Social Cognitive and Affective Neuroscience Advance Access published October 19, 2008

doi:10.1093/scan/nsn034

Knowing when to trust others: An ERP study of decision making after receiving information from unknown people

Cheryl Boudreau,[1,2] **Mathew D. McCubbins,**[2,3] **and Seana Coulson**[2,4]
[1]Political Science Department, University of California, Davis, CA 95616, [2]Kavli Institute for Brain and Mind, [3]Political Science Department, and [4]Cognitive Science Department, University of California, San Diego, CA 92093, USA

To address the neurocognitive mechanisms that underlie choices made after receiving information from an anonymous individual, reaction times (Experiment 1) and event-related brain potentials (Experiment 2) were recorded as participants played three variants of the coin toss game. In this game, participants guess the outcomes of unseen coin tosses after a person in another room (dubbed 'the reporter') observes the coin toss outcomes and then sends reports (which may or may not be truthful) to participants about whether the coins landed on heads or tails. Participants knew that the reporter's interests were aligned with their own (common interests), opposed to their own (conflicting interests) or opposed to their own, but that the reporter was penalized every time he or she sent a false report about the coin toss outcome (penalty for lying). In the common interests and penalty for lying conditions, participants followed the reporter's reports over 90% of the time, in contrast to <59% of the time in the conflicting interests condition. Reaction time results indicated that participants took similar amounts of time to respond in the common interests and penalty for lying conditions and that they were reliably faster than in the conflicting interests condition. Event-related potentials timelocked to the reporter's reports revealed a larger P2, P3 and late positive complex response in the common interests condition than in the other two, suggesting that participants' brains processed the reporter's reports differently in the common interests condition relative to the other two conditions. Results suggest that even when people behave as if they trust information, they consider communicative efforts of individuals whose interests are aligned with their own to be slightly more informative than those of individuals who are made trustworthy by an institution, such as a penalty for lying.

Keywords: social information transmission; P300; decision making; game theory; social cognition; trust

According to a Chinese proverb, ‘A wise man makes his own decisions, an ignorant man follows the public opinion.’ Unfortunately, as citizens in modern democracies, we do not always have enough time, energy and information to make independent decisions. Thus, in many political, legal and economic contexts, citizens must make decisions based on the statements of other people (Sniderman *et al.*, 1991; Lupia, 1992, 1994; Mondak, 1993; Lupia and McCubbins, 1998; Boudreau, 2006). For example, when choosing among different candidates for office, uninformed citizens must often base their decisions on the statements of politicians and the endorsements of interest groups. When deciding a question at a trial, jurors must rely upon the statements of competing attorneys and witnesses. Similarly, when choosing among products, consumers often rely on information provided by endorsers, such as Consumer Reports and the Better Business Bureau, as well as information provided by the sellers themselves. In making such decisions, people must evaluate information gleaned from people they do not personally know, and whose interests may or may not be aligned with their own.

Most research on the neural basis of decision making has focused on the neurocognitive mechanisms of reward processing, especially learning, and the way in which individuals modulate their behavior based on their evaluation of the outcomes of prior decisions (Knutson and Bossaerts, 2007; Wickens *et al.*, 2007). Only recently have cognitive neuroscientists begun to reckon with the fact that a major source of uncertainty in human decision making derives from the fact that decisions are made in social and institutional contexts. In neuroeconomics, for example, researchers have investigated the neural responses associated with cooperation and competition between people as they engage in interactive economic games, such as the ultimatum game (Sanfey *et al.*, 2003), the prisoner’s dilemma (Rilling *et al.*, 2004) and the trust game (McCabe *et al.*, 2001; de Quervain *et al.*, 2004; Zak *et al.*, 2004; Delgado *et al.*, 2005).

Such studies have shown that the brain treats money much as it does more biologically basic reinforcers, so that performance in social economic games can be partially explained by activity in the reward systems of the brain. For example, the decision to trust another individual has been linked to levels of the neuropeptide oxytocin

Received 24 August 2007; Accepted 5 September 2008

The authors thank the National Science Foundation (grant #SES-0616904) and the Kavli Institute for Brain and Mind for generously providing financial support for these experiments.

Correspondence should be addressed to Cheryl Boudreau, Department of Political Science, University of California, Davis, 1246 Social Sciences and Humanities, One Shields Avenue, Davis, CA 95616; San Diego, CA 92093. E-mail: clboudreau@ucdavis.edu.

(Kosfeld *et al*., 2005). The reciprocation of one's trust by another person results in greater striatal activation than does 'reciprocation' by a computer, suggesting that—over and above any financial benefits—interaction with a cooperative agent is itself rewarding (Rilling *et al*., 2002, 2004). Further, when people engage in multiple rounds of the same economic exchange game, activity in the caudate nucleus is positively correlated with the recognition of benevolent reciprocity by one's partner, as well as with subsequent trusting behavior, i.e. awarding one's partner a larger sum of money in the next round of play (King-Casas *et al*., 2005).

However, what happens when people exchange information with one another rather than money? In such anonymous exchanges, what are the conditions under which citizens can trust the statements of others in order to pursue their goals? In the present study, we address the cognitive and neural mechanisms that underlie choices made after receiving information from another individual. Our design is motivated by formal models in political science and economics that demonstrate the conditions under which people, first, trust the statements of individuals personally unknown to them (dubbed 'reporters' throughout this paper) and, second, base their decisions on the statements of these reporters. Specifically, mathematical models by Crawford and Sobel (1982) and Lupia and McCubbins (1998) demonstrate that, in equilibrium, common interests between a knowledgeable reporter and citizens induce the reporter to tell the truth and the citizens to trust the reporter's statements and base their choices upon them. However, when a knowledgeable reporter and citizens have conflicting interests, these models suggest that citizens should ignore the reporter's statements and make their decisions on their own (Lupia and McCubbins, 1998). Lupia and McCubbins (1998) have also shown that institutions can sometimes induce a reporter to tell the truth, even when his or her interests conflict with those of citizens. For example, an appropriately large penalty for lying can remove a reporter's incentive to lie and lead citizens to trust the reporter's statements (Lupia and McCubbins, 1998).

Lupia and McCubbins (1998) tested their game theoretic model in a series of behavioral experiments involving two kinds of participants: reporters and listeners. In those experiments, listeners were asked to guess the outcome of an unseen coin toss, and they were told that a person unknown to them and unseen by them (i.e. 'the reporter') would observe the coin toss outcome and send a report (which may or may not be truthful) to them about whether the coin landed on heads or tails. The presence of common vs. conflicting interests was varied by manipulating the financial incentives of both types of participants. In all conditions, the reporters earned money based on the listeners' performance, while the listeners earned money based on their own performance. For example, in the 'common interests' condition, listeners were told that both they and the reporter earned money when the listener correctly guessed the outcome of the coin toss. In the 'conflicting interests' condition, listeners were told that they would earn money when they, themselves, guessed correctly, but that the reporter would earn money only when the listeners guessed incorrectly. As predicted, reporters were much more likely to send truthful reports in the common than in the conflicting interests condition (Lupia and McCubbins, 1998). Further, listeners who perceived that the reporter shared common interests with them were significantly more likely to trust the reporter's statements than those who perceived that the reporter's interests conflicted with their own.

