Detecting deception: the scope and limits

Kamila E Sip\(^1,2\), Andreas Roepstorff\(^1,2\), William McGregor\(^2\) and Chris D Frith\(^1,3\)

\(^1\) Center for Functional Integrative Neuroscience, Århus University Hospital, Nørrebrogade 44, Building 30, DK–8000 Århus C, Denmark
\(^2\) Institute of Anthropology, Archeology and Linguistics, University of Århus, Moesgård, DK–8270, Højbjerg, Denmark
\(^3\) Wellcome Trust Centre for Neuroimaging at University College London, 12 Queen Square, London WC1N 3BG, UK

With the increasing interest in the neuroimaging of deception and its commercial application, there is a need to pay more attention to methodology. The weakness of studying deception in an experimental setting has been discussed intensively for over half a century. However, even though much effort has been put into their development, paradigms are still inadequate. The problems that bedevilled the old technology have not been eliminated by the new. Advances will only be possible if experiments are designed that take account of the intentions of the subject and the context in which these occur.

Deception in experimental settings

The neuroimaging of deception has raised considerable interest and controversy that reaches well beyond academic circles [1–4]. There has been much discussion of the validity and possible real-life application of this new and rapidly developing field. The brain imaging community must be prepared for the increasing commercialization of these techniques and their use in legal cases and by the security industry. We would argue the need for caution in putting the results of experimental studies into real-life application. For one thing, as has been examined in detail elsewhere [5,6], general considerations of the nature of deception raise problems with experimental paradigms. Here, we argue that the problem of studying deception with neuroimaging does not lie exclusively in the paradigms applied. Major stumbling blocks concern the concept of deceptive behaviour, the execution of supposedly deceptive behavior in experimental settings, reverse inference from brain activity to mental states, and the translation from the laboratory to real-life settings.

To begin with, it is essential to say something about the nature of deception. A useful characterization is provided by Vrij [7], who defined deception as follows: ‘a (…) deliberate attempt, without forewarning, to create in another a belief which the communicator considers to be untrue’. There are two important components of this definition. First, it captures the common understanding that deceptive behavior is not characterized merely by considerations of the truth value of propositions, but crucially involves deliberation on the part of the deceiver, and an intention to communicate misleading information. Second, deception typically occurs without ‘forewarning’. In other words, there is no instruction to lie or provision of a cue as to when precisely to utter a lie. Unfortunately, such instructions are a feature of nearly all studies of ‘deception’ in the literature. This eliminates the voluntary intention to deceive, in addition to its deliberate execution.

We would argue that such experiments do not, in fact, investigate deception but, at best, some of the complex executive functions indirectly associated with the phenomenon. In other words, the idealization of such experimental situations is not simply a matter of inadequate contextualization but goes to the heart of the act, eliminating its status as deceptive. The absence of this intentional aspect of deception in the experiments is thus more than a mere experimental confound.

But this is not all. There are also crucial contextual problems in the experimental designs. Deception is also socially rooted, and its processing is modified by our moral perception, which in turn depends on the existence or expectation of interactive consequences for our actions, negative or positive. Without such consequences, no harm can come to anyone, either the deceiving individual or the deceived, and we do not achieve a valid representation of the processes involved in deceptive acts. These problems of contextualization of human actions and choices, widely discussed within the field of the pragmatics of human communication [8], demand serious consideration in experimental design.

Cognitive processes associated with deception

Executive processes enable humans to navigate successfully through deceptive communicative interactions by going through a series of cognitive operations, which are repeated if necessary (Box 1). Our intentional goals are directed to satisfying our aims on the basis of cooperation and trust. If deception is a goal, the most basic scenario requires inhibition of prepotent truth responses to make others believe what we want them to believe. Furthermore, representation of truth under the pretence that it is a lie is an equally common form of deception in games, in political negotiations and in everyday life. Regardless of the nature of a deceptive act, the process involves a series of deliberate decisions based on several suboperations such as estimation of outcome, risk of punishment and reward expectation. Information exchange must be monitored and feedback assessed to build up reputation and trust. To use brain imaging to find markers of deception, we must
Box 1. Unpacking interpersonal deception

Figure 1 illustrates various social and cognitive processes that influence deceptive behaviour. All of these processes are closely correlated and mutually dependent. Our considerations are restricted to deceptive behaviour that is entered into deliberately and voluntarily. The decision to deceive is the first step in the deceptive process that is further composed of other interdependent mental operations that are not necessarily deceptive in their nature.

