frontiers in
NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 18 April 2012
doi: 10.3389/fnins.2012.00058

What if I get busted? Deception, choice, and decision-making in social interaction

Kamila E. Sip[1,2,3]*, Joshua C. Skewes[2], Jennifer L. Marchant[4,5], William B. McGregor[3], Andreas Roepstorff[2,6] and Christopher D. Frith[2,4]

[1] *Department of Psychology, Rutgers University, Newark, New Jersey, USA*
[2] *Center for Functionally Integrative Neuroscience, Aarhus University Hospital, Aarhus, Denmark*
[3] *Department of Aesthetics and Communication – Linguistics, University of Aarhus, Aarhus, Denmark*
[4] *Wellcome Trust Centre for Neuroimaging, University College London, London, UK*
[5] *Institute of Cognitive Neuroscience, University College London, London, UK*
[6] *Section for Anthropology and Ethnography, Department of Culture and Society, University of Aarhus, Aarhus, Denmark*

Edited by:
Gabriel José Corrêa Mograbi, Federal University of Mato Grosso, Brazil

Reviewed by:
Nobuhito Abe, Harvard University, USA
Julian Keenan, Montclair State University, USA
Daniel C. Mograbi, King's College London, UK

****Correspondence:***
Kamila E. Sip, Social and Affective Neuroscience Lab, Department of Psychology, Rutgers University, Smith Hall, Room 301, 101 Warren Street, Newark, NJ 07102, USA.
e-mail: ksip@psychology.rutgers.edu

Deception is an essentially social act, yet little is known about how social consequences affect the decision to deceive. In this study, participants played a computerized game of deception without constraints on whether or when to attempt to deceive their opponent. Participants were questioned by an opponent outside the scanner about their knowledge of the content of a display. Importantly, questions were posed so that, in some conditions, it was possible to be deceptive, while in other conditions it was not. To simulate a realistic interaction, participants could be confronted about their claims by the opponent. This design, therefore, creates a context in which a deceptive participant runs the risk of being punished if their deception is detected. Our results show that participants were slower to give honest than to give deceptive responses when they knew more about the display and could use this knowledge for their own benefit. The condition in which confrontation was not possible was associated with increased activity in subgenual anterior cingulate cortex. The processing of a question which allows a deceptive response was associated with activation in right caudate and inferior frontal gyrus. Our findings suggest the decision to deceive is affected by the potential risk of social confrontation rather than the claim itself.

Keywords: deception, confrontation, social interaction, decision-making

INTRODUCTION

Deception has been of interest to psychologists, forensic experts, and laymen (Woodruff and Premack, 1979; Whiten and Byrne, 1988; Saarni and Lewis, 1993; Bradley et al., 1996; Walters, 2000). It has triggered trans-disciplinary scientific investigations within anthropology; philosophy; cognitive, social, and forensic psychology; and recently, cognitive neuroscience. Among the reasons for studying deception, determining the motivation for deceptive behavior, and enhancing recognition of deceptive strategies appear to be of core interest. For deception to be successful, it needs to have some foundation in truth, such that people tend not to deceive with a cluster of deceptive messages, but instead incorporate deception while telling the truth (see e.g., Ekman, 1992; DePaulo et al., 1996; DePaulo and Kashy, 1998). Therefore, deception may be interwoven into a partially honest message, to secure the trust of interlocutors.

Complex social interaction typically requires the ability to make rapid decisions that take account of possible outcomes. This involves a broad set of cognitive processes, including the ability (i) to determine the possible courses of action and to identify how they could be coordinated with the interlocutor, (ii) to weigh these available courses of action against one another, and (iii) to choose which action to perform next in the interaction.

Deception is an example of a complex social interaction and thus involves the same set of cognitive processes (Sip et al., 2008) but has the goal to instill a false belief in the mind of the interlocutor so as to manipulate how the interaction unfolds. To deceive, therefore, consciously and/or subconsciously we must be able (i) to determine whether deception is one of the set of possible actions in the interaction, (ii) to weigh the advantage to be gained by deceiving against the risks and consequences of being detected, and (iii) to choose to perform the deceptive action. As argued by Sip et al. (2008) these key cognitive components of social decision-making, and not the telling of a falsehood as such, provide the main explanatory content for the neural activity associated with the production of deception. Here, we aim to explore decision-making in deception in terms of the costs and values of our day-by-day contexts, while providing a free choice within the limitations of decision-making in laboratory settings.

In deceptive encounters, the change in circumstances is connected not only to the decision *per se*, but also to the impact resulting from an attempt to modulate the perspectives and beliefs of others. Therefore, like all choices – especially in social interactions – deception is influenced by probable gains and losses. Usually, we choose to deceive because we believe that if our deception is successful, we shall be better off than if we had told the truth.

There are many variables to consider in making such a choice. Will our deception be detected? What are the consequences of detection? Will we gain something if we are falsely accused of telling a falsehood (see Sip et al., 2010)? Deception is not just a simple matter of truth and falsehood. The gains from deception can be large, but the actual calculation of relative gains and losses involves solving a complicated decision-making tree, which can, at best, only be approximated. In real-life, the cost of being caught red-handed can be enormous, in terms of loss of reputation, trust, power, or money. Consequently, the danger of being confronted with one's deceptive claims may share similarities with experiencing negative social consequences, such as rejection (Masten et al., 2009; Onoda et al., 2009).

There has been a significant lack of imaging literature that treats deception as a social phenomenon. Only recently, neuroimaging investigations started treating deception within a framework of social decision-making (see e.g., Abe et al., 2007; Barrios et al., 2008; Baumgartner et al., 2009; Greene and Paxton, 2009; Bhatt et al., 2010; Carrion et al., 2010; Sip et al., 2010). Abe and colleagues addressed the issue of instructed lies by introducing a clever twist in their instructions to participants (Abe et al., 2007). Using a temporary absence of experimenter 1, experimenter 2 secretly instructed participants to deceive experimenter 1 by providing opposite responses than those suggested by the experimenter 1. Interestingly, in this study, participants faced an externally introduced change to the set of rules, and therefore it might be problematic to account for that change as a result of both peripheral attentional load and deception activation that could have contributed to the final results. Bhatt et al. (2010) investigated the role of social image in strategic deception to manipulate others' beliefs about each other for gains in a bargaining game. Another study tested how participants would behave when faced with a possibility of being deceptive to gain monetary rewards (dishonest gain; Greene and Paxton, 2009).

Many earlier studies (see e.g., Ganis et al., 2003; Spence et al., 2004; Langleben et al., 2005) have tested the production of deception by instructing participants when to tell a falsehood. In this way, the truth or falsity of participants' claims have been treated as an *independent* variable in most experimental paradigms, such that in most experiments, whether a claim is true or false has been under the control of the experimenter and not the participant. This approach excludes social decision-making from the experimental equation (see Sip et al., 2008 and also Greely and Illes, 2007). Therefore, the purpose of the present study is to take an alternative approach that focuses more on the social decision-making processes involved in deception, rather than on deception as a "yes" or "no" response equated with an honest or deceptive response respectively. We were primarily interested in investigating how participants produced deception given a free choice to make deceptive claims when detection was a possible social consequence. Therefore, rather than treating deception as an *independent variable* coded in a balanced factorial design, we instead controlled the social context for deception by systematically varying both the possibility to deceive and the possibility of being detected. Then, within this context, we left participants free to decide when and if they should attempt to make deceptive claims. We thus treated the responses associated with the decision to deceive as a *modulatory variable*.