A third condition, the 'penalty for lying' condition, was similar to the conflicting interests condition in that reporters earned money whenever listeners guessed incorrectly. However, reporters lost money whenever their reports to the listener were not truthful. In some settings, the penalty for lying was more than the amount that reporters could earn for an incorrect response by the listener, so reporters in this condition were for the most part truthful in their reports to the listeners. Additionally, the listeners trusted the reporter's statements at a rate that was similar to that in the common interests condition.[1] Based upon these theoretical and experimental results, Lupia and McCubbins (1998) concluded that institutions can substitute for common interests because they, too, induce the reporter to make truthful statements and enable citizens to trust and learn from these statements.

Here we present two experiments that test the model presented by Lupia and McCubbins (1998). In Experiment 1, we collected reaction times from listeners in Lupia and McCubbins's coin toss game, and found that participants took a similar amount of time to respond to reporters with common interests and reporters in the penalty for lying condition. Moreover, reaction times in these two conditions were reliably faster than in the conflicting interests condition. In Experiment 2, we recorded event-related potentials (ERPs) from listeners in the coin toss game as they processed information conveyed by anonymous reporters whose trustworthiness was determined by these social conditions. Results indicated that participants' brains processed reports differently in the common interests condition relative to the other two conditions.

EXPERIMENT 1

As in the study by Lupia and McCubbins (1998), participants in Experiment 1 were asked to predict the outcomes of coin tosses that happened in a separate room. They were instructed that they would earn 50 cents for each correct prediction that they made, and nothing when they made an incorrect prediction. They were then told that a participant in another room (i.e. 'the reporter') would observe each coin toss outcome and then send a report to them via computer about whether the coin landed on heads or tails.

[1] Lupia and McCubbins (1998) also tested the effects of other external forces, such as the threat of verification by a third party and costly effort by the reporter.

Participants were also told that the reporter could either lie about the coin toss outcome or tell the truth. Thus, before participants made a prediction about each coin toss, they observed the reporter's report of whether the coin landed on heads or tails but did not know whether the report was truthful.

As in study by Lupia and McCubbins (1998), the key factor that was manipulated in the present study was the perceived trustworthiness of the reporter. To do so, two things were varied: the interests of the reporter and the participants, and the institutional context in which the reporter sent his or her report. In the common interests condition, participants were told that they would earn 50 cents each time they correctly predicted the coin toss outcome and that the reporter would also earn 50 cents each time participants correctly predicted the coin toss outcome. Participants were also instructed that neither they nor the reporter would earn any money when they (the participants) made an incorrect prediction.

In the conflicting interests condition, participants were told that they still earned 50 cents each time they made a correct prediction. However, they were also told that the reporter earned 50 cents each time they (the participants) made an incorrect prediction about the coin toss outcome, and earned nothing each time they (the participants) made a correct prediction.

In the penalty for lying condition, conflicting interests between the reporter and the participants were maintained, but an institution was also imposed upon the reporter, namely, a penalty for lying. Specifically, participants were told that $1 would be subtracted from the reporter's experimental earnings each time that he or she lied about the coin toss outcome. Because participants were told how the reporter earned money, they should in principle know that the $1 penalty would ensure that the reporter always had an incentive to tell the truth about the coin toss outcome.

METHODS

Participants

Forty-seven adults from the University of California, Davis community (29 men), aged 18–26, were paid based on the decisions they made in our experiment. All participants were healthy, and they earned, on average, $27.

Procedure

Upon entry into the experimental lab, participants were asked to predict whether several practice coin tosses landed on heads or tails and were paid 50 cents for each correct prediction they made. The purpose of these practice predictions was to ensure that participants understood that they would earn money based upon the choices they made in the experiment.

Following these initial coin tosses, participants were read the instructions for the common interests condition. Participants were told that a coin would be flipped in a different room and shown to another anonymous experimental participant who was referred to as the reporter. Upon seeing the outcome of the coin flip, the reporter would respond either HEADS or TAILS, and this message would appear on the computer monitor in front of the participant. Participants were told that they and the reporter would earn 50 cents every time they, the participants, correctly predicted the coin toss outcome, and nothing if they predicted incorrectly or failed to respond before the onset of the next trial. Participants were explicitly told that it was entirely the decision of the reporter as to whether he or she would send a truthful report via the computer.

Although participants were told that there was another person acting as 'the reporter' in another room, the reporter's reports of heads or tails in each experimental block were actually based upon the results of Lupia and McCubbins (1998) and were thus programmed into the computer before the experiment began. That said, many precautions were taken to ensure that participants believed that there was another person acting as the reporter. For example, the experimenter left the experimental laboratory between blocks, ostensively to make sure the reporter was ready to begin the next set of trials. The amount of time that it took for the reporter's reports to appear on participants' computer screen was long enough for us to credibly state that another participant was in another room sending reports via computer. Further, when debriefing participants at the end of the experiment, none expressed skepticism regarding the existence of a real reporter.

To ensure that participants fully understood the instructions, they were given several quiz questions in which they were asked to say how much money the reporter would earn under various circumstances. To motivate performance on the quiz, participants were paid 25 cents for each quiz question they answered correctly. When the experimenter was sure that participants understood how the reporter earned money in the common interests block, 10 experimental trials began.

Following the initial common interests trials, participants were read the instructions for the conflicting interests condition. Participants were told that their task would be the same as in the previous block of trials—to predict the outcome of an unseen coin toss after receiving a message from a reporter. Participants were told that while they themselves would still earn 50 cents for each correctly predicted coin toss and nothing for incorrect predictions, the reporter would now earn 50 cents for each incorrect prediction made by the participant. Participants were given a brief quiz about how much money the reporter would earn under various circumstances and were paid 25 cents for each correctly answered quiz question. When the experimenter was sure the participant understood how the reporter earned money, 10 conflicting interests trials began.

Following the initial conflicting interests trials, participants were read the instructions for the penalty for lying condition. Participants were told that as in the previous (conflicting interests) block, the reporter earned 50 cents for each of the participant's incorrect predictions, while the participant would earn 50 cents for each correct prediction. Participants were also told that every time the reporter sent a false report, $1 would be deducted from the reporter's earnings. Participants were given a brief quiz about how much money the reporter would earn under various circumstances and were paid 25 cents for each correctly answered quiz question. When the experimenter was sure the participant understood how the reporter earned money, 10 penalty for lying trials began.

Once the participants had completed 10 trials for all three conditions, we collected data for an additional block of 10 trials in each of our three conditions. In order to control for block order effects, half the participants completed the second block of trials in order 1 (common interests, conflicting interests, penalty for lying), while the other half completed the second block of trials in order 2 (penalty for lying, conflicting interests, common interests). In total, participants completed 20 trials in each of the three conditions.

