Neural Correlates of Different Types of Deception: An fMRI Investigation

Deception is a complex cognitive activity, and different types of lies could arise from different neural systems. We investigated this possibility by first classifying lies according to two dimensions, whether they fit into a coherent story and whether they were previously memorized. fMRI revealed that well-rehearsed lies that fit into a coherent story elicit more activation in right anterior frontal cortices than spontaneous lies that do not fit into a story, whereas the opposite pattern occurs in the anterior cingulate and in posterior visual cortex. Furthermore, both types of lies elicited more activation than telling the truth in anterior prefrontal cortices (bilaterally), the parahippocampal gyrus (bilaterally), the right precuneus, and the left cerebellum. At least in part, distinct neural networks support different types of deception.

Introduction

Deception occurs when one person attempts to convince another to accept as correct what the prevaricator believes is incorrect (typically in order to gain a benefit or avoid punishment; cf. Spence et al., 2001). Given the obvious importance of detecting deception, individuals as well as entire societies have long sought reliable methods for determining when a person is lying (Ekman, 1992, 2001). Traditionally, observers have tried to detect lies by noting subtle behavioral cues. Indeed, researchers have characterized a number of nonverbal cues that are associated with deception, but none of these cues is entirely diagnostic or reliable. For example, microexpressions, brief and incomplete changes in expression — such as head shakings or negative facial expressions — are among the most reliable nonverbal cues for deception (Mehrabian, 1971; Burgoon and Buller, 1994; Frank and Ekman, 1997); the pitch of the voice tends to be elevated when people are engaged in deception (Vrij, 1994; Zuckerman et al., 1979); the body posture is generally more rigid when one is lying than when one is telling the truth (Mehrabian, 1971; Vrij, 1994); and alterations in patterns of eye contact are also associated with deception (Hornath et al., 1994). These behavioral cues are generally thought to reflect increased physiological arousal during deception, which may arise because the individual feels guilty, is afraid of being detected, or is excited at the thought of deceiving others (Ekman, 1992).

In an effort to develop reliable objective measures of deception, researchers and criminologists have devised various machine-based techniques that typically attempt to measure arousal. The polygraph, which monitors physiological functions such as heart rate, breathing rate, and skin conductance, has been used in diverse ways in the service of detecting deception (Office of Technology Assessment, 1983). One class of methods (control question test, CQT) compares physiological responses to relevant questions (e.g. ‘Did you steal the car that was reported missing?’) with responses to irrelevant questions (e.g. ‘Were you born in California?’) and control questions (e.g. ‘During the first 15 years of your life, did you ever steal anything?’ (Horowitz et al., 1997; Podlesny and Raskin, 1977)).

Another class of polygraph methods, the Guilty Knowledge Test (GKT), has focused on detecting physiological changes in response to questions that could only be answered by the perpetrator of a crime (Lykken, 1974; MacLaren, 2001).

One general problem with all polygraph methods is that they detect increases in measures that reflect increased arousal, which are typically interpreted as reflecting guilt and fear. These measures can confound lie detection in two ways. First, guilt and fear can occur in many situations other than during deception, and hence the measures do not necessarily index deception per se. Second, if the liar does not feel guilty, he or she may not evince the physiological reaction. Thus, even though exceptions have been noted (Raskin and Hare, 1978), in general standard lie detection techniques might be characterized as ‘guilt detection techniques’.

In this study we focus on one potential method for circumventing these problems: Namely, we examine directly the organ that produces lies, the brain (other methods, not explored here, may include using behavioral or peripheral psychophysiological measures that correlate with cognitive processes of interest). This neurobiologically based strategy relies on identifying specific patterns of neural activation that underlie deception. The logic we adopt here has led researchers to use brain-monitoring techniques to try to develop improved lie detection techniques. For instance, researchers have used event-related potentials in the Guilty Knowledge Test, and have reported more accurate discrimination rates than is possible with polygraphic methods (Allen and Iacono, 1997). However, even these methods are not ideal because of the limited spatial resolution of the technique, making it difficult to disentangle complex cognitive processes occurring simultaneously. In the present study we used functional magnetic resonance imaging (fMRI) to monitor neural activation while people lied or told the truth. To date, three fMRI studies of deception or a related topic (e.g. malingering) have been published (Spence et al., 2001; Langleben et al., 2002; Lee et al., 2002). The results have not been consistent. Because the precise questions asked and methods employed in these studies are different from those used in our study, we will defer discussing them until after describing our own findings.