A novel design was implemented in an attempt to accommodate for free choice and potential confrontation. In a paradigm modified from a behavioral study of Keysar et al. (2000), participants were questioned by an interlocutor about their knowledge of the content of a display, and the interlocutor could sometimes challenge their responses. Rather than being instructed to deceive the interlocutor, questions were posed to participants so that deception was meaningful in some conditions and not in others, and so that any acts of deception could be detected in some conditions and not in others. Within this design, participants were left to choose for themselves when to deceive, and with that choice followed the possible consequence of being caught out in a lie. This allowed us to treat deception as an outcome of a social decision-making process, and, in our data analysis, to regress the decision to deceive with neural and behavioral measures. Given that deception is a social decision-making process, and that the anterior cingulate cortex (ACC) is involved in decision-making (see e.g., Botvinick, 2007; Dolan, 2007; Rushworth and Behrens, 2008; Croxson et al., 2009), we expected ACC to be active in conditions where it was necessary to balance a monetary reward for successfully deceiving the interlocutor against the risk of detection (e.g., Abe et al., 2006; Baumgartner et al., 2009).

Participants played both against (what they believed were) a human and a computer. This double partnership was motivated by previous social studies that showed that participants care whether their opponent is a human and attribute different behavior accordingly (see e.g., Gallagher et al., 2002). This aspect has not yet been tested in deception paradigms.

It bears clarifying that the primary aim of our study was not to observe how behavior and neural activity of individuals were affected by the *performance* of deception *per se*. Rather, the primary aim of our study was to investigate how individuals' decision to deceive modulates their behavior and neural activity given the social and informational context in which that decision is made. Our focus was therefore not on the production of deception as an act in and of itself, but rather on the social decision-making processes associated with the production of deception. This is why the participants' decision to deceive was treated as a free modulatory parameter in this study, and not as part of the study's factorial design. In this way, our study breaks with standard practice in the design of deception experiments for the purpose of addressing an important unresolved issue.

MATERIALS AND METHODS

SUBJECTS

Sixteen healthy, right-handed participants with no reported neurological or psychiatric disorders responded to an ad to volunteer in the experiment. Data from two participants were excluded. One told a falsehood at all times regardless of the context, while there were excessive movement artifacts in the fMRI data for the other. The remaining 14 participants (7 males) were aged between 20 and 45 years (mean = 26; SD = 6.9). Participants gave written informed consent to take part in the study, conducted according to the principles expressed in the Declaration of Helsinki, which was approved by the Joint Ethics Committee of the National Hospital for Neurology and Neuroscience (UCL NHS Trust) and Institute of Neurology (UCL).

STIMULI

Participants were presented with a two-dimensional representation of a three-dimensional box. The box was divided into 16 compartments (4 × 4 grid) or shelves (**Figure 1**). On each trial, each compartment could be empty or contain one of seven different objects. Each compartment was always represented as open to the front, but could be either open or closed to the back. From the front view, it was obvious if a particular object could also be seen from the back.

PROCEDURE

While in the scanner, participants were shown the front view of the stimulus, and were told an interlocutor was simultaneously being shown the back view. On each trial (see **Figure 2**), the interlocutor asked participants if they could see a target object on any of the shelves. The target object was randomized across trials. There was no restriction on whether the response should be true or false. Participants heard the questions via headphones and responded yes or no by button press.

The opponent could ask three types of question (A, B, and C). For *Question type A,* the target object was visible from the front and the back views, so that it was obvious to the participant that the interlocutor could easily detect deception (symmetrical knowledge; truth_eliciting question). For *Question type B,* the target object was only visible from the front view, so that it was obvious to the participant that it should be more difficult for the interlocutor to detect deception (asymmetrical knowledge, deception by omission; falsehood_eliciting question). For *Question type C,* the target object was not present in the box, so that it was more difficult for the interlocutor to detect deception, but this was not immediately obvious to the participant because it required visual search (asymmetrical knowledge, deception by commission; falsehood_eliciting question).

The experiment consisted of two sessions with different types of interlocutor (human or computer). Each session consisted of six blocks. In two blocks participants were informed that a computer program posed the questions and a computer-generated voice was used. In another two blocks participants were informed that the questions were posed online by the experimenter (K. Sip), whose voice they had heard, and with whom the participants had interacted with prior to the functional scans. In the two remaining blocks, participants were instructed to always state whether an object was present (answer truthfully with no motivation to deceive). These blocks were only used to check whether participants understood the task, and they were not used in the fMRI analysis. Unknown to the participants, the experimenter's voice was pre-recorded and the questions were posed in a predetermined order.

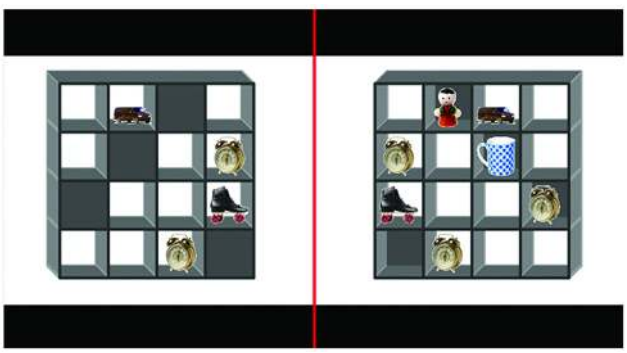

FIGURE 1 | An example of the stimulus display. The display represented the 16 compartment box with a typical set of objects used in the study. The participants were asked several different types of questions regarding the contents of the box, e.g., Question type A "Do you see a roller-skate?" (Truth_eliciting question), Question type B "Do you see a doll?" (Falsehood_eliciting question), Question type C "Do you see a giraffe?" (Falsehood_eliciting question).

In each of these situations, the interlocutor could confront participants about their responses in one block but not in the other. Although participants always knew which block they were in, they did not know which responses would be confronted. They were informed prior to the start of the confrontation block that the interlocutor was allowed to confront only some of their responses, usually up to four responses per block.

Each experimental trial could be rewarded or punished with a small amount (50 pence per event). Participants were informed that they would be rewarded for successful deception and penalized for unsuccessful attempts across all conditions. There was no monetary consequence for telling the truth when the object was visible for both players. The system of rewards was introduced to further motivate participants to try to avoid detection. Importantly, no monetary feedback was given to the participants during the functional scans at any point. Therefore, participants were not able to track their rewards on a trial to trial basis, instead allowing them to give priority to the decision about whether to be honest or not. This was important to ensure that participants were attentive in all conditions and refrained from giving only one type of response, e.g., always replying "yes" when confrontation was not possible. The total rewards were calculated at the end of experiment.