In all three conditions (common interests, conflicting interests and penalty for lying), participants were seated in a comfortable chair in front of a computer screen. As shown in Figure 1, all trials in our experiment began with the text '(Tossing Coin)' appearing in the center of a 19 in. color monitor for 3 s. Next, the text 'Showing outcome to reporter' was displayed on the monitor for 5 s. The text 'The reporter says' then appeared for 5 s. These first three prompts appeared on the screen for a longer amount of time (a total of 13 s) than subsequent prompts in order to promote the illusion that another experimenter was actually flipping a coin in another room and that another participant (acting as the reporter) was actually sending a report. The reporter's report comprised a 1 s presentation of the word 'HEADS' or 'TAILS'. Participants were given 6 s from the onset of the 'HEADS/TAILS' prompt to respond. Response was signaled via a button press in which a left-hand response indicated HEADS and a right-hand response indicated TAILS. This sequence was repeated for each of the 60 trials in our experiment, and on each trial, the time that elapsed between the presentation of the 'HEADS/TAILS' prompt and participants' responses was recorded. Participants were not told that their responses were being timed, and no feedback was given until the very end of the experiment.[2] Participants were also told that the reporter would not observe the outcome of individual trials.

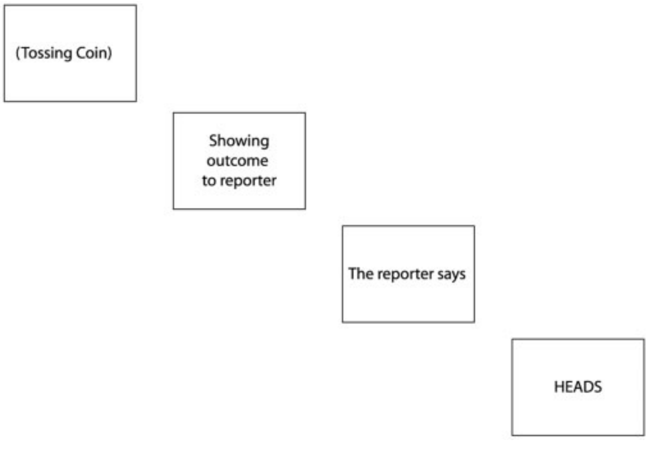

Fig. 1 Sample trial from Experiment 1.

RESULTS

Behavioral results

The extent to which participants trusted the reporter's reports was assessed by examining the percentage of times that their predictions were the same as what the reporter reported in each experimental condition (i.e. what percentage of the time did participants predict heads when the reporter reported heads and predict tails when the reporter reported tails in each experimental condition). One-sample *t*-tests were used to determine whether participants' predictions matched what the reporter reported >50% of the time. A 50% baseline was used because we tossed a fair coin; thus, if participants were simply choosing heads or tails randomly, then we would expect their predictions to match the reporter's reports 50% of the time. If participants trusted the reporter's reports, then we should observe their predictions matching the reporter's reports >50% of the time.

As shown in Figure 2, when participants were told that the reporter shared common interests with them, their predictions matched what the reporter reported 92% of the time, a figure that is significantly greater than our 50% baseline ($t = 18.93$, $P < 0.001$). Similarly, in the penalty for lying condition, participants' predictions matched the reporter's reports 93% of the time, which is also significantly >50% ($t = 18.99$, $P < 0.001$). However, in the conflicting interests condition, participants' predictions matched what the reporter reported only 58% of the time. Although this percentage is significantly greater than our 50% baseline ($t = 2.13$, $P < 0.05$), it is significantly less than the percentage of times that participants' predictions matched the reporter's reports in the common interests and penalty for lying conditions (when compared to the common interests condition: $t = 7.79$, $P < 0.001$; when compared to the penalty for lying condition: $t = 8.03$, $P < 0.001$). These results are largely consistent with those of Lupia and McCubbins (1998).[3]

Reaction time results

The reaction time results were consistent with our behavioral results. That is, participants in the common interests and

[2] Because participants were not given feedback about their choices, they could not verify whether the reporter's reports were truthful. Thus, there was no opportunity for the reporter to develop a reputation in our experiments. This aspect of our experimental design differs from those of other scholars, who did allow reputations to develop in their experiments (see, e.g. Andreoni, 1988; Palfrey and Prisbrey 1996; McCabe *et al*., 2001; Milinski *et al*., 2002; King-Casas *et al*., 2005; Izuma *et al*., 2008).

[3] Our data were also analyzed with a random effects logit model, and the same results were obtained. The random effects model was used to account for participants making multiple choices in each experimental condition. That the same results were obtained suggests that the results are not driven by unobserved individual differences (i.e. the errors appear to be approximately Gaussian).

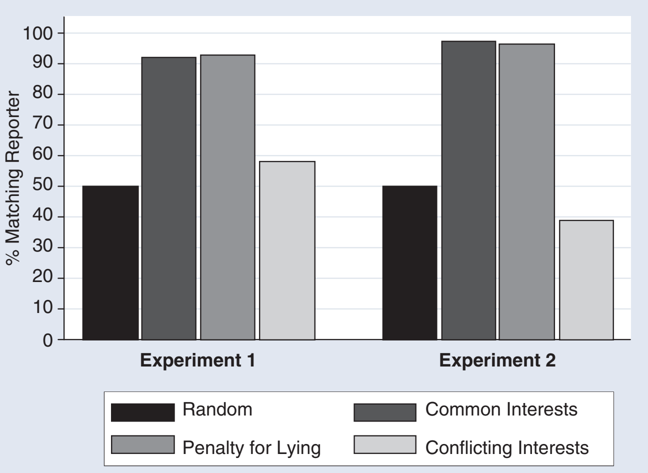

Fig. 2 Percentage of participants' predictions that matched the reporter's reports in the common interests, conflicting interests and penalty for lying conditions in Experiments 1 and 2.

penalty for lying conditions took similar amounts of time to respond after viewing the 'HEADS/TAILS' prompt. Further, participants in the conflicting interests condition were slower to respond than were participants in the other two conditions. Specifically, participants in the common interests condition took, on average, 1191 ms to register their response of 'heads' or 'tails', while participants in the penalty for lying condition took, on average, 1157 ms to register their response. This difference is not statistically significant ($t = 0.41$). Participants in the conflicting interests condition, however, took, on average, 1318 ms to respond, which is significantly slower than participants in penalty for lying and common interests conditions (when compared to the penalty for lying condition: $t = 1.84$, $P < 0.05$; when compared to the common interests condition: $t = 1.44$, $P < 0.1$).

DISCUSSION

Results of Experiment 1 replicate and extend the findings of Lupia and McCubbins (1998). Participants' responses suggest they mirror the information sent by reporters in the common interests and penalty for lying conditions, and largely disregard information sent by reporters in the conflicting interests condition. Moreover, listeners' response times were similar in the common interests and penalty for lying conditions and were reliably faster in these blocks than when they received information from reporters with conflicting interests.