**Information management**
This is crucial during an interaction to maintain deception. It is necessary to keep track of the false beliefs being engendered in the interlocutor. Furthermore, one must try to anticipate the verbal or non-verbal responses of the interlocutor on the basis of what one thinks they know, feel and intend. In particular, mentalizing, the human ability to ‘...predict and explain the mental states of other people, such as desires, beliefs and intentions’ [27], is required to implement information management.

**Impression management**
This is used to generate trust-building strategies. Building a reputation for being trustworthy is a standard deceptive procedure but it has a cost. There is the cost of the initial cooperation that precedes deception. There is the cost of controlling verbal and non-verbal signals to indicate trustworthiness. There is the cost of constantly monitoring feedback and adjusting behaviour to maintain the appearance of trustworthiness.

**Risk management**
This is a necessary component of deception. One must be sure that one gains more than one loses from deception. One must also consider the risk of not being believed, and take into account the consequences of such failure. As became apparent in the Watergate affair, the effort needed to maintain a deception is unpredictable and can develop over time, contingent on the actions of opponents. The cost of being caught red-handed can be enormous in terms of loss of reputation, trust, power and money.

**Reputation management**
This is needed to convince both oneself and one’s peer group that the deception was intended to achieve some greater good and was therefore justified.

consider how these different components of deception might be implemented at a neural level.

**The view from neuroscience: evidence for many systems**
As has been demonstrated extensively over the past decade, the cognitive processes involved in social interactions, such as occur during deception, map onto a wide range of cortical and subcortical networks. Furthermore, the activation of a specific region of the human brain is rarely affiliated to only one such cognitive process.

**Neural basis of decision making, risk taking and reward processing**
Estimation of rewards and risks are, as mentioned earlier, fundamental for deceptive acts. The medial frontal cortex
and adjacent anterior cingulate cortex (ACC) are believed to have an essential role in relating actions to their consequences, representing the expected value of actions and the likelihood of errors [9]. Risky decision making when rewards and losses are uncertain is associated with activation in the ACC, orbitofrontal regions and striatum [10]. More generally, ACC activity is associated with problem solving, motivation and the anticipation of the consequences of action.

**Neural basis of mentalizing**

A major requirement for successful deception is mentalizing, the ability to read and manipulate the mental states of others, in particular their intentions and beliefs [11,12]. An extended network of brain regions, including the dorsal and ventromedial prefrontal cortex (dMPFC and vMPFC), dorsal anterior cingulate cortex (dACC), posterior superior temporal sulcus (pSTS), temporo-parietal junction (TPJ) and posterior cingulate cortex, are activated when mentalizing, particularly when thinking about intentions and beliefs [13]. Regions of the (dMPFC) in particular are activated when people make decisions that have a role in social interactions involving cooperation and competition [14]. This region might also have a special role in the reputation management exercised by the deceiver to ensure that he is trusted by his victim [15].

**Neural basis of response inhibition and control**

During deception, we also need to monitor and control our responses carefully. Inhibition of prepotent responses is associated with activity in the dorsolateral prefrontal cortex (DLPFC), inferior frontal cortex and dACC [16,17]. Such activation is elicited, for example, in GO/NO-GO tasks, where subjects must inhibit responses on NO-GO trials, and in the Stroop task, where subjects must inhibit the automatic response of reading the word rather than naming the ink colour [18]. In both of these cases, continuous monitoring is needed to prevent the production of inappropriate responses. It can thus be questioned whether activity in specific brain regions can be reliably correlated with deception.