The same reward pattern was used for unchecked trials in the confrontation blocks. However, in the few predetermined checked trials (four per block), participants were penalized if they were caught telling a falsehood, and were compensated for being wrongly accused of telling a falsehood when they made a truthful response.

Question trials were randomized within the blocks. Block and session order were counterbalanced using a 2 × 2 Latin Square. After the experiment was completed, the participants were debriefed, which revealed that all believed they had interacted with a human during the human sessions, and that all had actively tried to deceive her.

ANALYSIS AND DESIGN

A three-way factorial design was used with question type (3) × confrontation (2) × interlocutor (2) as factors, with response type included as a covariate and response time as a dependent variable. In data analysis, participants' decision to answer truthfully or to try to deceive the interlocutor was added as a modulator [as a covariate for the response times and a parametric modulation for the blood oxygenation level-dependent (BOLD) signal]. This allowed us to determine the influence of participants'

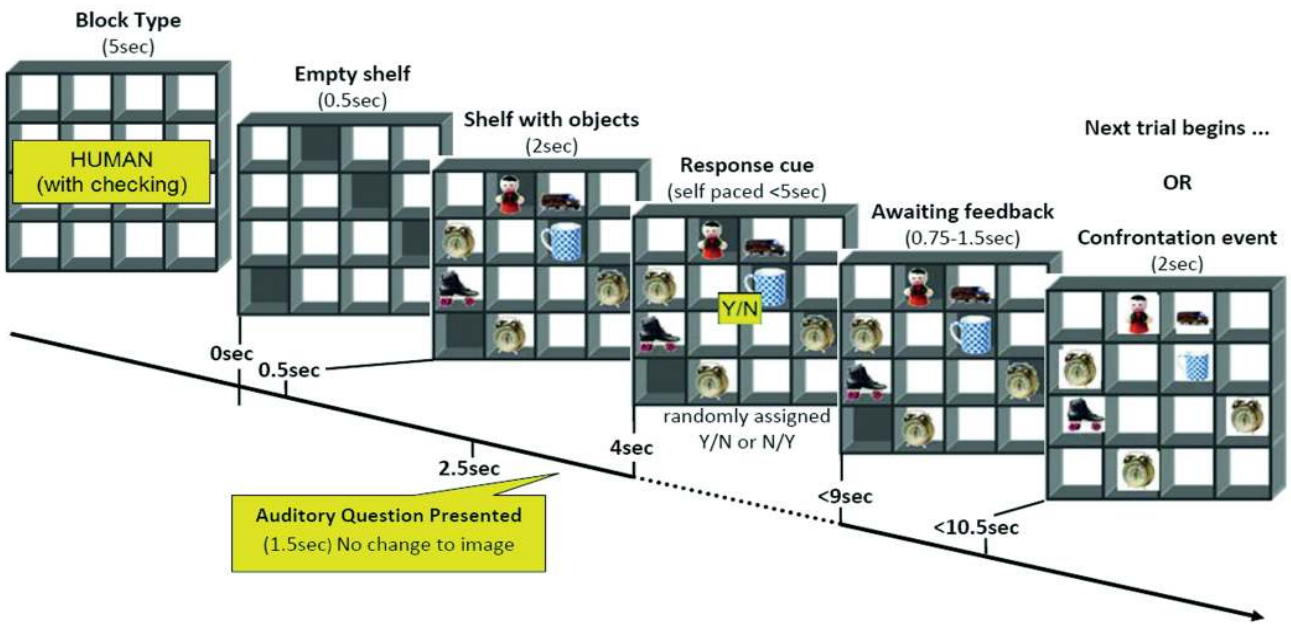

FIGURE 2 | An example of the events per trial in a confrontation block. This figure presents the sequence and duration of events in a trial in blocks where the participants could be confronted with their claims. Participants responded on average 2.2 ± 0.23 s after the question was asked and the response cue was removed from the screen. In trials where verbal feedback was given to the participants after confronting their choices, the feedback lasted 1.7 ± 0.27 s.

active social decision-making on their behavior and neural activity when performing deception.

The approach to include participants' decision to deceive as a modulatory variable deviates from the usual approaches of treating variables of interest as controlled experimental factors to be analyzed with analysis of variance. However, our choice is justified, both in principle and empirically, from the perspective of our experimental design. The truth or falsity of participants' responses were not experimentally controlled, but intentionally left under participant control, so that the choice to deceive was not an independent variable in our study. In principle, therefore, the choice to deceive is not a valid target for inclusion as a separate factor in our analysis. Moreover, because participants were free to decide when they should make deceptive claims, they attempted to deceive more often in some conditions than in others. Empirically, therefore, participants' decision to deceive is not sufficiently balanced across conditions, so that treating this variable as a factor would violate one of the core assumptions of analysis of variance. It should also be recalled in this context that our reason for designing the study in this way was that we were not interested in deception in itself as an isolated speech act, but in the social decision-making processes involved in deception. Participants' free decision to deceive was thus conceived in our experimental design as a modulatory variable, and is analyzed as such.

fMRI SCANNING PARAMETERS

A 1.5T Siemens Sonata MRI scanner (Siemens, Erlangen, Germany) was used to acquire T1-weighted anatomical images and T2∗-weighted echo-planar functional images with blood oxygenation level-dependent (BOLD) contrast (35 axial slices, 2 mm slice thickness with 1 mm gap, 3×3 resolution in plane, slice TE = 50 ms, volume TR = 3.15 s, 64×64 matrix, 192×192 mm FOV, 90° flip angle). Two functional EPI sessions of up to 345 on average whole brain volumes (range 300–364 depending on participants response speed) were acquired and the first four volumes were discarded to allow for T1 equilibrium effects.

Image processing was carried out using SPM5 (Statistical Parametric Mapping software, Wellcome Trust Centre for Neuroimaging, UCL)[1] implemented in MATLAB (The Mathworks Inc., Massachusetts)[2]. EPI images were realigned and unwarped to correct for movements, slice time corrected, spatially normalized to standard space using the Montreal Neurological Institute EPI template (voxel size of 2 mm × 2 mm × 2 mm) and spatially smoothed with a 8 mm full-width half maximum Gaussian kernel.

IMAGING DATA ANALYSIS

All events were modeled using the standard hemodynamic response function of SPM5. The design matrix comprised a column for each experimental condition, with separate events defined by their onset time and duration (based on participants' response times). In keeping with our statistical approach of treating the participants' decision to deceive as a modulatory variable, participants' truthful, and deceptive responses in each condition were

[1] www.fil.ion.ucl.ac.uk/spm
[2] www.mathworks.com

added as separate parametric modulations of each column of the design matrix. The fit to the data was estimated for each participant using a general linear model (Friston et al., 1995) with a 128 s high-pass filter, global scaling, and modeling of serial autocorrelations.