EXPERIMENT 2

Experiment 1 showed that, behaviorally, participants act similarly on information from reporters who share common interests with them and information from reporters who are made trustworthy by an external institution, such as a penalty for lying. However, previous research in cognitive neuroscience has shown that similar behavioral outcomes can be subserved by different neural mechanisms, as for example when older adults engage more brain regions than younger adults in order to achieve similar performance levels on perceptual or memory tasks (Grady *et al.*, 1992; Reuter-Lorenz *et al.*, 2000). Further, whereas reaction times measure the endpoint of participants' decision process, event-related brain potentials can provide a measure of various perceptual and cognitive processes involved in the generation of participants' responses.

Thus, in Experiment 2, we recorded ERPs from listeners in the coin toss game in order to investigate the brain's real-time response to information provided by a reporter with common interests, a reporter with conflicting interests and a reporter who was made trustworthy by an external institution, namely a penalty for lying. Our predictions focused on the amplitude of the P3 (P3b) component of the ERPs. As a broadly distributed positivity with a centroparietal maximum, the P3 is associated with stimulus evaluation and is typically elicited for stimuli that require a binary response (Donchin and Coles, 1988).

The most influential models of the psychological processes underlying the P3 include Johnson's (1986, 1988) triarchic model and Donchin's context updating model (Donchin, 1981; Donchin and Coles, 1988). In the triarchic model, P3 amplitude is sensitive to subjective probability, stimulus meaning and information transmission. In the context updating model, P3 amplitude is proportional to the level of uncertainty in participants' prior expectations about the stimulus. In both models, then, P3 amplitude is related to the information value of the stimulus.

Based on the results of Lupia and McCubbins (1998) and those in Experiment 1, we predicted that participants would be more likely to trust the reporter's statements when the reporter shared common interests with them than when the reporter's interests conflicted with their own, and thus reports would elicit a larger P3 response in the common interests than in the conflicting interests condition. Further, because Lupia and McCubbins (1998) demonstrated that institutions can substitute for common interests, we predicted that the P3 response to reports in the penalty for lying condition would be similar in amplitude to that in the common interests condition. In sum, because we assumed that participants would consider the reporter's reports to be more informative in the common interests and penalty for lying conditions, we expected a larger P3 in those conditions than in the conflicting interests condition, where participants largely ignore the reporter's reports.

METHODS

Participants

Twelve adults from the University of California, San Diego, community (eight men), aged 19–28, were paid based on the decisions that they made in our experiment. All participants were healthy, and they earned, on average, $60.

Procedure

The procedure used in Experiment 2 was nearly identical to the procedure used in Experiment 1, with the main difference being that participants' electroencephalogram (EEG) was recorded during Experiment 2. This difference necessitated other minor modifications to the procedure used in Experiment 1. Specifically, participants in Experiment 2 completed the experiment one at a time from within an EEG booth, whereas participants in Experiment 1 completed the experiment in an experimental laboratory, which seated between two and six other participants. Further, after predicting the outcomes of 10 coin tosses in the common interests, conflicting interests, and penalty for lying conditions, participants in Experiment 2 participated in two additional blocks of 20 trials in each of our three conditions (as opposed to one additional block of 10 trials in Experiment 1). Larger trial numbers were used in Experiment 2 in order to ensure an acceptable signal to noise ratio in the ERPs. Thus, participants in Experiment 2 completed a total of 50 trials in each of the three conditions (as opposed to 20 trials in each condition in Experiment 1). As in Experiment 1, in order to control for block order effects, half the participants completed the remaining blocks of trials in order 1 (common interests, conflicting interests, penalty for lying), while the other half completed the remaining blocks of trials in order 2 (penalty for lying, conflicting interests, and common interests).

In Experiment 2, participants were seated in a comfortable chair approximately 37 in. from a computer screen. All trials began with the text '(Tossing coin)' appearing in the center of a 19 in. color monitor for 3 s, followed by a variable interstimulus interval (ISI) that ranged from 4 to 1000 ms. Variable ISI was used in Experiment 2 because of the larger number of trials that participants completed; that is, because participants predicted the outcomes of 150 coin tosses, they likely would have noticed if the experimenter always took the same amount of time to flip the coin or if the reporter always took the same amount of time to send his or her report. Thus, in order to promote the illusion that another experimenter was actually flipping a coin in another room and that another participant was actually acting as the reporter, the amount of time that it took for the coin to be tossed and for the reporter's reports to appear on participants' computer screen was randomly varied. Next, the text 'Showing Coin Toss Outcome to the Reporter' was displayed on the monitor for 5 s, followed by a variable ISI of 4–1000 ms. The text 'The Reporter Says . . .' then appeared for 1 s followed by 500 ms of blank screen. The reporter's report comprised a 500 ms presentation of the word 'HEADS' or the word 'TAILS,' followed by a 300 ms ISI. The reporter's report of heads or tails was followed by a prompt that read 'Your Guess?' for 500 ms. Participants were given 4500 ms from the onset of the 'Your Guess' prompt to respond. This sequence was repeated for each of the 150 trials in Experiment 2, and on each trial, the amount of time that elapsed between the presentation of the 'Your Guess?' prompt and participants' responses was recorded.

EEG recording and analysis

EEG, sampled at 250 Hz, was collected from 29 tin electrodes arranged in an expanded version of the 10–20 system (Nuwer *et al.*, 1998), referenced to the left mastoid. Blinks and eye movements were monitored via an electrode beneath the right eye and one electrode at each of the outer canthi (the electrooculogram, EOG). Average artifact rejection rate was 31% (s.e. = 17%). The EEG and EOG were recorded and amplified with a set of 32 bioamplifiers from SA Instruments (San Diego, CA), with half-amplitude cutoffs at 0.01 and 40 Hz and digitized on a PC. Informed consent was obtained, and all procedures conformed to ethical requirements of the University of California, San Diego.

ERPs were timelocked to the onset of the reporter's report in each of the three sorts of experimental blocks (common interests, conflicting interests and penalty for lying). The 100 ms interval preceding stimulus onset served as the baseline. ERPs were assessed via mean amplitude measurements in intervals designed to capture various components of interest (such as the N1, P2 and P3 components). Values were subjected to three sorts of repeated measures analysis of variance (ANOVA). Midline analyses involved measurements taken from channels FPz, FCz, Cz, CPz, Pz and Oz, and included within-participants factors: trustworthiness (common interests, conflicting interests and penalty for lying) and electrode site (six levels). Medial analyses involved measurements taken from channels FP1, F3, FC3, C3, CP3, P3, O1 and their RH counterparts. Factors included trustworthiness (three levels), hemisphere (left, right) and anterior/posterior (seven levels). Analogously, lateral analyses involved factors trustworthiness (three levels), hemisphere (two levels) and anterior/posterior (four levels), and utilized measurements from channels F7, FT7, TP7, T5 and their RH counterparts. Where appropriate, the Huynh–Feldt correction (Huynh and Feldt, 1978) has been applied. We report corrected *P*-values, but the original degrees of freedom have been maintained for clarity.

