AgCl electrodes (Brain Products, Gilching, Germany) at 30 scalp sites using the AFz as ground electrode. The horizontal electrooculogram (EOG) was recorded from electrodes placed at the outer canthus of each eye. The vertical EOG was recorded from electrodes placed at FP1 and under the left eye. The activity of all scalp electrodes was initially referenced to FCz and rereferenced offline to averaged mastoids. Biosignals were recorded continuously with a BrainAmp DC amplifier with a sampling rate of 500 Hz. EEG data were further analyzed using VisionAnalyzer 2 software (Brain Products). A 0.1 Hz–30 kHz band-pass filter was applied offline. Trials containing voltage steps $> 300 \mu\text{V}$ were removed prior to the removal of eye movement artifacts. Eye movements were corrected via independent component analysis (ICA), using the ocular correction ICA tool in Vision Analyzer 2. EEG epochs extended from 100 ms prior to stimulus onset until 1,000 ms after stimulus onset. Responses to two types of stimuli were analyzed. First, we analyzed the response to the feedback about a partner's behavior (cheating, cooperation, or break-off) in Round 1 and Round 4 of the sequential prisoner's dilemma game. Second, we examined the ERP responses to the onset of the (cheater, cooperator, control, or new) faces in the source memory test. Trials with remaining artifacts (voltage

steps $> 20 \mu\text{V/ms}$, voltage differences $> 200 \mu\text{V}$ within a 2,600-ms time window) were excluded based on a second, semiautomatic artifact removal. Mean amplitudes were computed for each participant. The electrodes were clustered according to their caudality (anterior, central, posterior) and laterality (left, medial, right) into nine clusters of electrodes (left anterior [LA; F7, F3, FC5], medial anterior [MA; Fz, FC1, FCZ, FC2], right anterior [RA; F4, F8, FC6], left central [LC; T7, C3, CP5], medial central [MC; Cz, CP1, CP2], right central [RC; C4, T8, CP6], left posterior [LP; P7, O1], medial posterior [MP; P3, Pz, P4, Oz], and right posterior [RP; P8, O2]), which has proven to be a good compromise between spatial resolution and reliability of the measures (Bell, Dentale, Buchner, & Mayr, 2010; Mayr, Niedeggen, Buchner, & Pietrowsky, 2003). The data of one participant were excluded from all ERP analyses because of artifacts in almost every trial.

Design

In summary, participants first played three rounds of a sequential prisoner's dilemma game with 150 partners. Partner behavior (cooperation, cheating, control) was the repeated measures independent variable. Next, a source memory test was required in which all (cooperator, cheater, control) faces were presented, intermixed with 150 new faces. Then, the last round of the game followed with an underlying 2×2 design with partner behavior (cooperation, cheating) and expectation violation (expected, unexpected) as independent variables. A multivariate approach was used for all general linear model within-subject comparisons. In the present application, all multivariate test criteria correspond to the same (exact) F statistic, which is reported. The level of α was set to .05 for all analyses. Given a sample size of $N = 35$ in the ERP analyses and the requirement that $\alpha = \beta = .05$, and assuming a correlation among the levels of the independent variable of $p = .80$, it was possible to detect effects of size $f = 0.18$ for the three-level partner behavior variable in Round 1 and of size $f = 0.2$ for the two-level partner behavior and expectation violation variables in Round 4 [i.e., between small ($f = 0.1$) and medium ($f = .25$) size effects in terms of Cohen's (1988) conventions] (Faul, Erdfelder, Lang, & Buchner, 2007).

Results and Interpretation

Behavioral Data

Investments in the social dilemma game. A 3 (Partner Behavior: cheating, control, cooperation) $\times$ 3 (Round: Rounds 1–3) multivariate analysis of variance (MANOVA) on the mean investments in the social dilemma game revealed a main effect of partner behavior, $F(2,34) = 39.68, p < .01, \eta_p^2 = .70$, that was qualified by a significant interaction with round of the game, $F(4,32) = 23.64, p < .01, \eta_p^2 = .75$. As expected, the cooperator condition did not differ from the control condition in the first round of the game (in which participants had no prior experience with the interaction partners, and no knowledge about their character), $t(35) = -0.26, p = .80$, but investments were higher in the cooperator condition than in the control condition in Round 2, $t(35) = 3.55, p < .01$, and Round 3, $t(35) = 8.25, p < .01$. Likewise, the cheater condition did not differ from the control condition in the first round of the game, $t(35) = -0.68, p = .50$, but investments were lower in the cheater condition than in the control condition in Round 2, $t(35) = 4.01, p < .01$, and Round 3, $t(35) = 4.40, p < .01$. These findings show

Table 1. Mean Investments (in Euro Cents) in the Social Dilemma Game as a Function of Partner Behavior (Cooperation, Control, Cheating) and Round (1–4)

	Cooperation		Control		Cheating	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Round 1	20.63	(2.77)	20.68	(2.79)	20.85	(2.32)
Round 2	20.43	(2.74)	19.53	(2.75)	18.84	(2.45)
Round 3	22.18	(2.37)	19.27	(2.65)	18.12	(1.91)
Round 4	22.06	(2.44)	–	–	18.16	(2.32)

Note. In Round 4, only the faces of the cooperators and the cheaters were presented. Two thirds of the partners switched strategies (i.e., partners with a history of cooperation cheated, and partners with a history of cheating cooperated). Here, cooperation and cheating refer to the previous history of cheating or cooperation, not to the behavior in Round 4 of the game.

that participants remembered their partner's behaviors, and adjusted their investments accordingly.