**Neuroimaging studies of deception**

As can be seen in the tasks listed in Table 1 [19], much ingenuity has gone into developing novel formats for the laboratory study of deception. For example, comparisons have been made between spontaneous and prepared lies, with additional personal involvement implemented by contrasting autobiographical and non-autobiographical experience [19,20]. Perhaps the study that gets closest to real life is that of Abe et al. [21]. This study introduced a novel twist, in which one experimenter instructed the subject to deceive the other experimenter. However, the study still remains essentially unrealistic because there was a narrow choice for wrong-doing and a lack of unpleasant consequences if the deception was detected. Langleben’s group [22,23] (see also Phan et al. [24] and Gamer et al. [25]) used a modification of the Guilty Knowledge Test to investigate the neural correlates of information inhibition and response suppression. Subjects were given two playing cards and were instructed to lie about possession of one particular card. The problem still remains that the subjects had been instructed to lie, even though they could choose what to lie about.

In spite of the wide range of different experimental paradigms used, the pattern of activations associated with deception is strikingly similar (Table 1), typically including DLPFC, medial prefrontal cortex (MPFC) and ACC. These are precisely the regions that would be expected, given the cognitive processes listed above that are likely to be involved during deception.

The results of these studies are entirely consistent with the idea that deception is a cognitively demanding process that relies on a range of high-level functions including decision making, response monitoring and mentalizing, in addition to more traditional executive functions such as working memory, attention and conflict monitoring. At first sight, the similar pattern of neural activity seen across the studies [ACC, DLPFC, MPFC, inferior frontal gyrus (IFG), amygdala and caudate] might suggest that there is a characteristic pattern of brain activity associated with deception that is robust to variations in the precise paradigm used. The problem is that all of these regions are also activated during the performance of tasks involving the same cognitive processes in the absence of deception. For example, activation of ACC is seen during inhibition of information in memory retrieval tasks that do not involve deception [17]. The same applies to the more primitive regions concerned with basic survival, rather than cognition, such as the amygdala and caudate, where activity is observed during deception [20,21,26,27]. Such activity is also observed during interactions that do not involve deception. The problem is that deception involves many cognitive processes, none of which are unique to it.

When using functional magnetic resonance imaging (fMRI) (or any other physiological measure) to detect deception, we are confronted with the problem of making reverse inferences about cognitive processes from patterns of brain activity [28]. Even if deception reliably activates a certain brain region, we cannot logically conclude that, if that brain region is activated, deception is taking place. Further issues are highlighted in Box 2.

**What can fMRI add in the search for neural markers of deception?**

As rightly stressed by Wolpe et al. [5], traditional paradigms used to detect deceptive behavior, such as the polygraph, have many disadvantages. Can the newer techniques of brain imaging make the breakthrough from the laboratory and provide key evidence about deception for use in criminal cases? In our opinion, the problems with which the old technology has struggled have not been eliminated by the new technology (fMRI, electroencephalography, near-infrared etc.). There is no doubt that the new techniques can study the human brain along many dimensions. We know that fMRI can detect emotional responses [29], activity associated with decision making under risk [30] and anticipated regret [31]. We also know that fMRI can detect activity associated with cognitive load [32] and with mentalizing [15]. Of particular relevance, we know that fMRI can detect activity concerned with the top-down control of emotions [33]. Given this latter ability, it
Table 1. Summary of results from brain imaging studies of deception