RESULTS

Behavioral results

As in Experiment 1, the extent to which participants trusted the reporter's reports was assessed by examining the percentage of times that their predictions were the same as what the reporter reported in each experimental condition. One-sample *t*-tests were again used to determine whether participants' predictions matched what the reporter reported >50% of the time. As shown in Figure 2, when participants were told that the reporter shared common interests with them, their predictions matched what the reporter reported 97% of the time, a figure that is significantly greater than our 50% baseline ($t = 39.28$, $P < 0.001$). However, in the conflicting interests condition where the reporter's interests conflicted with those of participants, participants' predictions matched what the reporter reported only 39% of the time, which is not significantly different from 50% ($t = 1.9$). In the penalty for lying condition, participants' predictions matched the reporter's reports 96% of the time, which is significantly >50% ($t = 23.65$, $P < 0.001$). These results are largely consistent with those of Lupia and McCubbins (1998) and those from Experiment 1.

Reaction time results

Reaction times were measured from the onset of the 'Your Guess' prompt, and analyzed with repeated measures ANOVA. In the common interests condition, average response time was 841 ms; in the conflicting interests condition, average response time was 910 ms; and in the penalty for lying condition, average response time was 901 ms. Our analysis revealed no effect of experimental condition [$F(2, 22) = 0.83$]. The absence of reaction time effects in Experiment 2 is likely attributable to our instructions to participants to wait until the 'Your Guess' prompt appeared before pressing the response button. Intended to minimize the presence of motor preparation effects in ERPs to the reporter's reports (HEADS vs. TAILS), this aspect of our design served to decrease the variance in participants' reaction times.

ERP results

Grand average ERPs to the reporter's reports in each of the three conditions can be seen in Figure 3. Prominent portions of the waveform included a negativity peaking ~100 ms poststimulus over frontal electrodes (the AN1), a frontal positivity peaking ~200 ms poststimulus (the P2), a more broadly distributed positivity peaking at ~500 ms (the P3), a negative-going peak at 600 ms (the medial negativity) and subsequent slow wave activity we refer to as the late positive complex (LPC).

AN1 component

The anterior N1 component was assessed by measuring the mean amplitude of ERPs elicited between 80 and 110 ms poststimulus. In this portion of the waveform, ERPs to stimuli in the common interests condition were less negative than ERPs to stimuli in the other two conditions (Figure 4). Measured at medial sites, the AN1 in the common interests condition was $-0.3\,\mu V$ vs $-1.2\,\mu V$ in the conflicting interests condition and $-0.7\,\mu V$ in the penalty for lying condition. The interaction between trustworthiness, hemisphere and the anterior–posterior factor (Table 1) results because the N1 response was largest at the anterior medial electrode sites and was slightly larger over the left hemisphere.

P2 component

The P2 component was assessed by measuring the mean amplitude of ERPs between 180 and 250 ms poststimulus. As shown in Table 1, repeated measures ANOVA revealed a main effect of trustworthiness at the midline and lateral electrode sites. ERPs were most positive in the common interests condition (4.9 μV at midline electrodes), compared to 3.63 μV (at midline electrodes) in the conflicting interests condition, and 3.26 μV (at midline electrodes) in the penalty for lying condition.

P3 component

The P3 component was assessed by measuring the mean amplitude of ERPs from 400 to 600 ms poststimulus. Analysis suggested that ERPs to stimuli in the common interests condition were more positive than ERPs to stimuli in either the conflicting interests condition or the penalty for lying condition (Table 1). Follow-up analysis of data recorded from midline sites revealed that the mean amplitude of ERPs in the common interests condition was 3.29 μV, which was significantly more positive than 2.27 μV in the penalty for lying condition [$F(1, 11) = 5.32$, $P < 0.05$] and more positive than 1.62 μV in the conflicting interests condition [$F(1, 11) = 15.43$, $P < 0.01$]. The amplitude difference between the penalty for lying and the conflicting interests conditions only approached significance [$F(1, 11) = 3.35$, $P = 0.09$].

Medial negativity

Medial negativity was examined by measuring the mean amplitude of ERPs from 550 to 650 ms poststimulus. Analysis suggested the mean amplitude of the ERPs was more negative in the conflicting interests condition than in either the common interests or the penalty for lying conditions (Table 1). The mean amplitude of ERPs at the midline sites in the conflicting interests condition was 0.21 μV, which was significantly more negative than 2.05 μV in the common interests condition [$F(1, 11) = 26.08$, $P < 0.01$] and 1.44 μV in the penalty for lying condition [$F(1, 11) = 7.92$, $P < 0.05$].[4]

Late positive complex

The LPC was assessed by measuring the mean amplitude of ERPs from 600 to 900 ms poststimulus. In this interval, the mean amplitude of ERPs was significantly more positive in the common interests condition, relative to the conflicting interests and the penalty for lying conditions, at both the midline and lateral sites (Table 1). *Post hoc* comparisons of measurements at midline sites revealed that ERPs in the common interests condition measured 3.18 μV, which was significantly more positive than 1.92 μV in the conflicting interests condition [$F(1, 11) = 11.28$, $P < 0.01$] and 1.73 μV in the penalty for lying condition [$F(1, 11) = 6.11$, $P < 0.05$]. The amplitude of the LPC to stimuli in the conflicting

[4] In experimental paradigms where the offset of the stimulus precedes a behaviorally significant response, stimulus offset is associated with readiness potentials in the ERPs related to response preparation (Spantekow *et al.*, 1999). As stated in the discussion section, the medial negativity observed in the present study may be associated with response preparation. In fact, both the medial negativity and the LPC may have been triggered by the stimulus offset, rather than the stimulus onset. Because of the ambiguity caused by the timing of the stimulus offset (which may have triggered ERPs that were convolved with the stimulus-generated ERPs), the medial negativity and LPC results should be interpreted with caution.

interests and the penalty for lying conditions did not reliably differ [midline: $F(1, 11) = 0.21$, $P = 0.65$].

DISCUSSION

We briefly discuss how our manipulation of the social conditions for trustworthiness affected the amplitude of ERP components elicited by the reporter's reports.

AN1 component

Remarkably, our manipulation of the trustworthiness of the reporter (via the instructions in the three different experimental conditions) affected the amplitude of ERP waveforms within 100 ms of the appearance of the reporter's report. Although the visually presented stimuli were identical in each of the three conditions, the anterior N1 (AN1) was