In Round 4, only the faces of the cooperators and the cheaters were presented. Some of the partners switched strategies. In the following analysis, the partner behavior refers to the previous history of cheating and cooperation, not to the behavior in Round 4 of the game. A 2 (Partner Behavior: cheating, cooperation) $\times$ 2 (Round: Rounds 3, 4) MANOVA revealed that the investments did not differ between Rounds 3 and 4 of the game in response to cooperator and cheater faces, $F(1,35) = 0.02, p = .89, \eta_p^2 < .01$. There was also no interaction between round and partner behavior, $F(1,35) = 0.11, p = .75, \eta_p^2 < .01$, but there was a significant effect of partner behavior, $F(1,35) = 135.10, p < .01, \eta_p^2 = .79$. Investments were lower when the partners were associated with a history of cheating than when they were associated with a history of cooperation. These findings suggest that the learned reputations continued to inform the investment decisions in Round 4 of the game (Table 1).

Old-new recognition and source memory. In the memory test, old-new face recognition (Table 2) differed significantly as a function of partner behavior, $F(2,34) = 7.51, p < .01, \eta_p^2 = .31$. Old-new face recognition did not differ between cooperator faces and cheater faces, $t(35) = 1.42, p = .16$, but cheater faces and cooperator faces were better recognized than control faces, $t(35) = 2.84, p < .01$, and $t(35) = 3.85, p < .01$.

Next, we wanted to know how well participants remembered the association between the face and the game outcome (cheating, cooperation, and control). When examining source memory, it is important to use a measure that does not confound source memory with old-new recognition and guessing biases (Bröder & Meiser, 2007). Therefore, we analyzed the data using the multinomial source monitoring model of Bayen, Murnane, and Erdfelder (1996) to decompose the classification of the faces presented in the source memory test into the cognitive processes involved in this

Table 2. Old-New Recognition in Terms of P_r (Hit Rates – False Alarm Rates) as a Function of Partner Behavior (Cooperation, Control, Cheating)

	Cooperation		Control		Cheating	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
	0.76	(0.18)	0.72	(0.17)	0.75	(0.17)

Figure 2. Estimates of the source memory parameter d , representing the conditional probability of remembering the partner's behavior given that the partner's face was correctly recognized, as a function of the partner's behavior. The error bars represent the 95% confidence intervals.

classification, and to estimate independent parameters of item memory, guessing, and source memory from the observed response frequencies in the source memory test (Moshagen, 2010). The source monitoring model is an extension of the two-high threshold model of old-new recognition (Snodgrass & Corwin, 1988), and has the advantage that it provides a measure of source memory that is not affected by old-new recognition and guessing, as confirmed by validation studies (Bayen & Murnane, 1996). Multinomial models are widely used in research on source monitoring (Batchelder & Riefer, 1999; Erdfelder et al., 2009), and have been applied in most previous studies on reputational memory (Bell, Buchner, & Musch, 2010; Buchner et al., 2009; Giang et al., 2012; Kroneisen et al., 2015). We used the same model as in Bell et al.'s (2010) Experiment 1, which fit the data well, $G^2(1) = 0.25$, $p = .62$. Source memory, as reflected in source memory parameter d (which represents the conditional probability of remembering the negative, neutral, or positive reputation associated with a face given that the face was correctly recognized), did not differ between cheater faces and cooperator faces, $\Delta G^2(1) = 1.31$, $p = .25$, but cheaters and cooperators were better remembered than control faces, $\Delta G^2(1) = 21.97$, $p < .01$, and $\Delta G^2(1) = 56.62$, $p < .01$, respectively (Figure 2). These analyses were done using multiTree (Moshagen, 2010).

ERP Data

Feedback processing. In previous studies, feedback indicating monetary reward elicited an increased P300 amplitude (Wu & Zhou, 2009; Yeung & Sanfey, 2004). Here, we analyzed the ERP response to the feedback about the partner's behavior in the game. Consistent with previous studies, the P300 component was most pronounced at medial central and medial posterior electrodes between 300–600 ms after feedback onset (Figure 3). At medial central electrodes, there was a main effect of partner behavior on mean amplitude, $F(2,33) = 14.72$, $p < .01$, $\eta_p^2 = .47$. Feedback about cooperation and cheating elicited an enhanced P300 in comparison to reputation-irrelevant control feedback, $F(1,34) = 30.34$, $p < .01$, $\eta_p^2 = .47$, but the P300 amplitude did not differ between cheating and cooperation, $F(1,34) = 0.75$, $p = .39$, $\eta_p^2 = .02$. The

same effect was obtained at medial posterior electrodes, $F(2,33) = 14.68$, $p < .01$, $\eta_p^2 = .47$. Cheating and cooperation were associated with a more pronounced P300 in comparison to the control condition, $F(1,34) = 30.10$, $p < .01$, $\eta_p^2 = .47$, but cooperation and cheating elicited similar responses, $F(1,34) = 0.30$, $p = .59$, $\eta_p^2 < .01$. Thus, feedback that was diagnostic about a partner's character elicited an enhanced P300, but the P300 amplitude did not differ between cheating and cooperation, consistent with previous findings showing that the P300 is insensitive to feedback valence (Hajcak, Holroyd, Moser, & Simons, 2005; Yeung & Sanfey, 2004).