Individual T-contrasts related to the different conditions within our factorial design were created from the parameter estimates (beta weights). T-contrasts were computed within subjects for the main effect of confrontation and the main effect of partner, for the effects of question types A, B, and C, and for the relevant interactions. These were then used in separate second level random effects analyses in order to facilitate inferences about group effects (Friston et al., 1995).

Unless specified otherwise, whole brain results are reported for clusters with at least 10 voxels and a threshold of $p < 0.005$ uncorrected for multiple comparisons, the most commonly reported threshold for social neuroimaging studies (Wager et al., 2007). This threshold allows for an appropriate balance between Type I and Type II errors especially in complicated designs involving socio-cognitive decision-making (see e.g., Lieberman and Cunningham, 2009). Additionally, we indicate several areas which survive a more stringent FWE correction for multiple comparisons.

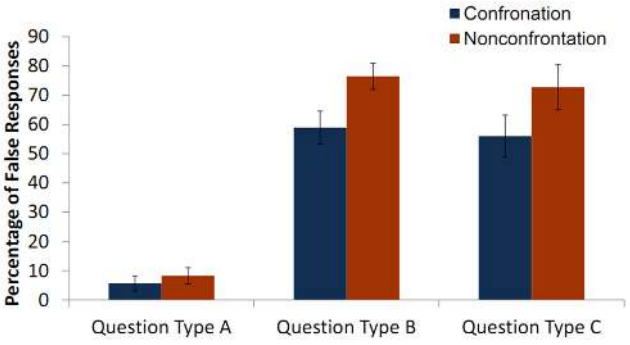

FIGURE 3 | Mean percentage of false claims across conditions. For illustration purposes, this graph shows the mean percentage of false claims across question type and confrontation. In the confrontation condition participants gave 58.95% (SE = 5.63) false responses to Question Type B (the target object was only visible from the front view), 56.04% (SE = 7.15) false responses to Question Type C (the target object was not present in the box), and 8.3% (SE = 2.76) false responses to Question Type A (the target object was visible from the front and the back views). In the non-confrontation condition they gave 76.45% (SE = 4.49) false responses to Question Type B, 72.74% (SE = 7.62) false responses to Question Type C, and 5.6% (SE = 2.61) false responses to Question Type A.

RESULTS

BEHAVIORAL RESULTS

A 2 (partner) × 2 (possibility of being confronted) × 3 (type of question) repeated measures ANOVA revealed significant main effects of confrontation [$F(1,13) = 16.23$, $p = 0.001$] and question type [$F(2,26) = 61.72$, $p < 0.001$] on producing false responses. The main effect of partner was not significant [$F(1,13) = 1.49$, $p = 0.24$]. The test revealed a significant interaction between confrontation and question type on the percentage of false claims [$F(2,26) = 3.65$, $p = 0.04$] There were fewer false responses in the confrontation condition, but this was only the case for the falsehood_eliciting question types (see **Figure 3**). There was no significant interaction between partner and question type [$F(2,26) = 1.56$, $p = 0.23$] and partner and confrontation [$F(1,13) = 0.11$, $p = 0.75$] on producing false responses. The three-way interaction was not significant [$F(2,26) = 0.024$, $p = 0.97$].

When the decision to deceive was added as a covariate, a 2 (type of interlocutor) × 2 (possibility of being confronted by the interlocutor) × 3 (type of question asked) repeated measures ANCOVA on response time revealed a significant main effect of question type [$F(2,12) = 13.26$, $p = 0.001$], and a significant interaction between the question type factor and the response type covariate [$F(2,12) = 4.98$, $p = 0.03$]. A marginally significant interaction between confrontation and question type [$F(2,12) = 3.84$, $p = 0.05$] was also revealed.

Figure 4 (see **Figure 4**) shows that (i) when participants and interlocutors had the same knowledge about the presence of an object in the box, participants were faster to give a true response, regardless of the possibility of confrontation; (ii) when there was obviously asymmetric knowledge between participants and the interlocutor, participants were slower to give a true response, but only when there was no possibility of being confronted; and (iii) when participants knew more about the stimulus but greater attention was required to take advantage of this knowledge, they were slower to give a true than a false response, regardless of the possibility of being confronted. These effects were not significant, however, if the covariate coding participants' decision to respond truthfully or falsely on each trial was removed from the analysis.

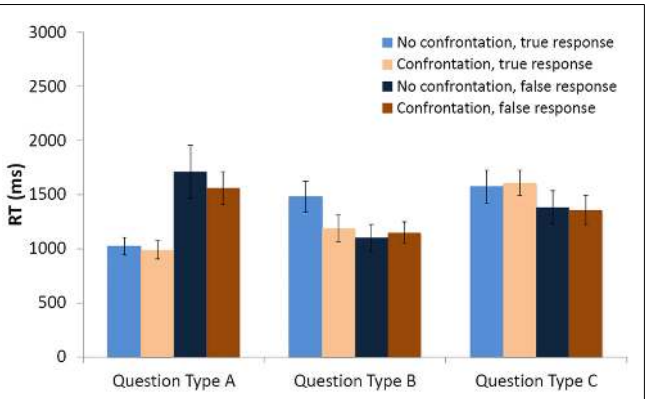

FIGURE 4 | Mean response times (RT) to answer the opponent's question. Separate means are given for false and true responses, and for responses given both when the opponent could and could not confront the response. Error bars represent one SEM.

NEUROIMAGING RESULTS

When the decision to deceive was added as a parametric modulator, the main effect of confrontation showed increased activity in subgenual anterior cingulate cortex (subACC) when participants' responses could not be confronted (**Figure 5**; see **Table 1**).

There was also a significant main effect of question type. For question type B, we observe increased activation in right caudate and inferior frontal gyrus (IFG; **Figure 6**). For question type A, we

observed increased activity in right putamen, superior temporal gyrus (auditory cortex), and occipital cortex.

DISCUSSION

The current investigation allowed participants the choice to deceive by creating a context in which deception was sometimes possible, but ran into the risk of being punished if it was detected. Our paradigm captures the idea that when people attempt to deceive others, they face a demanding task, based on balancing the tensions between choice and potential outcomes. The paradigm allowed us to treat deception as the outcome of social decision-making, and in our data analysis, to regress the choices participants made with the neural and behavioral measures taken.

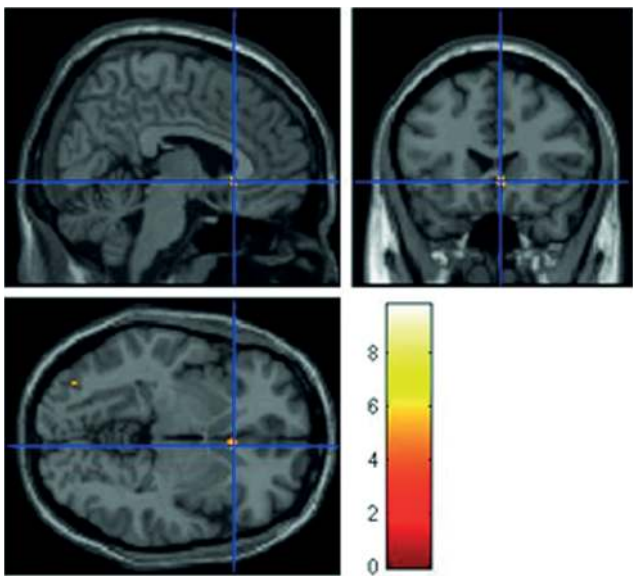

FIGURE 5 | Main effect of possibility of confrontation on response type. The peak activation is localized in subgenual ACC (6 22 −4, $p < 0.023$, FEW corrected). The color-bar corresponds to *T*-values.