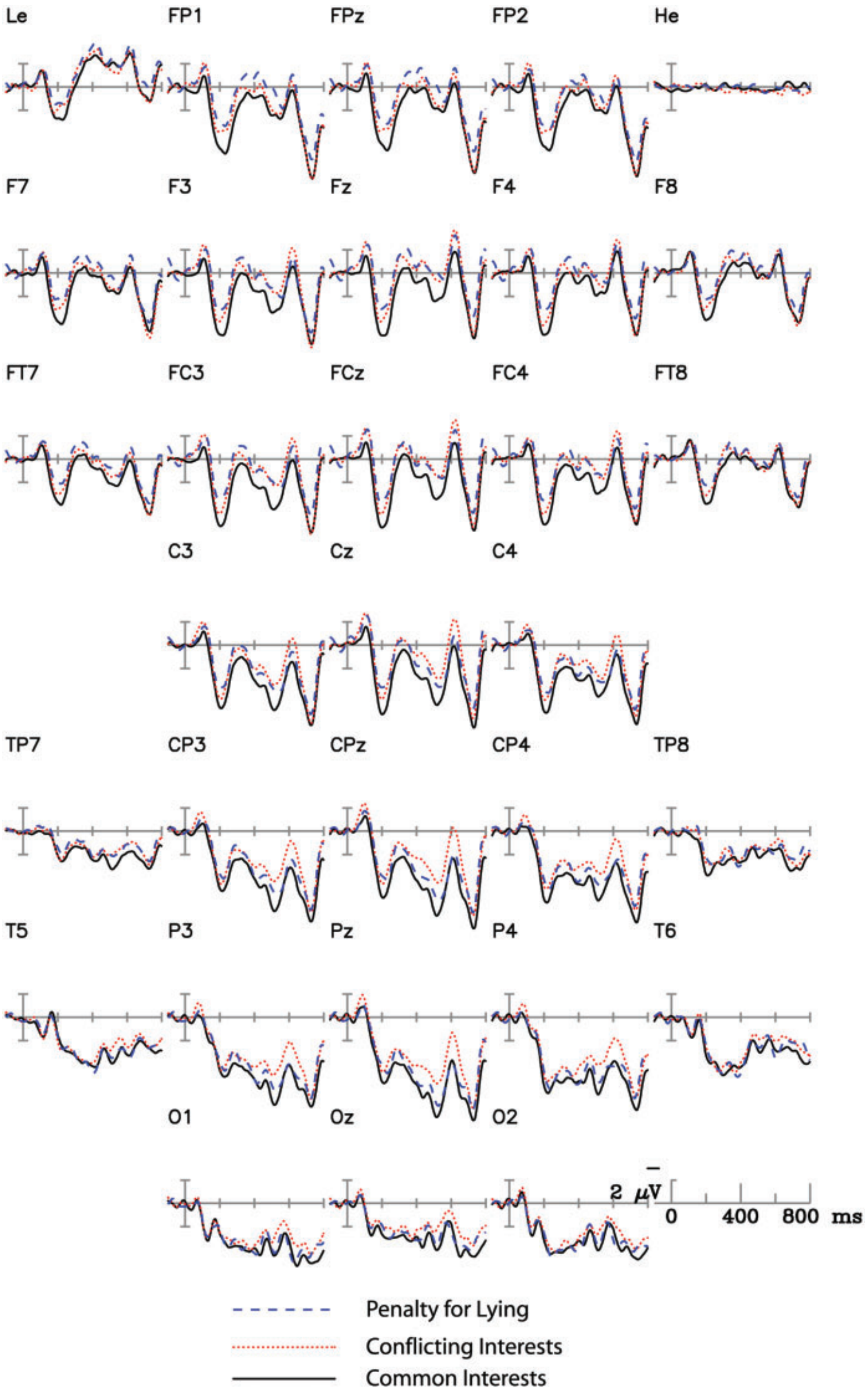

Fig. 3 Grand average ERPs to the reporter's reports in the common interests (solid line), conflicting interests (dotted line) and penalty for lying (dashed line) conditions. Negative voltage is plotted up.

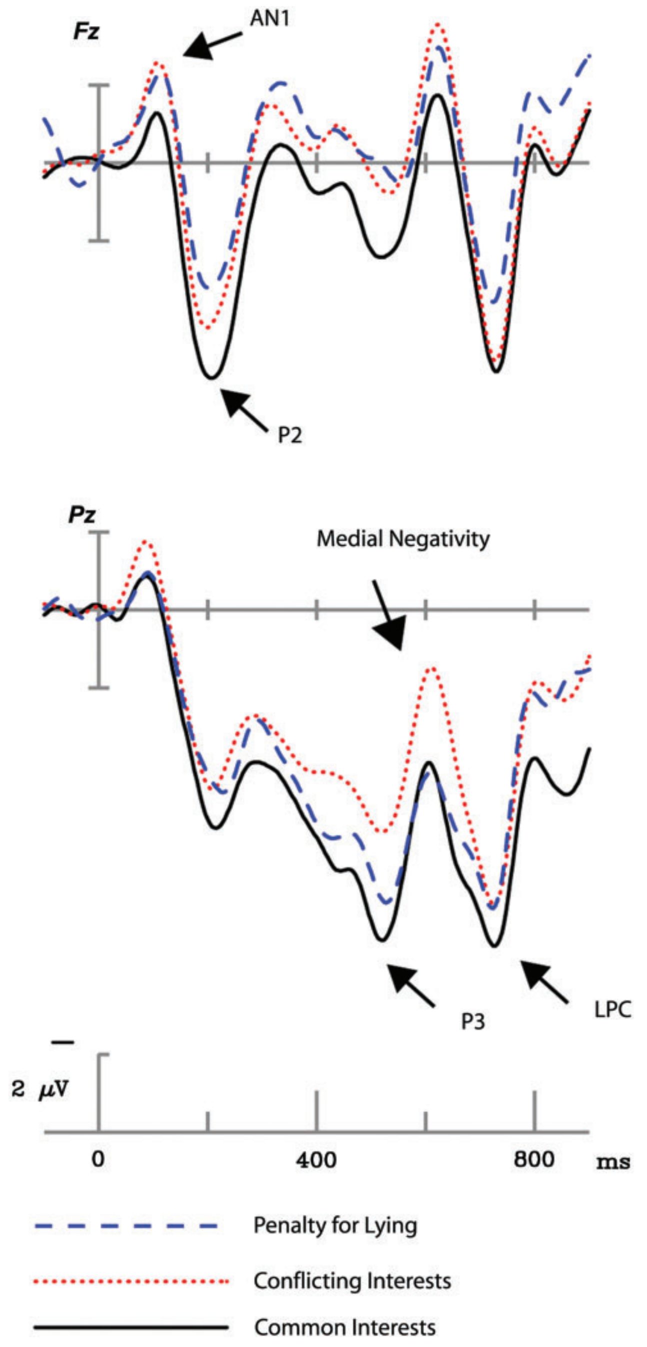

Fig. 4 Grand average ERPs recorded from the midline frontal (Fz) and parietal (Pz) electrode sites in the common interests, conflicting interests, and penalty for lying conditions. ERPs are timelocked to the reporter's reports in each experimental condition. Negative voltage is plotted up.

larger in both the conflicting interests and the penalty for lying conditions than it was in the common interests condition. The AN1 is a negativity peaking over frontocentral electrodes ~100 ms after the onset of a visually presented stimulus (Luck, 1995). It peaks slightly earlier than the posterior N1 (or N170), which is also elicited by visual stimuli but is most prominent at occipitotemporal electrode sites (Hillyard and Anllo-Vento, 1998). The amplitude of both visual N1 components is modulated by attention, being larger for stimuli in attended than in unattended locations (Clark and Hillyard, 1996). But while the posterior N1 has been shown to index visual discrimination processes, enhanced amplitude of the AN1 has been argued to reflect anticipatory motor processes related to response preparation (Vogel and Luck, 2000).