In line with previous studies (Wu & Zhou, 2009), we examined mean amplitudes at medial anterior electrodes 250–300 ms after feedback onset to examine the feedback negativity. The feedback negativity was not well defined in Round 1 of the game. Descriptively, mean amplitude was more negative for feedback about a partner's cheating than for the feedback about a partner's cooperation, but this difference was not statistically significant, $F(1,34) = 1.29$, $p = .26$, $\eta_p^2 = .04$. Amplitudes were more negative in the control condition in comparison to the other two conditions, $F(1,34) = 11.85$, $p < .01$, $\eta_p^2 = .26$. Speculatively, these findings could be interpreted as indicating that the feedback about the technical failure of the Internet connection may have initially caused a feeling of surprise or irritation because it was unexpected, but received less further processing due to its low social informational value, as suggested by the P300 results.

Expectancy violation effects. In Round 4 of the game (which followed the source-memory test), only faces of cheaters and cooperators were presented. Some partners switched strategies and unexpectedly cheated when they had previously cooperated, and vice versa. We were most interested in examining the effects of expectancy violations on the P300. The data of four participants had to be excluded because of too few (<15) artifact-free trials. As shown in Figure 4, unexpected cheating was associated with a more pronounced P300 than expected cheating. The topography of this effect was similar to the P300 effect reported above. At medial central electrodes, there was no main effect of behavior, $F(1,30) < 0.01$, $p = .96$, $\eta_p^2 < .01$, but there was an interaction between behavior and expectation, $F(1,30) = 18.02$, $p < .01$, $\eta_p^2 = .38$, which was due to the fact that there was a more pronounced difference between expected and unexpected cheating, $F(1,30) = 18.64$, $p < .01$, $\eta_p^2 = .38$, but no difference between expected and unexpected cooperation, $F(1,30) = 1.04$, $p = .32$, $\eta_p^2 = .03$. The same interaction was obtained at medial posterior electrodes, $F(1,30) = 12.13$, $p < .01$, $\eta_p^2 = .29$, again reflecting a pronounced difference between expected and unexpected cheating, $F(1,30) = 17.83$, $p < .01$, $\eta_p^2 = .37$, but no difference between expected and unexpected cooperation, $F(1,30) = 0.27$, $p = .61$, $\eta_p^2 = .01$.

Given that the feedback negativity is known to reflect both the valence of a feedback (Hajcak et al., 2006) and prediction errors (Bellebaum & Daum, 2008; Hajcak et al., 2007), it is interesting to examine the feedback negativity to expected and unexpected cooperation and cheating in the last round of the game. A 2×2 MANOVA with behavior (cheating vs. cooperation) and expectation (expected vs. unexpected) as independent variables was performed. There were no effects of behavior, $F(1,30) = 1.46$, $p = .24$, $\eta_p^2 = .05$, and expectation, $F(1,30) = 0.03$, $p = .86$, $\eta_p^2 < .01$, but there was an interaction between behavior and expectation, $F(1,30) = 4.57$, $p = .04$, $\eta_p^2 = .13$. There was a difference between cooperation and cheating for expected behaviors, $t(30) = 2.99$,

Figure 3. Grand-averaged ERPs in response to the feedback about the partner's behavior in Round 1 of the social dilemma game at left anterior (LA), medial anterior (MA), right anterior (RA), left central (LC), medial central (MC), right central (RC), left posterior (LP), medial posterior (MP), and right posterior (RP) sites.

$p < .01$, but not for unexpected behaviors, $t(30) = 0.28$, $p = .78$. As Figure 5 shows, the feedback negativity was less pronounced for the expected cooperation condition in comparison to all other conditions. This finding can be interpreted to support the idea that the feedback negativity is a correlate of the undesirability of a feedback (Hajcak et al., 2006). Being cheated on is an undesirable outcome regardless of whether it is expected or unexpected because it is associated with a negative payoff. Although unexpected cooperation could be considered a pleasant surprise at first sight, in the present context it may elicit a feedback negativity because people often chose not to trust persons associated with a history of cheating (Table 1), which can be considered a prediction error. Not trusting (or being uncooperative towards) a helpful person may elicit negative self-related feelings such as regret and guilt (Rilling et al., 2002). Unexpected cooperative or altruistic acts are often perceived as negative to the extent that disproportionate cooperation is even punished in many societies (Henrich et al., 2006). Thus, the only condition that is unambiguously positive is the expected cooperation condition.