Our results suggest that social feedback can only be seen to mediate responses to the question being asked if we take seriously the variance introduced by the free choice the participants are given.

Although this is not the first study to explore deception in social interaction (see Baumgartner et al., 2009; Sip et al., 2010), it is one of the first to provide a context in which participants run the risk of being socially confronted in case their deception is detected (see also Baumgartner et al., 2009; Sip et al., 2010). Participants were allowed to decide whether or not to deceive the partner on any given trial. We found activation in subgenual ACC when the partner could not check the truthfulness of the participants' response. Activation in right caudate and IFG was observed when participants were deciding how to respond to a question that allowed deception. Surprisingly, there were neither behavioral nor neural effects of partner (human vs. computer). This is surprising because one would expect that (1) participants would consider a computer of less importance and thus exhibit a very different pattern of behavior in contrast to that toward human; and (2) participants would try to attribute intentions and causality of actions to people, but not to computers (see e.g., Gallagher et al., 2002). We speculate that the lack of partner effect results from the paradigm placing the main focus on confrontation. Even though participants played with a computer, the machine still exposes their deception to the people observing the task outside the scanner.

The activations in right caudate and IFG strongly suggest that when participants are in the position to make a false claim, presumably they have to decide whether or not to do so given the ratio between the effort invested in the action and its potential rewards. The right IFG has been typically associated with response inhibition tasks in which participants typically need to inhibit their natural response (e.g., Aron et al., 2004). Interestingly, this area has also been implicated in risk aversion, and is suggested to play a role in inhibition of accepting a risky option (Christopoulos et al., 2009). Additionally, the area BA47 (see **Table 1**) has also been implicated in comprehending spoken language (Petrides and Pandya, 2002), which suggests that participants in the current study had to focus on what they were asked about before giving a response. The activation of caudate – well-known for processing

Table 1 | Brain regions showing activation in decision-making.

Brain region	Cluster size	*x*	*y*	*z*	*T*-value	*Z*-value
MAIN EFFECT OF CONFRONTATION (NON-CONFRONTATION > CONFRONTATION)						
Right subgenual ACC (BA25)*	16	6	22	−4	9.81	5.18
MAIN EFFECT OF QUESTION TYPE (FALSE ELICITING QUESTION > TRUTH ELICITING QUESTION)						
Right superior frontal gyrus (SMA, BA 6)	26	4	6	66	7.89	4.70
Right caudate	47	14	12	10	6.56	4.29
Right inferior frontal gyrus (IFG; BA 47)	36	42	20	−12	5.58	3.92
MAIN EFFECT OF QUESTION TYPE (TRUTH ELICITING QUESTION > FALSE ELICITING QUESTION)						
Right superior temporal gyrus (BA 22)*	35	48	−8	0	9.33	5.07
Right putamen*	54	22	−2	6	10.16	5.25
Left occipital lobe	44	−8	−72	5	4.93	3.54

The coordinates are given according to the MNI space, together with T-scores, Z-scores, and significant thresholds $p < 0.005$ uncorrected for multiple comparisons with a cluster extent threshold of 10 voxels, corrected at the cluster level. We indicate with an asterisk (*) the areas which survive more stringent threshold of FWE correction of $p < 0.05$ at the voxel level.

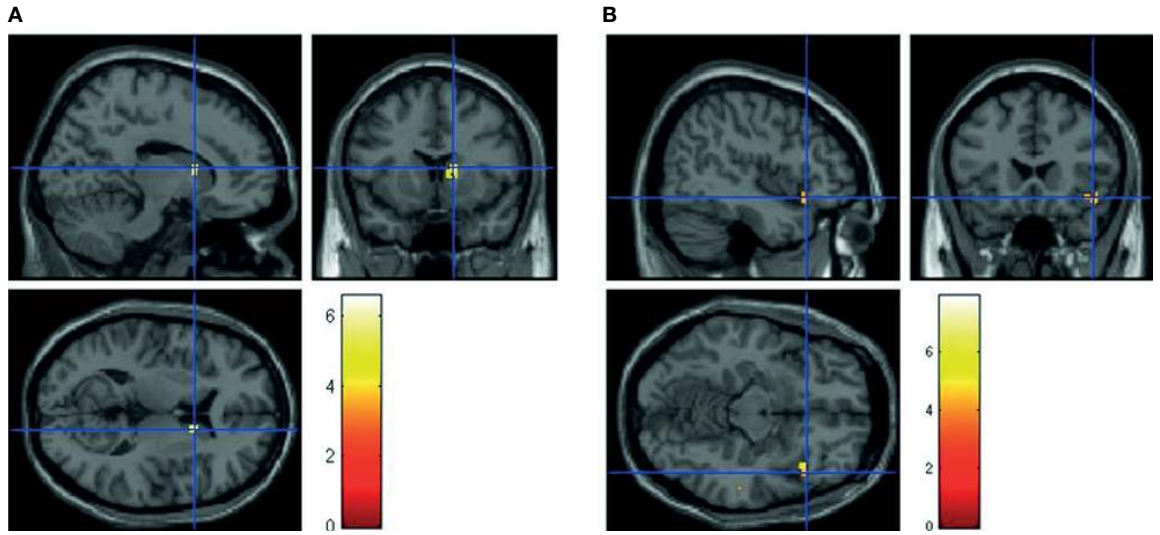

FIGURE 6 | Main effect of falsehood-eliciting question (Question Type B) on response type. The peak activations are in **(A)** the right caudate (14 12 10) and **(B)** right inferior frontal gyrus (IFG; 42 20–12), $p < 0.005$, uncorrected. The color-bar corresponds to *T*-values.

effort to engage in an action/choice selection (Croxson et al., 2009; Kurniawan et al., 2010) – and dorsal putamen – reported in prediction error, memory, and affective learning (Delgado, 2007) – suggests that the choice of making either a false or true claim may elicit the feeling of reward, reward anticipation, or the feeling of control when making a choice (Leotti et al., 2010). While giving a response, participants needed to also account for previous choices as well as indirectly learn from the interaction what would be their best strategy to exercise deception. Interestingly, activation of dorsal putamen and caudate nucleus may indicate that memory and learning facilitated the choice participants were faced with in our task.