Enhanced AN1 observed here to stimuli presented in the conflicting interests and penalty for lying conditions indicates that participants' brains processed the reporter's reports differently in the common interests condition, relative to the other two conditions. This processing difference may result from greater anticipatory activity in the former two conditions, as opposed to enhanced visual attention. This suggests that participants in the common interests condition may have attempted to more fully process the reporter's reports before preparing their response.

P2 component

Similarly, the P2 component was larger in the common interests condition than in the other two conditions. Although the functional significance of this component is not completely agreed upon, the P2 has been argued to reflect some aspect of high-level perceptual processing (Kranczioch *et al.*, 2003). In target detection paradigms, the P2 is larger for targets than non-targets as defined by a variety of visual features (Hillyard and Muente, 1984; Kenemans *et al.*, 1993), leading to the suggestion that it indexes a multidimensional feature detection process (Luck and Hillyard, 1994). Noting that the P2 is modulated in overt and covert target detection paradigms, others have suggested it is primarily sensitive to the task relevance of perceptual information, and consequently argued that the P2 indexes the integration of motivational and perceptual information (Potts, 2004; Potts *et al.*, 2004, 2006). The enhanced P2 that we observed in the common interests condition is, thus, consistent with our claim that participants' brains processed the reporter's reports differently in the common interests condition, relative to both the penalty for lying and conflicting interests conditions. It may also indicate that the reporter's report was more perceptually and motivationally salient in the common interests condition, relative to the other two conditions.

P3 and LPC components

Contrary to our expectations, participants exhibited a significantly larger P3 response when they were exposed to a reporter who shared common interests with them, relative to when they were exposed to a reporter who was subject to a penalty for lying or who had conflicting interests with them. Interestingly (and unexpectedly), the size of the P3 response was more similar in the penalty for lying and conflicting interests conditions than in the penalty for lying and common interests conditions. An identical pattern of results was found for the LPC. We discuss these results further in the General Discussion section.

Table 1 Mean amplitude analysis of the AN1, P2, P3, Medial negativity and LPC components, measured at midline, medial and lateral electrode sites

	Midline sites		Medial sites		Lateral sites	
	F	*P*	*F*	*P*	*F*	*P*
AN1						
Trustworthiness × Hemis × Ant-Pos	N/A	N/A	*F* (10, 110) = 3.12	<0.05*	*F* (6, 66) = 0.76	NS
P2						
Trustworthiness	*F* (2, 22) = 4.63	<0.05*	*F* (2, 22) = 3.11	NS	*F* (2, 22) = 6.27	<0.01*
P3						
Trustworthiness	*F* (2, 22) = 9.14	<0.01*	*F* (2, 22) = 6.56	<0.01*	*F* (2, 22) = 2.47	NS
Medial negativity						
Trustworthiness	*F* (2, 22) = 9.32	<0.01*	*F* (2, 22) = 7	<0.01*	*F* (2, 22) = 2.68	NS
LPC						
Trustworthiness	*F* (2, 22) = 5.76	<0.05*	*F* (2, 22) = 3.48	NS	*F* (2, 22) = 4.45	<0.05*

The * indicates a *P*-value less than or equal to our alpha level of 0.05; NS: not significant at 0.05; N/A: not applicable (there was no Hemisphere factor in the analysis of midline electrodes).

Medial negativity

The only ERP component that showed the predicted pattern of a similar response for the common interests and the penalty for lying conditions, i.e. the two in which participants behaved similarly, was the negative waveform evident 550–650 ms postonset that we have dubbed the medial negativity. In this interval, reports in the conflicting interests condition elicited more negative ERPs over centroparietal electrodes than did reports in the other two conditions. Given its timing during the interval between the report and the response prompt, it may reflect participants' preparation to respond to actionable information.

GENERAL DISCUSSION

The present study addressed the brain's real-time response to information conveyed by reporters whose trustworthiness was determined by social conditions. We did so by recording reaction times (Experiment 1) and ERPs (Experiment 2) from participants who played the role of listeners in Lupia and McCubbins's (1998) coin toss game. In this game, participants guess the outcome of an unseen coin toss after they receive information from an anonymous reporter who knows the outcome of the coin toss, but is under no obligation to communicate it truthfully. ERPs were timelocked to the onset of the reporter's report. We expected to observe similar ERP responses to stimuli in the common interests and penalty for lying conditions, where the reporter's reports were presumed to be trustworthy, and that both would differ from ERPs to stimuli in the conflicting interests condition. Further, because reports in the common interests and penalty for lying conditions were presumed to be more informative than those in the conflicting interests condition, we predicted the latter would elicit a smaller amplitude P3 (P3b) than the other two conditions.

Results, however, indicated that while participants behaved as if reporters in the common interests and penalty for lying conditions were equally trustworthy, their brain response suggested that they processed reports differently in these two conditions. As can be seen in Figure 2, participants in both the common interests and penalty for lying conditions almost always based their predictions on the reporter's report, while participants apparently ignored the reports in the conflicting interests condition. Further, participants' reaction times were similar in the common interests and penalty for lying conditions and were faster than participants' reaction times in the conflicting interests condition. However, P3 amplitude was larger for reports in the common interests condition than it was for the other two conditions. In fact, P3 amplitude to reports in the penalty for lying condition was more similar to that of the conflicting interests condition than to the common interests condition.[5]

Based purely on the behavioral responses, one might conclude that our participants were equally likely to trust a reporter who shared common interests with them and a reporter who was made trustworthy by an institution, namely a penalty for lying. In contrast, ERPs to the reporter's report in the common interests condition tended to differ from both the conflicting interests condition (as predicted) and the penalty for lying condition (contrary to our expectations). Given the relatively small sample size in the ERP experiment, the significant differences we observed between the common interests condition and the other two conditions are quite remarkable. Indeed, the small sample size increases the risk of a type II error (that is, failing to observe a difference between two conditions when there is in fact a difference between them).

The fact that reports in the common interests condition elicited a larger P3 component than in the penalty for lying condition may reflect the fact that participants perceived reports from the common interest reporter to be slightly more informative than those in the penalty for lying

[5] Our failure to observe statistically significant P3 amplitude differences between the penalty for lying and the conflicting interests conditions ($P = 0.09$) likely reflects power limitations of our relatively small sample size ($N = 12$). This was not the case, however, for the more robust differences between the common interests condition and each of the other two conditions.

condition. Taken together, these results indicate that participants' brains differentially processed information in the common interests condition, relative to the other two conditions. More broadly, these results may reflect different attentional processes resulting from the target in the three conditions.

As for the implications of these results, they indicate that even when socially transmitted information induces similar behavior, it may be processed differently depending on the manner in which the source is made trustworthy. Specifically, even though the reporter was, theoretically and behaviorally, equally trustworthy in the common interests and penalty for lying conditions, participants processed information differently when it came from a reporter who was trustworthy by virtue of sharing common interests with them vs. a reporter who was made trustworthy by an external institution. In this way, our results suggest that even though institutions substitute for common interests behaviorally, they are not necessarily cognitive substitutes for common interests. Of course, the question of whether and when the cognitive differences that we observed lead to changes in citizens' propensity to trust others is an empirical question that should be explored in future research.