Retrieval of negative and positive reputations. Next, we analyzed the ERPs to the faces presented in the memory test. Given that cheaters represent an evolutionary relevant threat, it seemed possible to postulate that an early modulation of face perception by negative reputations may result in a quick detection of cheater faces (Galli, Feurra, & Viggiano, 2006). The right-lateralized occipitotemporal N170 is seen as a correlate of the perceptual encoding of a face (Bentin, Allison, Puce, Perez, & McCarthy, 1996), which has often been characterized as being insensitive to attentional modulations (Cauquil, Edmonds, & Taylor, 2000), but there are some reports showing a modulation of early configural face processing by emotional context information (Galli et al., 2006; Montalan et al., 2008). For instance, Galli et al. (2006) examined the ERP in response to faces embedded in emotionally valent contexts (newspaper articles with headlines about moral or immoral actions committed by the person), and found a decreased N170 in response to faces previously associated with negative in comparison to positive contexts, which was interpreted as evidence for an immediate perception advantage of threatening faces that could be beneficial

Figure 4. Grand-averaged ERPs at medial central (MC) electrodes in response to the feedback about the partner's behavior (cheating or cooperation) in the last round of the social dilemma game as a function of whether the behavior was expected (consistent with the partner's behavior in previous rounds) or unexpected (inconsistent with the partner's previous behavior).

from an evolutionary perspective. However, these findings stand in contrast to the prevalent view that the N170 reflects structural face processing, and is not associated with person recognition or context retrieval (Schweinberger, 2011). Therefore, it seemed interesting to test whether the N170 differed as a function of partner behavior at right posterior electrodes (Figure 6). Our results showed that mean amplitude between 150–200 ms after face onset did not differ as a function of partner behavior, $F(2,33) = 0.56$, $p = .58$, $\eta_p^2 = .03$, providing evidence against a cheater detection module operating at early stages of perceptual processing.

In addition, we looked at the retrieval of the social context information (i.e., at the processes governing retrieval of a partner's cheating or cooperative behavior). Previous studies (e.g., Paller et al., 2000) suggested that memory for the association between a face and episodic context information is reflected in an anterior positivity 400–600 ms after face onset. Consistent with these studies, there was a main effect of partner behavior on mean amplitudes at medial anterior electrodes in this time window, $F(2,33) = 4.28$, $p = .02$, $\eta_p^2 = .21$ (Figure 6). Post hoc contrast showed that responses to cheater and cooperator faces differed from those to control faces, $F(1,34) = 7.36$, $p = .01$, $\eta_p^2 = .18$, but responses to

Figure 5. Grand-averaged ERPs at medial anterior (MA) electrodes in response to the feedback about the partner's behavior (cheating or cooperation) in the last round of the social dilemma game as a function of whether the behavior was expected (consistent with the partner's behavior in previous rounds) or unexpected (inconsistent with the partner's previous behavior).

cheater and cooperator faces did not differ from each other, $F(1,34) = 0.86$, $p = .36$, $\eta_p^2 = .02$. This analysis shows that the difference between faces presented with and without reputation-relevant information is reflected in an anterior positivity, but there is no difference between cheater or cooperator faces.

However, it remains to be shown whether this component is indeed a correlate of the retrieval of reputational context information (i.e., retrieval of the association between the face and cheating or cooperation). To examine this, we compared the correctly recognized faces associated with source hits (i.e., faces for which the positive, neutral, or negative reputation could be correctly identified) with the correctly recognized faces associated with source errors (i.e., faces correctly identified as old for which the previous history of interactions was not correctly identified) to obtain a correlate of source memory (Figure 7). In line with previous studies examining source memory for faces (MacKenzie & Donaldson, 2007, 2009; Paller et al., 2000, 2003), we included a contrast between hits (correctly identified old faces) and correct rejections (correctly identified new faces) to examine the old-new effect for the facial stimuli used in the present paradigm. Four participants had to be excluded because of too few data points (<15) in one of the conditions. The difference between correctly recognized faces associated with source hits and those associated with source errors was most pronounced at medial anterior electrodes 400–600 ms after face onset. The effect of face type (source hit, source error, correct rejection) at medial anterior electrodes was significant, $F(2,29) = 11.11$, $p < .01$, $\eta_p^2 = .43$. Mean amplitude was more positive for correctly recognized faces associated with source hits than for correctly recognized faces associated with source errors, $F(1,30) = 6.04$, $p = .02$, $\eta_p^2 = .17$, but there was no difference between correctly recognized faces associated with source errors in comparison to correctly rejected new faces, $F(1,30) = 1.34$, $p = .26$, $\eta_p^2 = .04$. At medial posterior electrodes, the opposite pattern emerged. There was no difference between correctly

Figure 6. Grand-averaged ERPs in response to the faces in the memory test as a function of face type (new, control, cooperation, cheating) at left anterior (LA), medial anterior (MA), right anterior (RA), left central (LC), medial central (MC), right central (RC), left posterior (LP), medial posterior (MP), and right posterior (RP) sites.

recognized faces associated with source hits and correctly recognized faces associated with source errors, $F(1,30) = 2.26$, $p = .14$, $\eta_p^2 = .07$, but the old-new effect reflected in the difference between correctly recognized faces associated with source errors and the correctly rejected new faces was significant, $F(1,30) = 9.39$, $p < .01$, $\eta_p^2 = .24$. This replicates previous studies using face stimuli (MacKenzie & Donaldson, 2007, 2009; Paller et al., 2000, 2003) that have identified a posterior old-new effect as a correlate for old-new face recognition, and an anterior correlate of context memory for faces.