Anterior cingulate cortex has been implicated in social–affective processes involved in decision-making (Dolan, 2007; Rushworth and Behrens, 2008; Croxson et al., 2009). ACC is believed to store associations between past behaviors and rewards (for reviews see Paus, 2001; Rudebeck et al., 2008) and to process choices in dynamic and open-ended contexts (Walton et al., 2007). It subserves response and cognitive conflict monitoring (Botvinick, 2007), calculates cost–benefit evaluations (Croxson et al., 2009), reward expectations (Delgado et al., 2005; Etkin et al., 2006) as well as action selection (for review see e.g., Rushworth et al., 2004; Rushworth et al., 2007). The dorsal and rostral portions of ACC have been associated with choice, conflict monitoring (Rushworth et al., 2004) and representations of beliefs and expectations (Petrovic et al., 2005). The more ventral part of ACC has been reported in processing the value of possible choices in relation to expected reward (Bush et al., 2000). Because of anatomical and functional connections with orbitofrontal cortex (OFC; for review see e.g., Paus, 2001) and ventral striatum (Balleine et al., 2007; Delgado, 2007), ACC functions are strongly modulated by social and emotional context (Rushworth et al., 2007; Rushworth and Behrens, 2008). Multiple ACC functions are therefore likely to be implicated in the decision to deceive (e.g., Ganis et al., 2003; Abe et al., 2006; Baumgartner et al., 2009).

Our finding that ACC is active in a task involving deception is not surprising. Surprisingly though, in other studies an increased activation in ACC has been reported in very different portions of this large area. Several groups reported the activation of dorsal ACC (BA 24/32; Ganis et al., 2003; Kozel et al., 2005; Langleben et al., 2005) in association with the production of deception. However, the tasks used in these experiments were quite different from the task used in the present study (for discussion see Greely and Illes, 2007; Sip et al., 2008; Christ et al., 2009), and the activations were located more dorsally. For example, Ganis et al. (2003) found activation in the dorsal ACC (BA32, 4 6 39; among other areas) by contrasting activity associated with the production of "spontaneous lies" that do not necessarily fit into a coherent story with the production of well-rehearsed falsehoods accommodated in a prepared story. Kozel et al. (2005) observed right ACC activation (ACC, 3 18 60) in a mock-crime experiment in which the subjects were asked to deny possession of a "stolen" object. This activation was associated with monitoring a deceptive response by inhibiting truth-telling. In another study, Abe et al. (2006) observed increased activation of right ACC (BA 24/32) when participants engaged in deception about past events. Only recently was ACC (BA 24) activation reported in an ecologically valid study (Baumgartner et al., 2009), where it was associated with breaking a previously expressed promise in a trust game.

Our observation that the subgenual ACC is active when the decision to deceive does not have immediate social consequences is, however, interesting. Subgenual ACC has previously been implicated in studies of social rejection (8 22 −4 and 10 20 −8 in Masten et al., 2009) and social pain (10 32 −10 in Onoda et al., 2009). Our imaging findings, supported by our behavioral results, therefore suggest that ACC subserves social monitoring when the decision to deceive does not depend upon possible confrontation. In the confrontation condition, the decision to deceive or not will be based largely on utilities, for example the value of deception, and

the likely hood of being detected. In the non-confrontation condition these considerations are irrelevant. Rather, the decision not to deceive, even when deception cannot be detected, would be based on moral considerations. To our knowledge, this role of subgenual ACC has not been implicated in other deception studies. Our results confirm our hypothesis (also expressed in Sip et al., 2008) that social feedback – and consequently a potential social rejection – affects production of deception. We speculate that subACC, caudate, and IFG play an important role in mediating a decision to deceive based on the context, rather than in producing false statements.

SOCIAL AND MORAL CONSIDERATION IN EXERCISING DECEPTION

For many of us, social rejection may also be based on moral values (Greene et al., 2001; Raine and Yang, 2006) and expectations. Thus deception is interestingly related to moral emotions, such as guilt and shame. However, a moral belief that we should not deceive others may be dismissed in contexts in which deception is allowed or even expected, as in most game scenarios and controlled experimental settings (Sip et al., 2010). This means that although there is an important relationship between deception and morality, when deception is sanctioned by the context, it is possible for people to perform genuine deception without experiencing any of the moral emotions one might expect to experience otherwise. Nevertheless, other social consequences of being detected must still be weighted accordingly when one is faced with the choice to deceive, even when moral concerns are made irrelevant to the decision.

We did not observe activation in an emotional network (e.g., insula or amygdala) as in another ecological study of deception (Baumgartner et al., 2009). The reason for this difference may be a difference in focus. Our participants did not declare (promise) to their interlocutor whether they would be honest or deceptive on specific trials. Therefore, the component of explicit social commitment is not involved in our study, such that we should not expect a similar emotional reaction as observed in Baumgartner's study (Baumgartner et al., 2009). This might be because the choice of whether to perform a morally sanctioned act of deception in a game and the more morally loaded choice of whether to break a promise, involve different social phenomena – rejection (van Beest and Williams, 2006) and guilt respectively. Nevertheless, it is challenging to evoke and accurately assess guilt associated with deception in real-life interrogations (Bashore and Rapp, 1993; Pollina et al., 2004), let alone in experimental settings.

Additionally, given that most neuroimaging studies of deception use a researcher as a recipient of deception (and this is known to the subjects), one may argue that this could weaken participants' attempts at deception. In our experiment, however, participants do not act against the experimenter, but rather act within the normative context of the experiment, which implies that the same behavior would not be processed differently toward a stranger. In other words, if participants believe they play with another human in the context of this experiment, this entails an oppositional behavior. Therefore, moral emotions are canceled out by the fact that immoral behavior is sanctioned by the context. Additionally, based on the post-scan debriefing, we are confident that participants tried their best to deceive the experimenter, where in many cases this was a matter of gaining an upper hand over somebody more experienced in the topic.

THE ROLE OF INSTRUCTIONS

In experimental settings, instructions given to the participants not only determine their behavior, but also frame how they think about others' actions, mental states, and expectations. In complicated studies of social decision-making, there is a discrepancy between what the instructions say, what the participants agree to do, and what they actually do while lying still in the MR chamber. This is specifically relevant to experimental tasks based on explicit forced-choice instructions, in which the execution of deception is often presumed to be intelligible independently of the choice and intention to instill a false belief in another person (Sip et al., 2008). These social cognitive processes, functioning in the context of the instructions, constrain the concrete task of executing deception, thus posing conceptual problems for interpreting results produced by any experimental design that does not incorporate them. Ideally, then, task instructions (1) must not define too specifically for the participants when to be deceptive or truthful, and (2) they should not overly limit the quantity and the quality of the choices made by the participants.

In human behavioral and psychological experiments more generally, the interaction between the experimenter and the participant involves sharing a specific script that is aimed to facilitate the execution of an experimental task (Roepstorff and Frith, 2004). In order words, the experimenter communicates the nature of the paradigm to the participant, who acts according to the instructions, or more precisely, to her own understanding of what they entail. In the ideal situation, it is then up to the subject to make the choice of whether or not to comply. However, if the instructions tell the participants to "lie" about events in one condition and to be honest about other events in another (Sip et al., 2008), then the executive role of the participant in choosing to act is essentially left out. Thus, an interesting aspect of deception, namely the social cognitive processes involved in the decision to deceive, are excluded unless participants are able to achieve a certain degree of freedom in response selection, which is not controlled by the experimenter.