<table>
<thead>
<tr>
<th>Study type</th>
<th>Authors</th>
<th>Task</th>
<th>Neural activation</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>fMRI studies, autobiographical events</td>
<td>Spence et al. 2001</td>
<td>S instructed to lie about autobiographical events</td>
<td>DLPFC, VLPFC, ACC, MPFC</td>
<td>Feigning a memory impairment was associated with robust activation in regions associated with executive functions such as attention. These activations were robust across different groups and paradigms, but are not unique markers of deception [51,52].</td>
</tr>
<tr>
<td></td>
<td>Ganis et al. 2003</td>
<td>S instructed to lie about autobiographical events</td>
<td>ACC, left parietal cortex</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nunez et al. 2005</td>
<td>Comparison of spontaneous and prepared lies</td>
<td>DLPFC, ACC, caudate, thalamic nuclei, ACC, DLPFC</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>S instructed to lie about autobiographical and non-autobiographical events</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PET studies/experienced event</td>
<td>Abe et al. 2006</td>
<td>S instructed to lie about experienced events</td>
<td>ACC, SFG, anterior parietal cortex</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Abe et al. 2007</td>
<td>S instructed to lie about autobiographical events</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>In half of the trials, S was told by one experimenter to disobey the other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fMRI studies, Guilty Knowledge Test</td>
<td>Langleben et al. 2002</td>
<td>Guilty Knowledge Test</td>
<td>ACC, SFG, anterior parietal cortex</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>S instructed to conceal the identity of a playing card they had been given. Monetary reward for successful deception</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Langleben et al. 2005</td>
<td>Guilty Knowledge Test</td>
<td>DLPFC, ACC, left IFG, insula, parietal cortex, precuneus, thalamus</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Phan et al. 2005</td>
<td>Similar paradigm to Langleben's study</td>
<td>VLPFC, DLPFC, DMFPC, STS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gamer et al. 2007</td>
<td>Guilty Knowledge Test</td>
<td>Right IFG, cingulate, parietal cortex</td>
<td></td>
</tr>
<tr>
<td>fMRI studies, malingering</td>
<td>Lee et al. 2002</td>
<td>S trained to feign memory impairments; forced-choice word recognition task. S instructed to lie</td>
<td>Bilateral DLPFC, inferior parietal, middle temporal and posterior cingulate together with bilateral caudate</td>
<td>Feigning a memory impairment was associated with activation in regions associated with executive functions such as attention. These activations were robust across different groups and paradigms, but are not unique markers of deception [51,52].</td>
</tr>
<tr>
<td></td>
<td>Lee et al. 2005</td>
<td>S trained to feign memory impairments; forced-choice word and digit recognition task. S instructed to lie</td>
<td>Bilateral DLPFC, cingulate, supramarginal and parietal.</td>
<td></td>
</tr>
<tr>
<td>fMRI studies, mock-crime scenario</td>
<td>Kozel et al. 2005</td>
<td>Mock-crime scenario and Guilty Knowledge Test. S instructed to ‘steal’ a ring or a watch. Monetary reward for successful deception</td>
<td>DLPFC, ACC, MTG, inferior OFG</td>
<td>In the mock-crime scenario, subjects take part in an event in which they are intellectually and emotionally involved, even though it is still in an experimental setting. In the study of Kozel et al. [53], subjects had a free, although necessarily limited choice of what to ‘steal’ and hence what to lie about. Mohamed et al. [54] directly compared a polygraph test with fMRI. However, the paradigm was weakened by the explicit instruction to lie.</td>
</tr>
<tr>
<td></td>
<td>Mohamed et al. 2006</td>
<td>Mock-crime scenario and Guilty Knowledge Test. One group of S (the guilty group) instructed to fire a starter pistol and then deny having done so. Monetary reward for successful deception</td>
<td>ACC, MTG, FFG, insula, hippocampus</td>
<td></td>
</tr>
<tr>
<td>fMRI studies, detecting non-verbal deceptive intentions</td>
<td>Grèzes et al. 2004</td>
<td>Subjects used non-verbal cues to detect deception</td>
<td>ACC, amygdala</td>
<td>In the first study of Grèzes et al. [26], amygdala activity only occurred when the subject was personally involved in the deception.</td>
</tr>
<tr>
<td></td>
<td>Grèzes et al. 2006</td>
<td>Subjects observed a video of an event in which they or someone else had been deceived</td>
<td>ACC and STS, amygdala and fusiform gyrus</td>
<td></td>
</tr>
</tbody>
</table>