Discussion

Coupling the sequential prisoner's dilemma game with psychophysiological measures allowed us to examine the motivational and cognitive reactions underlying reciprocal social exchange. Specifically, the P300 is often considered as reflecting the motivational salience of information in social dilemma games (Wang et al.,

2013). Several theories of reciprocal cooperation imply that information about another person's cheating should be prioritized over information about another person's cooperation. Specifically, social contract theory (Cosmides & Tooby, 1989, 1992) postulates that the human brain includes neurocognitive mechanisms that are specialized in the detection of cheaters. Inconsistent with the negativity hypothesis, we found that the ERP correlates of the outcome evaluation were very similar for the cheater and the cooperator condition. In particular, the P300 had a similar shape and amplitude in the cheater condition and in the cooperator condition in Round 1 of the sequential prisoner's dilemma game. Nevertheless, participants' P300 was greater when the opponent cooperated or cheated than when his or her Internet connection broke off. The P300 findings suggest that people devote increased attention to information that could be helpful in social exchange, but do not generally prioritize cheating over cooperation.

The results obtained in Round 4 of the sequential prisoner's dilemma game provide insight into ways in which the processing

Figure 7. Grand-averaged ERPs at medial electrodes in response to correctly rejected new faces, correctly recognized faces associated with source errors, and correctly recognized faces associated with source hits at medial anterior (MA), medial central (MC), and medial posterior (MP) sites.

of cooperation and cheating is affected by expectancy violations. The P300 was larger in response to unexpected cheating in comparison to expected cheating. Given that the P300 is known to be enhanced for less probable events, this finding could be attributed to an orienting reaction elicited by the violation of an expectation (Polich, 2007). However, no corresponding effect was found for unexpected cooperation. This pattern of results suggests that in the present context the P300 does not only reflect the low probability of the eliciting event, but may instead be reflective of the informational value of the social behavior. Honest individuals will always

be fair and cooperative, regardless of the circumstances. Dishonest persons, in contrast, will not always show antisocial behavior. Instead, they tend to show whatever behavior helps them to maximize their own payoff. Therefore, even dishonest persons will show cooperative behavior if cooperation yields a better payoff because cheating is likely to be punished (Hilbig, Zettler, & Heydasch, 2012). Specifically, in a repeated social dilemma game, it may pay off to make cooperative decisions in the first rounds of the game because it is likely that partners retaliate against cheating by reducing their investments (which reduces the payoff in later rounds and, thereby, the incentive to cheat). However, in the last round of the game, there is no risk of retaliation. Therefore, selfish individuals who cooperated merely out of fear of retaliation will switch to an uncooperative strategy (e.g., Keser & van Winden, 2000). Thus, cheating in the last round of a repeated social dilemma game may be particularly diagnostic of a partner's character, which may explain why participants paid special attention to this information. This finding corroborates previous conclusions (Wang et al., 2013) that the feedback P300 is not only determined by expectancy violation, but is also sensitive to the affective and motivational implications of social dilemma situations. Considering that the P300 did not differ between feedback about cooperation and feedback about cheating in Round 1, the general pattern of results suggests that people do not always prioritize negative over positive social information, but instead flexibly allocate their attention to whatever information seems more relevant to evaluate another person (i.e., is more diagnostic about a partner's character) in a specific context.

The feedback negativity is often elicited by feedback indicating either monetary losses or monetary changes. Based on these and similar results, it has been suggested that the feedback negativity may reflect either a feedback's valence or the degree to which the feedback is salient because it violates expectations. In Round 1 of the sequential prisoner's dilemma game, the participant arguably had no specific expectations about the partner's future behavior, and the feedback negativity was poorly developed. The feedback about the failure of the Internet connection seems to have elicited a feedback negativity, which could indicate that participants were surprised or irritated about the initially unexpected technical failure. Note that, despite this initial irritation, the control condition had less social informational value than the other two conditions, which may explain why this feedback received less higher-order attentional processing as reflected in the P300, and was less well remembered than cooperation and cheating. In the fourth round of the game, cheating was associated with a more pronounced feedback negativity than expected cooperation. Interestingly, a partner's unexpected cooperation elicited a similar feedback negativity as cheating, despite the fact that it was associated with a monetary reward. This is consistent with recent studies suggesting that, in social situations, the feedback negativity is more determined by the social implications of an outcome than by its monetary consequences (Boksem & De Cremer, 2010; Polezzi et al., 2008; Wang et al., 2013). In particular, it has been argued that defecting against a cooperative partner is perceived as aversive (Rilling et al., 2002). In the present study, investing only a little money into a joint business with a partner who later chooses to cooperate may have been perceived as a prediction error, and may have therefore been associated with a feedback negativity. This finding corroborates the general impression that participants responded sensitively to their partner's cooperation.

A necessary requirement for reciprocal cooperation is that cooperative individuals learn to discriminate between cheaters and