Interestingly, in the current study, even though experimental instructions implicitly suggested telling a falsehood, participants did not tell a falsehood 100% of the time when deception was possible (**Figure 3**). This suggests that even when there was no direct danger of being caught in a lie in the non-confrontation condition, participants still mimic a real-life situation in this context, in which the ratio of true and false claims is not predetermined across contexts. Another interesting result was that there were several trials in which participants decided to tell a falsehood in response to questions in which the object was visible to both parties (**Figures 3** and **4**). Peculiar as it sounds; this suggests that mistakes aside, participants did exercise their free choice, even in a situation that was not beneficial to them. Additionally, **Figure 4** shows an interesting pattern of reaction times relative to the question type and response type. One possibility is that the slower RTs of true claims are concerned with less plausible responses that perhaps require more thought. For example, the somewhat irrational responses of telling a falsehood in response to question type A, and telling the

truth when deception cannot be detected in question type B, are similarly slowed.

LIMITATIONS

Because of our effort to account for a natural deceptive interaction in laboratory settings, this study faces certain limitations: (a) free choice in deceptive decision-making give rise to a range of behavior that is difficult to predict prior to the experiment, (b) unbalanced numbers of events that are then included in imaging analysis, (c) interpersonal differences that cause inter- and intra-subject variability in recorded data. Additionally, our study might be underpowered due to the small sample size to detect activations associated with moral emotions. Therefore, one may speculate alternative explanations for the lack of moral and emotional networks, such that it is plausible that the presence of moral emotions was merely diminished instead of canceled out. Further ecological studies are called for to allow better understanding of neural and behavioral processes that facilitate deceptive behavior.

Overall, our findings suggest that production of deception depends upon an effort-based affective–motivational network rather than merely higher-level cognitive processes as has been suggested thus far. Given that potential social consequences affect decisions to deceive, we argue that real-life deception may be interpreted as a decision with costs, benefits and losses. The gain from the deception must be evaluated as greater than the cost of the deception. Similarly, the gain made possible by the deception must be balanced against the cost of being found out. As in all such decisions, the costs are monitored according to what the other person knows and does not know, in relation to what the deceptive agents know. We suggest that the fields of neuroeconomics and deception intersect (see e.g., Baumgartner et al., 2009) and could offer an interesting contribution to further understanding of deception.

ACKNOWLEDGMENTS

We are grateful to Ian Apperly for his help with the development of the paradigm.

REFERENCES

Abe, N., Suzuki, M., Mori, E., Itoh, M., and Fujii, T. (2007). Deceiving others: distinct neural responses of the prefrontal cortex and amygdala in simple fabrication and deception with social interactions. *J. Cogn. Neurosci.* 19, 287–295.

Abe, N., Suzuki, M., Tsukiura, T., Mori, E., Yamaguchi, K., Itoh, M., and Fujii, T. (2006). Dissociable roles of prefrontal and anterior cingulate cortices in deception. *Cereb. Cortex* 16, 192–199.

Aron, A. R., Robbins, T. W., and Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends Cogn. Sci. (Regul. Ed.)* 8, 170–177.

Balleine, B. W., Delgado, M. R., and Hikosaka, O. (2007). The role of dorsal striatum in reward and decision-making. *J. Neurosci.* 27, 8159–8160.

Barrios, V., Kwan, V. S. Y., Ganis, G., Gorman, J., Romanowski, J., and Keenan, J. P. (2008). Elucidating the neural correlates of egoistic and moralistic self-enhancement. *Conscious. Cogn.* 17, 451–356.

Bashore, T. R., and Rapp, P. E. (1993). Are there alternatives to traditional polygraph procedures? *Psychol. Bull.* 113, 3–22.

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

Bhatt, M. A., Lohrenz, T., Camerer, C. F., and Montague, P. R. (2010). Neural signatures of strategic types in a two-person bargaining game. *Proc. Natl. Acad. Sci. U.S.A.* 107, 19720–19725.

Botvinick, M. M. (2007). Conflict monitoring and decision making: reconciling two perspectives on anterior cingulate function. *Cogn. Affect. Behav. Neurosci.* 7, 356–366.

Bradley, M. T., MacLaren, V. V., and Carle, S. B. (1996). Deception and nondeception in guilty knowledge and guilty actions polygraph tests. *J. Appl. Psychol.* 81, 153–160.

Bush, G., Luu, P., and Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends Cogn. Sci. (Regul. Ed.)* 4, 215–222.

Carrion, R. E., Keenan, J. P., and Sebanz, N. (2010). A truth that's told with bad intent: an ERP study of deception. *Cognition* 114, 105–110.

Christ, S. E., Van Essen, D. C., Watson, J. M., Brubaker, L. E., and McDermott, K. B. (2009). The contributions of prefrontal cortex and executive control of deception: evidence from activation likelihood estimate meta-analyses. *Cereb. Cortex* 19, 1557–1566.

Christopoulos, G. I., Tobler, P. N., Bossaerts, P., Dolan, R. J., and Schultz, W. (2009). Neural correlates of value, risk, and risk aversion contributing to decision making under risk. *J. Neurosci.* 26, 6469–6472

Croxson, P. L., Walton, M. E., O'Reilly, J. X., Behrens, T. E., and Rushworth, M. F. (2009). Effort-based cost-benefit valuation and the human brain. *J. Neurosci.* 29, 4531–4541.

Delgado, M. R. (2007). Reward-related responses in the human striatum. *Ann. N. Y. Acad. Sci.* 1104, 70–88.

Delgado, M. R., Frank, R. H., and Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nat. Neurosci.* 8, 1611–1618.

DePaulo, B. M., and Kashy, D. A. (1998). Everyday lies in close and casual relationships. *J. Pers. Soc. Psychol.* 74, 63–79.

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., and Epstein, J. A. (1996). Lying in everyday life. *J. Pers. Soc. Psychol.* 70, 979–995.

Dolan, R. J. (2007). The human amygdala and orbital prefrontal cortex in behavioural regulation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 362, 787–799.

Ekman, P. (1992). *Telling Lies: Clues to Deceit in the Marketplace, Politics and Marriage.* New York: W W Norton.

Etkin, A., Egner, T., Peraza, D. M., Kandel, E. R., and Hirsch, J. (2006). Resolving emotional conflict: a role for the rostral anterior cingulate cortex in modulating activity in the amygdala. *Neuron* 51, 871–882.

Friston, K. J., Holmes, A. P., Poline, J. B., Grasby, P. J., Williams, S. C., Frackowiak, R. S., and Turner, R. (1995). Analysis of fMRI time-series revisited. *Neuroimage* 2, 45–53.