REFERENCES

Andreoni, J. (1988). Why free ride? Strategies and learning in public goods experiments. *Journal of Public Economics*, *37*, 291–304.

Boudreau, C. (2006). Jurors are competent cue-takers: how institutions substitute for legal sophistication. *International Journal of Law in Context*, *2*(3), 293–304.

Clark, V.P., Hillyard, S.A. (1996). Spatial selective attention affects early extrastriate but not striate components of the visual evoked potentials. *Journal of Cognitive Neuroscience*, *8*, 387–402.

Crawford, V., Sobel, J. (1982). Strategic information transmission. *Econometrica*, *50*, 1431–51.

De Quervain, D.J.F., Fischbacher, U., Treyer, V., et al. (2004). The neural basis of altruistic punishment. *Science*, *305*, 1254–9.

Delgado, M.R., Frank, R.H., Phelps, E.A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nature Neuroscience*, *8*, 1611–8.

Donchin, E. (1981). Surprise! . . . Surprise? *Psychophysiology*, *18*, 493–513.

Donchin, E., Coles, M. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, *11*, 357–427.

Grady, C.L., Haxby, J.V., Horwitz, B., et al. (1992). Dissociation of object and spatial vision in human extrastriate cortex: age-related changes in activation of regional cerebral blood flow measured with [15O] water and positron emission tomography. *Journal of Cognitive Neuroscience*, *4*, 23–34.

Hillyard, S.A., Anllo-Vento, L. (1998). Event-related brain potentials in the study of visual selective attention. *Proceedings of the National Academy of Sciences USA*, *95*, 781–7.

Hillyard, S.A., Muente, T. (1984). Selective attention to color and location: an analysis with event-related brain potentials. *Perception & Psychophysics*, *36*, 185–98.

Huynh, H., Feldt, L.S. (1978). Estimation of the box correction for degrees of freedom from sample data in the randomized block and split plot designs. *Journal of Educational Statistics*, *1*, 69–82.

Izuma, K., Saito, D.N., Sadato, N. (2008). Processing of social and monetary rewards in the human striatum. *Neuron*, *58*, 284–294.

Johnson, R., Jr. (1986). A triarchic model of P300 amplitude. *Psychophysiology*, *23*, 211–24.

Johnson, R. (1988). The amplitude of the P300 component of the event-related potential: review and synthesis. In: Ackles, P., Jennings, J.R., Coles, M.G.H., editors. *Advances in Psychophysiology: A Research Annual, Vol. 3*. Greenwich, CT: JA1 Press, Inc., pp. 69–137.

Kenemans, J., Kok, A., Smulders, F. (1993). Event-related potentials to conjunctions of spatial frequency and orientation as a function of stimulus parameters and response requirements. *Electroencephalography and Clinical Neurophysiology*, *88*, 51–63.

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

Knutson, B., Bossaerts, P. (2007). Neural antecedents of financial decisions. *Journal of Neuroscience*, *27*, 8174–7.

Kosfeld, M., Heinrichs, M., Zak, P.J., Fischbacher, U., Fehr, E. (2005). Oxytocin increases trust in humans. *Nature*, *435*, 673–6.

Kranczioch, C, Debner, S., Engel, A. (2003). Event-related potential of the attentional blink phenomenon. *Cognitive Brain Research*, *17*, 177–187.

Luck, S.J. (1995). Multiple mechanisms of visual-spatial attention: recent evidence from human electrophysiology. *Behavioral Brain Research*, *71*, 113–123.

Luck, S.J., Hillyard, S.A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology*, *31*(3), 291–308.

Lupia, A. (1992). Busy voters, agenda control, and the power of information. *American Political Science Review*, *86*, 390–404.

Lupia, A. (1994). Shortcuts versus encyclopedias: information and voting behavior in California insurance reform elections. *American Political Science Review*, *88*, 63–76.

Lupia, A., McCubbins, M.D. (1998). *The Democratic Dilemma: Can Citizens Learn What They Need to Know?* Cambridge: Cambridge University Press, pp. 1–229.

McCabe, K., Houser, D., Ryan, L., Smith, V., Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the National Academy of Sciences USA*, *98*, 11832–11835.

Milinski, M., Semmann, D., Krambeck, H. J. (2002). Reputation helps solve the "tragedy of the commons." *Nature*, *415*, 424–6.

Mondak, J.J. (1993). Source cues and policy approval: The cognitive dynamics of public support for the Reagan agenda. *American Journal of Political Science*, *37*, 186–212.

Nuwer, M., Comi, G., Emerson, R., et al. (1998). IFCN standards for digital recording of clinical EEG. *Electroencephalography and Clinical Neurophysiology*, *106*, 259–61.

Palfrey, T.R., Prisbrey, J.E. (1996). Altruism, reputation, and noise in linear public goods experiments. *Journal of Public Economics*, *61*, 409–427.

Potts, G.F. (2004). An ERP index of task relevance evaluation of visual stimuli. *Brain & Cognition*, *56*, 5–13.

Potts, G.F., Martin, L.E., Burton, P., Montague, P.R. (2006). When things are better or worse than expected: the medial frontal cortex and the allocation of processing resources. *Journal of Cognitive Neuroscience*, *18*, 1112–1119.

Potts, G.F., Patel, S.H., Azzam, P.N. (2004). Impact of instructed relevance on the visual ERP. *International Journal of Psychophysiology*, *52*, 197–209.

Reuter-Lorenz, P.A., Jonides, J., Smith, E.E., et al. (2000). Age differences in the frontal lateralization of verbal and spatial working memory revealed by PET. *Journal of Cognitive Neuroscience*, *12*, 174–87.

Rilling, J.K., Gutman, D.A., Zeh, T.R., Pagnoni, G., Berns, G.S., Kilts, C.D. (2002). A neural basis for social cooperation. *Neuron*, *35*, 395–405.

Rilling, J.K., Sanfey, A.G., Aronson, J.A., Nystrom, L.E., Cohen, J.D. (2004). Opposing BOLD responses to reciprocating and unreciprocated altruism in putative reward pathways. *Neuroreport*, *15*, 2539–2543.

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*, 1755–8.

Sniderman, P.M., Brody, R.A., Tetlock, P.E. (1991). *Reasoning and Choice: Explorations in Political Psychology*. New York: Cambridge University Press, pp. 1–272.

Spantekow, A., Krappmann, P., Everling, S., Flohr, H. (1999). Event-related potentials and saccadic reaction times: Effects of fixation point offset or change. *Experimental Brain Research*, *127*, 291–297.

Vogel, E.K., Luck, S.J. (2000). The visual N1 component as an index of a discrimination process. *Psychophysiology*, *37*, 190–203.

Wickens, J.R., Horvitz, J.C., Costa, R.M., Killcross, S. (2007). Dopaminergic mechanisms in actions and habits. *Journal of Neuroscience*, *27*, 8181–3.

Zak, P.J., Kurzban, R., Matzner, W.T. (2004). The neurobiology of trust. *Annals of the New York Academy of Science*, *1032*, 224–7.