cooperators. Thus, reciprocal cooperation requires good memory for social interaction partners. In particular, cooperative individuals have to have good source memory for the association between a face and the behavior in the prisoner's dilemma game (Buchner et al., 2009). Therefore, we were not only interested in the outcome evaluation stage, but we were also interested in how well these previous episodes could be retrieved in a memory test. We expected that the retrieval of social context information (a person's behavior in the social dilemma game) would be associated with the same ERPs as the retrieval of nonemotional context information. Previous studies have identified a posterior correlate of face recognition, and an anterior correlate of context memory for faces (MacKenzie & Donaldson, 2007, 2009; Paller et al., 2000, 2003). In line with these previous findings, we observed a posterior effect that differentiated between correctly recognized old faces and new faces, and an anterior effect that differentiated between faces whose contexts were correctly remembered and faces whose contexts were not correctly remembered. This anterior effect between 400 and 600 ms after face onset was more pronounced for cheater and cooperator faces in comparison to control faces, which suggests that retrieval was enhanced for socially relevant information that helped to discriminate between cheaters and cooperators. In contrast to earlier theorizing (Buchner et al., 2009; Mealey et al., 1996), there was no indication that memory for cheating was enhanced over memory for cooperation. In principle, both memory for cheating and memory for cooperation are useful for deciding whether to cooperate with another person or not. From a functional point of view, it should not matter if people attend to cheating or cooperation, as long as they are able to discriminate between cheaters and cooperators (Barclay, 2008). Previous studies have shown that people do not generally have enhanced memory for cheaters but instead focus on both cheating and cooperation, whichever information is more relevant in a given situation (Barclay, 2008; Bell, Buchner, & Musch, 2010; Kroneisen et al., 2015). The present study extends these findings by showing similar ERP correlates for the retrieval of negative and positive reputational information in a source mem-

ory test. The symmetric nature of the retrieval effect is in line with other recent neuropsychological research showing that the processing of a stimulus is determined by its motivational and behavioral relevance rather than by its negativity (e.g., Brosch, Sander, Pourtois, & Scherer, 2008).

In summary, we examined the feedback negativity and the P300 in response to feedback about another person's cheating or cooperation. Replicating results that have been obtained with other experimental dilemma games (Wang et al., 2013), we found that both components were not only determined by outcome valence and magnitude (i.e., monetary loss or gain), but also by their social implications. There was no consistent evidence for a negativity bias, that is, enhanced allocation of attention to feedback about another person's cheating in comparison to feedback about another person's cooperation. Instead, participants seemed to focus on both positive and negative information, but flexibly adjusted their processing biases to the diagnosticity of the information. This general conclusion was corroborated by the ERP correlates of memory retrieval. Both information about a partner's cooperation as well as information about a partner's cheating were preferentially retrieved. These results are inconsistent with models of social exchange postulating that people should preferentially attend to, and remember, cheaters (Cosmides & Tooby, 1989, 1992), and are more consistent with alternative models postulating that people preferentially attend to, and remember, both cheating and cooperation, whichever information is more relevant in a given context. Indeed, it can be questioned whether attending selectively to the negative aspects of social exchange (cheating) would represent an adaptive facet of human cognition. Finding reliable cooperation partners is an important task, just as important as detecting cheaters. Presumably, a system that is hardwired to only attend to, and remember, cheaters would be ill-equipped to deal with the complexities of social exchange. A system that evaluates the relevance of social information, and adapts to the diagnosticity of this information in a given situation, is probably much better suited to manage the intricate demands of human cooperation.

References

- Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, 211, 1390–1396. doi: 10.1126/science.7466396
- Barclay, P. (2008). Enhanced recognition of defectors depends on their rarity. *Cognition*, 107, 817–828. doi: 10.1016/j.cognition.2007.11.013
- Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin and Review*, 6, 57–86. doi: 10.3758/BF03210812
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, 5, 323–370. doi: 10.1037/1089-2680.5.4.323
- Bayen, U. J., & Murnane, K. (1996). Aging and the use of perceptual and temporal information in source memory tasks. *Psychology and Aging*, 11, 293–303. doi: 10.1037/0882-7974.11.2.293
- Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 197–215. doi: 10.1037/0278-7393.22.1.197
- Bell, R., Buchner, A., & Musch, J. (2010). Enhanced old-new recognition and source memory for faces of cooperators and defectors in a social-dilemma game. *Cognition*, 117, 261–275. doi: 10.1016/j.cognition.2010.08.020
- Bell, R., Dentale, S., Buchner, A., & Mayr, S. (2010). ERP correlates of the irrelevant sound effect. *Psychophysiology*, 47, 1182–1191. doi: 10.1111/j.1469-8986.2010.01029.x
- Bellebaum, C., & Daum, I. (2008). Learning-related changes in reward expectancy are reflected in the feedback-related negativity. *European Journal of Neuroscience*, 27, 1823–1835. doi: 10.1111/j.1460-9568.2008.06138.x
- Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience*, 8, 551–565. doi: 10.1162/jocn.1996.8.6.551
- Boksem, M. A. S., & De Cremer, D. (2010). Fairness concerns predict medial frontal negativity amplitude in ultimatum bargaining. *Social Neuroscience*, 5, 118–128. doi: 10.1080/17470910903202666
- Bröder, A., & Meiser, T. (2007). Measuring source memory. *Zeitschrift für Psychologie/Journal of Psychology*, 215, 52–60. doi: 10.1027/0044-3409.215.1.52
- Brosch, T., Sander, D., Pourtois, G., & Scherer, K. R. (2008). Beyond fear: Rapid spatial orienting toward positive emotional stimuli. *Psychological Science*, 19, 362–370. doi: 10.1111/j.1467-9280.2008.02094.x
- Buchner, A., Bell, R., Mehl, B., & Musch, J. (2009). No enhanced recognition memory, but better source memory for faces of cheaters. *Evolution and Human Behavior*, 30, 212–224. doi: 10.1016/j.evolhumbehav.2009.01.004
- Carlisle, E., & Shafir, E. (2005). Questioning the cheater-detection hypothesis: New studies with the selection task. *Thinking and Reasoning*, 11, 97–122. doi: 10.1080/13546780442000079
- Cauquil, A. S., Edmonds, G. E., & Taylor, M. J. (2000). Is the face-sensitive N170 the only ERP not affected by selective attention? *NeuroReport*, 11, 2167–2171. doi: 10.1097/00001756-200007140-00021
- Chiappe, D., Brown, A., Dow, B., Koontz, J., Rodriguez, M., & McCulloch, K. (2004). Cheaters are looked at longer and remembered better than cooperators in social exchange situations. *Evolutionary Psychology*, 2, 108–120. doi: 10.1177/147470490400200117