Gallagher, H. L., Jack, A. I., Roepstorff, A., and Frith, C. D. (2002). Imaging the intentional stance in a competitive game. *Neuroimage* 16, 814–821.

Ganis, G., Kosslyn, S. M., Stose, S., Thompson, W. L., and Yurgelun-Todd, D. A. (2003). Neural correlates of different types of deception: an fMRI investigation. *Cereb. Cortex* 13, 830–836.

Greely, H. T., and Illes, J. (2007). Neuroscience-based lie detection: the urgent need for regulation. *Am. J. Law Med.* 33, 377–431.

Greene, J. D., and Paxton, J. M. (2009). Patterns of neural activity associated with honest and dishonest moral decisions. *Proc. Natl. Acad. Sci. U.S.A.* 106, 12506–12511.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., and Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science* 293, 2105–2108.

Keysar, B., Barr, D. J., Balin, J. A., and Brauner, J. S. (2000). Taking perspective in conversation: the role of mutual knowledge in comprehension. *Psychol. Sci.* 11, 32–38.

Kozel, F. A., Johnson, K. A., Mu, Q., Grenesko, E. L., Laken, S. J., and George, M. S. (2005). Detecting deception using functional magnetic resonance imaging. *Biol. Psychiatry* 58, 605–613.

Kurniawan, I. T., Seymour, B., Talmi, D., Yoshida, W., Chater, N., and Dolan, R. J. (2010). Choosing to make an effort: the role of striatum in signaling physical effort of a chosen action. *J. Neurophysiol.* 104, 313–321.

Langleben, D. D., Loughead, J. W., Bilker, W. B., Ruparel, K., Childress, A. R., Busch, S. I., and Gur, R. C. (2005). Telling truth from lie in individual subjects with fast event-related fMRI. *Hum. Brain Mapp.* 26, 262–272.

Leotti, L. A., Iyengar, S. S., and Ochsner, K. N. (2010). Born to choose: the origins and value of the need for control. *Trends Cogn. Sci. (Regul. Ed.)* 14, 457–463.

Lieberman, M. D., and Cunningham, W. A. (2009). Type I and type II error concerns in fMRI research: re-balancing the scale. *Soc. Cogn. Affect. Neurosci.* 4, 423–428.

Masten, C. L., Eisenberger, N. I., Borofsky, L. A., Pfeifer, J. H., McNealy, K., Mazziotta, J. C., and Dapretto, M. (2009). Neural correlates of social exclusion during adolescence: understanding the distress of peer rejection. *Soc. Cogn. Affect. Neurosci.* 4, 143–157.

Onoda, K., Okamoto, Y., Nakashima, K., Nittono, H., Ura, M., and Yamawaki, S. (2009). Decreased ventral anterior cingulate cortex activity is associated with reduced social pain during emotional support. *Soc. Neurosci.* 4, 443–454.

Paus, T. (2001). Primate anterior cingulate cortex: where motor control drive and cognition interface. *Nat. Neurosci.* 2, 417–424.

Petrides, M., and Pandya, D. N. (2002). Comparative cytoarchitectonic analysis of the human and the macaque ventrolateral prefrontal cortex and corticocortical connection patterns in the monkey. *Eur. J. Neurosci.* 16, 291–310.

Petrovic, P., Dietrich, T., Fransson, P., Andersson, J., Carlsson, K., and Ingvar, M. (2005). Placebo in emotional processing – induced expectations on anxiety relief activate a generalized modulatory network. *Neuron* 46, 957–969.

Pollina, D. A., Dollins, A. B., Senter, S. M., Krapohl, D. J., and Ryan, A. H. (2004). Comparison of polygraph data obtained from individuals involved in mock crimes and actual criminal investigations. *J. Appl. Psychol.* 89, 1099–1105.

Raine, A., and Yang, Y. (2006). Neural foundations to moral reasoning and antisocial behavior. *Soc. Cogn. Affect. Neurosci.* 1, 201–213.

Roepstorff, A., and Frith, C. D. (2004). What's at the top in the top-down control of action? Script-sharing and "top-top" control of action in cognitive experiments. *Psychol. Res.* 68, 189–198.

Rudebeck, P. H., Bannerman, D. M., and Rushworth, M. F. (2008). The contribution of distinct subregions of the ventromedial frontal cortex to emotion, social behavior, and decision making. *Cogn. Affect. Behav. Neurosci.* 8, 485–497.

Rushworth, M. F., Behrens, T. E., Rudebeck, P. H., and Walton, M. E. (2007). Contrasting roles for cingulate and orbitofrontal cortex in decisions and social behaviour. *Trends Cogn. Sci. (Regul. Ed.)* 11, 168–176.

Rushworth, M. F., and Behrens, T. E. J. (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. *Nat. Neurosci.* 11, 389–397.

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

Saarni, C., and Lewis, M. (1993). *Lying and Deception in Everyday Life (1-30)*. New York,: The Guilford Press.

Sip, K. E., Lynge, M., Wallentin, M., McGregor, W. B., Frith, C. D., and Roepstorff, A. (2010). The production and detection of deception in an interactive game. *Neuropsychologia* 48, 3619–3626.

Sip, K. E., Roepstorff, A., McGregor, W. B., and Frith, C. D. (2008). Detecting deception: the scope and limits. *Trends Cogn. Sci. (Regul. Ed.)* 12, 48–53.

Spence, S. A., Hunter, M. D., Farrow, T. F., Green, R. D., Leung, D. H., Hughes, C. J., and Ganesan, V. (2004). A cognitive neurobiological account of deception: evidence from functional neuroimaging. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 359, 1755–1762.

van Beest, I., and Williams, K. D. (2006). When inclusion costs and ostracism pays, ostracism still hurts. *J. Pers. Soc. Psychol.* 91, 918–928.

Wager, T. D., Lindguist, M., and Kaplan, L. (2007). Meta-analysis of functional neuroimaging data: current and future directions. *Soc. Cogn. Affect. Neurosci.* 2, 150–158.

Walters, S. B. (2000). *The Truth About Lying. How to Spot a Lie and Protect Yourself from Deception*. Naperville, IL: Sourcebook Inc.

Walton, M. E., Rudebeck, P. H., Bannerman, D. M., and Rushworth, M. F. (2007). Calculating the cost of acting in frontal cortex. *Ann. N. Y. Acad. Sci.* 1104, 340–356.

Whiten, A., and Byrne, R. W. (1988). Tactical deception in primates. *Behav. Brain Sci.* 11, 233–244.

Woodruff, G., and Premack, D. (1979). Intentional communication in the chimpanzee: the development of deception. *Cognition* 7, 333–362.

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

Received: 11 January 2012; accepted: 01 April 2012; published online: 18 April 2012.

Citation: Sip KE, Skewes JC, Marchant JL, McGregor WB, Roepstorff A and Frith CD (2012) What if I get busted? Deception, choice, and decision-making in social interaction. Front. Neurosci. ***6****:58. doi: 10.3389/fnins.2012.00058*

This article was submitted to Frontiers in Decision Neuroscience, a specialty of Frontiers in Neuroscience.