*FFG, fusiform gyrus; MTG, middle temporal gyrus; OFG, orbitofrontal gyrus; SFG, superior frontal gyrus.*
Box 2. Four problems for interpreting imaging studies of deception

First problem: making reverse inferences from physiology to mental states

We would expect to see activity in brain areas associated with emotion, mentalizing and risk taking during deception. However, if we see activity in these areas, we cannot logically make the reverse inference and conclude that deception is taking place [28,39]. Observation of brain activity is not enough, on its own, to make reverse inferences.

Second problem: making inferences about individuals on the basis of group studies

Laboratory studies typically base their conclusions on data from groups of subjects. For forensic purposes, the techniques have to be applied to single cases. As yet, there is no established practice for ensuring that fMRI results from a single subject are reliable and consistent with group data [5,40]. Furthermore, the individuals involved in forensic cases are likely to be drawn from different populations from those typically investigated in academic studies.

Third problem: studying choice in the laboratory

Explicit instruction to lie lift moral sanctions against lying and remove the need for decision making under conflict, and are crucial components of deception. One of the few examples of a deception that is not elicited by the experimenter occurs when the subject decides to disobey the rules of the experiment [41]. However, this might only be revealed if the perpetrator chooses to talk about it afterwards [42].

Fourth problem: interpretation must take into account the context

The factors that determine a decision to deceive (Box 1) are not about states of the world. They are about beliefs in the mind of the protagonist. Hence, different people will have different attitudes to the same act of deception. Thus, brain activity cannot properly be interpreted unless the attitudes and beliefs of the person studied are taken into account. Psychopathy provides an extreme example of this. The lack of emotional response observed in psychopaths, when they tell a lie, stem from their attitude to the victim of the deception. They have no empathy for the potential suffering that their actions might cause [43].

might be possible to detect that someone is using counter-measures to control their emotional responses. Finally, using the novel technique of voxel pattern classification [34], fMRI signals seem to reveal a person’s covert intentions [35], at least in a highly constrained experimental setting, and even identify sensory processing of which the person is not aware [36]. Such techniques might, perhaps, be able to detect when someone is lying in the laboratory if used with valid paradigms [37] (Box 1). It is another matter, however, to translate these possibilities into practical use in a forensic setting (Box 2).

Recently, Spence et al. [38] gave a thorough account of the current state of the evidence, treating some of the aspects that we also raise here. Unfortunately, this seems to have had little impact on the field. Acknowledging a problem is not a solution to it. Experimental paradigms still fail to overcome the key problems.

The future: how can the study of deception be improved?

Given the current results of neuroimaging studies of deception and the current controversies surrounding the topic, we suggest that there are two key factors that need to be considered to improve the study of deception. First, we must recognize that deception involves a social dimension and needs to be studied as such (Boxes 2 and 3). In particular, the pragmatics of social interaction need to be considered in further research [8]. We must recognize that deception can occur in the absence of overt lying. Second, we must develop paradigms in which subjects have a real choice as to whether and when to lie. The intention to deceive someone is crucial. Only by using such paradigms can we start to study realistic deception and learn more about the neural basis of deceptive behaviour.

Acknowledgements

Our work is supported by the Danish Research Council for Culture and Communication, The Danish National Research Foundation and the Wellcome Trust.

References

Free journals for developing countries

In 2002, the WHO and six medical journal publishers launched the Health InterNetwork Access to Research Initiative, which enabled nearly 70 of the world’s poorest countries to gain free or reduced-cost access to biomedical literature through the internet. Currently more than 70 publishers are participating in the program, providing access to over 2000 journals.

Gro Harlem Brundtland, former director-general for the WHO, said that this initiative was “perhaps the biggest step ever taken towards reducing the health information gap between rich and poor countries”.

For more information, visit www.who.int/hinari