- Clark, K., & Sefton, M. (2001). The sequential prisoner's dilemma: Evidence on reciprocity. *Economic Journal*, 111, 51–68. doi: 10.1111/1468-0297.00588
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.
- Cosmides, L., & Tooby, J. (1989). Evolutionary psychology and the generation of culture: II. Case study: A computational theory of social exchange. *Ethology and Sociobiology*, 10, 51–97. doi: 10.1016/0162-3095(89)90013-7
- Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 163–228). New York, NY: Oxford University Press.
- Dugatkin, L. A. (2002). Animal cooperation among unrelated individuals. *Naturwissenschaften*, 89, 533–541. doi: 10.1007/s00114-002-0379-y
- Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). FACES—A database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior Research Methods*, 42, 351–362. doi: 10.3758/BRM.42.1.351
- Erdfelder, E., Auer, T.-S., Hilbig, B. E., Abfal, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models. A review of the literature. *Zeitschrift für Psychologie/Journal of Psychology*, 217, 108–124. doi: 10.1027/0044-3409.217.3.108
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191. doi: 10.3758/BF03193146
- Gächter, S., & Thöni, C. (2005). Social learning and voluntary cooperation among like-minded people. *Journal of the European Economic Association*, 3, 303–314. doi: 10.1162/jeea.2005.3.2-3.303
- Galli, G., Feurra, M., & Viggiano, M. P. (2006). “Did you see him in the newspaper?”: Electrophysiological correlates of context and valence in face processing. *Brain Research*, 1119, 190–202. doi: 10.1016/j.brainres.2006.08.076
- Giang, T., Bell, R., & Buchner, A. (2012). Does facial resemblance enhance cooperation? *PLoS ONE*, 7, e47809. doi: 10.1371/journal.pone.0047809
- Goeleven, E., De Raedt, R., Leyman, L., & Verschuere, B. (2008). The Karolinska directed emotional faces: A validation study. *Cognition and Emotion*, 22, 1094–1118. doi: 10.1080/02699930701626582
- Hajcak, G., Holroyd, C. B., Moser, J. S., & Simons, R. F. (2005). Brain potentials associated with expected and unexpected good and bad outcomes. *Psychophysiology*, 42, 161–170. doi: 10.1111/j.1469-8986.2005.00278.x
- Hajcak, G., Moser, J. S., Holroyd, C. B., & Simons, R. F. (2006). The feedback-related negativity reflects the binary evaluation of good versus bad outcomes. *Biological psychology*, 71, 148–154. doi: 10.1016/j.biopsycho.2005.04.001
- Hajcak, G., Moser, J. S., Holroyd, C. B., & Simons, R. F. (2007). It's worse than you thought: The feedback negativity and violations of reward prediction in gambling tasks. *Psychophysiology*, 44, 905–912. doi: 10.1111/j.1469-8986.2007.00567.x
- Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., ... Ziker, J. (2006). Costly punishment across human societies. *Science*, 312, 1767–1770. doi: 10.1126/science.1127333 16794075
- Hilbig, B. E., Zettler, I., & Heydasch, T. (2012). Personality, punishment and public goods: Strategic shifts towards cooperation as a matter of dispositional honesty-humility. *European Journal of Personality*, 26, 245–254. doi: 10.1002/per.830
- Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, 109, 679–709. doi: 10.1037/0033-295X.109.4.679
- Hruschka, D. J., & Henrich, J. (2006). Friendship, cliquishness, and the emergence of cooperation. *Journal of Theoretical Biology*, 239, 1–15. doi: 10.1016/j.jtbi.2005.07.006
- Keser, C., & van Winden, F. (2000). Conditional cooperation and voluntary contributions to public goods. *Scandinavian Journal of Economics*, 102, 23–39. doi: 10.1111/1467-9442.00182
- Kroneisen, M., Woehe, L., & Rausch, L. S. (2015). Expectancy effects in source memory: How moving to a bad neighborhood can change your memory. *Psychonomic Bulletin & Review*, 22, 179–89. doi: 10.3758/s13423-014-0655-9
- Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud faces database. *Cognition and Emotion*, 24, 1377–1388. doi: 10.1080/02699930903485076
- Linden, D. E. J. (2005). The p300: Where in the brain is it produced and what does it tell us? *Neuroscientist*, 11, 563–76. doi: 10.1177/1073858405280524
- Lundqvist, D., Flykt, A., & Öhman, A. (1998). *The Karolinska directed emotional faces—KDEF*. Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, Solna, Sweden.
- MacKenzie, G., & Donaldson, D. I. (2007). Dissociating recollection from familiarity: Electrophysiological evidence that familiarity for faces is associated with a posterior old/new effect. *NeuroImage*, 36, 454–463. doi: 10.1016/j.neuroimage.2006.12.005
- MacKenzie, G., & Donaldson, D. I. (2009). Examining the neural basis of episodic memory: ERP evidence that faces are recollected differently from names. *Neuropsychologia*, 47, 2756–2765. doi: 10.1016/j.neuropsychologia.2009.05.025 19505485
- Mayr, S., Niedeggen, M., Buchner, A., & Pietrowsky, R. (2003). ERP correlates of auditory negative priming. *Cognition*, 90, 11–21. doi: 10.1016/S0010-0277(03)00142-2
- Mealey, L., Daoud, C., & Krage, M. (1996). Enhanced memory for faces of cheaters. *Ethology and Sociobiology*, 17, 119–128. doi: 10.1016/0162-3095(95)00131-X
- Milinski, M., & Wedekind, C. (1998). Working memory constrains human cooperation in the prisoner's dilemma. *Proceedings of the National Academy of Sciences*, 95, 13755–13758. doi: 10.1073/pnas.95.23.13755
- Miner, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments & Computers*, 36, 630–633. doi: 10.3758/BF03206543
- Montalan, B., Caharel, S., Personnaz, B., Le Dantec, C., Germain, R., Bernard, C., ... Rebai, M. (2008). Sensitivity of N170 and late positive components to social categorization and emotional valence. *Brain Research*, 1233, 120–128. doi: 10.1016/j.brainres.2008.07.055 18691559
- Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods*, 42, 42–54. doi: 10.3758/BRM.42.1.42
- Paller, K. A., Gonsalves, B., Grabowecy, M., Bozic, V. S., & Yamada, S. (2000). Electrophysiological correlates of recollecting faces of known and unknown individuals. *NeuroImage*, 11, 98–110. doi: 10.1006/ning.1999.0521
- Paller, K. A., Ranganath, C., Gonsalves, B., LaBar, K. S., Parrish, T. B., Gitelman, D. R., ... Reber, P. J. (2003). Neural correlates of person recognition. *Learning & Memory*, 10, 253–260. doi: 10.1101/lm.57403
- Polezzi, D., Daum, I., Rubaltelli, E., Lotto, L., Civai, C., Sartori, G., & Rumiati, R. (2008). Mentalizing in economic decision-making. *Behavioural Brain Research*, 190, 218–223. doi: 10.1016/j.bbr.2008.03.003
- Polich, J. (2007). Updating p300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, 118, 2128–2148. doi: 10.1016/j.clinph.2007.04.019
- Rilling, J., Gutman, D., Zeh, T., Pagnoni, G., Berns, G., & Kilts, C. (2002). A neural basis for social cooperation. *Neuron*, 35, 395–405. doi: 10.1016/S0896-6273(02)00755-9
- Sanfey, A. G. (2007). Social decision-making: Insights from game theory and neuroscience. *Science*, 318, 598–602. doi: 10.1126/science.1142996 17962552
- Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science*, 300, 1755–1758. doi: 10.1126/science.1082976 12805551
- Schweinberger, S. R. (2011). Neurophysiological correlates of face recognition. In J. Calder, G. Rhodes, M. H. Johnson, & J. V. Haxby (Eds.), *The handbook of face perception* (pp. 345–366). Oxford, UK: Oxford University Press. doi: 10.1093/oxfordhb/9780199559053.001.0001
- Skowronski, J. J., & Carlston, D. E. (1987). Social judgment and social memory: The role of cue diagnosticity in negativity, positivity, and extremity biases. *Journal of Personality and Social Psychology*, 52, 689–699. doi: 10.1037/0022-3514.52.4.689
- Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, 117, 34–50. doi: 10.1037/0096-3445.117.1.34
- Sperber, D., Cara, F., & Girotto, V. (1995). Relevance theory explains the selection task. *Cognition*, 57, 31–95. doi: 10.1016/0010-0277(95)2895-9

- Suzuki, A., & Suga, S. (2010). Enhanced memory for the wolf in sheep's clothing: Facial trustworthiness modulates face-trait associative memory. *Cognition*, *117*, 224–229. doi: 10.1016/j.cognition.2010.08.004
- Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., . . . Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, *168*, 242–249. doi: 10.1016/j.psychres.2008.05.006 19564050
- Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, *46*, 35–57. doi: 10.1086/406755
- Wang, Y., Roberts, K., Yuan, B., Zhang, W., Shen, D., & Simons, R. (2013). Psychophysiological correlates of interpersonal cooperation and aggression. *Biological Psychology*, *93*, 386–391. doi: 10.1016/j.biopsycho.2013.04.008 23618597
- Wu, Y., & Zhou, X. (2009). The P300 and reward valence, magnitude, and expectancy in outcome evaluation. *Brain Research*, *1286*, 114–122. doi: 10.1016/j.brainres.2009.06.032 19539614
- Yeung, N., & Sanfey, A. G. (2004). Independent coding of reward magnitude and valence in the human brain. *Journal of Neuroscience*, *24*, 6258–6264. doi: 10.1523/JNEUROSCI.4537-03.2004
- Yovel, G., & Paller, K. A. (2004). The neural basis of the butcher-on-the-bus phenomenon: When a face seems familiar but is not remembered. *NeuroImage*, *21*, 789–800. doi: 10.1016/j.neuroimage.2003.09.034

(RECEIVED May 13, 2015; ACCEPTED September 16, 2015)