In addition to the limitations inherent in previous lie detection techniques, a second — and in some respects deeper — problem with all prior lie detection methods is that they rest on the assumption that there is only one type of lie. One of the strengths of contemporary theory in cognitive neuroscience is that it distinguishes subtypes of a given function. For example, ‘memory’ may be decomposed into working, episodic, semantic memory, and so on. In the present work, we applied the same approach to study deception and asked what types of processing
differences might distinguish different types of lies. We focused on two orthogonal dimensions. Along the first dimension we differentiated between spontaneous and memorized lies. A spontaneous lie is constructed based on stored information, probably using a mixture of semantic and episodic knowledge. For instance, one could lie about what one ate for lunch by retrieving a specific episode about what one ate some other time (episodic memory), or one could think about what foods one plausibly could eat for lunch (semantic memory).

In contrast, if a lie is memorized in advance, one needs only to retrieve it from memory. We hypothesize that one defining feature of memorized lies is that they are not as rich in detail or as well learned as truthful knowledge. Real experiences are rich in incidental perceptual detail, whereas lies may often consist only of bare-bones descriptions. Moreover, actual experience may be registered in multiple modalities, and hence is subject to ‘dual encoding’ (Paivio, 1971). If so, then all else being equal, representations of lies should be more difficult to retrieve than truthful knowledge.

In the second dimension, we differentiated between lies that may be isolated versus lies that fit into a scenario (i.e. a coherent story). Spontaneous lies that are isolated are easier to generate than coherent lies because one does not have to cross-check details to ensure that they fit into a larger scheme. In terms of underlying neurocognitive processes, this translates into working memory’s being more engaged when one generates a coherent lie than an isolated lie because more information has to be held in mind and evaluated (Smith and Jonides, 1998). In contrast, for memorized lies, those that fit into a coherent scenario may be easier to generate because it is easier to recall a lie when more retrieval cues are present (Schacter, 1996). Treating these two dimensions as orthogonal, we can identify four distinct types of lies – and each type should be associated with a different pattern of neural processing.

In this study we focused on two extremes from this taxonomy: Spontaneous-Isolated (SI) lies and Memorized-Scenario (MS) lies (Fig. 1). To construct a SI lie one needs to retrieve information from semantic and/or episodic memory and generate a viable lie rapidly, keeping in mind many possibilities (i.e. the truth, so to be able to avoid it, together with a number of potential lies) and selecting among them. Accordingly, we hypothesized that, relative to telling the truth, telling SI lies should result in stronger activation in neural structures underlying: (i) semantic and episodic retrieval [e.g. ventrolateral prefrontal cortex and anterior prefrontal cortex, respectively (Duncan and Owen, 2000; Fletcher and Henson, 2001), precuneus (Krause et al., 1999), and possibly ventral stream regions, if these retrieval operations are also accompanied by visual imagery (Kosslyn et al., 2001)]; (ii) working memory [e.g. dorsolateral prefrontal cortex (Smith and Jonides, 1998)]; and (iii) response inhibition and conflict monitoring [e.g. the anterior cingulate (Braver et al., 2001; Ruff et al., 2001)]. In contrast, to generate MS lies, one needs only to access knowledge stored in episodic memory. Thus, for MS lies we hypothesized that we would find increased activation (relative to telling the truth) in brain regions associated with retrieving information from episodic memory [anterior prefrontal cortex (Buckner et al., 1998; Duncan and Owen, 2000) and the precuneus (Krause et al., 1999)].

Materials and Methods

Participants
Three males and seven females, between the ages of 20 and 30 (mean age 25), volunteered to participate for pay. All participants had normal or corrected-to-normal vision. The study was conducted with the informed consent of each participant and the approval of the Harvard University and McLean Hospital Institutional Review Boards.

Pre-testing Procedure and Stimuli
We began by interviewing all participants to obtain detailed information about two of their actual experiences (Fig. 1a). We asked them to write down details about the most memorable work experience they ever had (‘work’ situation) and about the most memorable vacation they ever took (‘vacation’ situation). Participants returned to the lab on average 1 week after the initial session, which allowed us time to prepare and record questions based on their specific experiences. We first reviewed in general terms the reports they provided earlier, to refresh their memories. Then we instructed the participants to generate an alternative scenario for one of the situations (Fig. 1b). The investigator helped them to
constrain the scenario so that it fit the prepared questions. For example, if they actually took their vacation in Florida, we instructed them to pretend that the vacation took place in another location in the United States (e.g. California); if they traveled there by car, we instructed them to pretend they used another means of transportation (e.g. by plane), and so on. We helped the participants in this process by ensuring that the scenarios they generated were coherent and internally consistent. Participants were then asked to rehearse and memorize this ‘false scenario’, so that they could answer questions based on it. This scenario was then used to assess MS lies.

Following this, the participants took part in a practice session outside the scanner, which simulated what they would experience during the actual test session. During this practice session, participants first were given instructions and asked to paraphrase them to ensure that they understood the task. The instructions reiterated the importance of remaining still during each scan; to minimize the possibility of random responses during the SI one-word scans, when participants produced verbal responses, we told them that their responses (without the condition label) would be recorded and reviewed later by a separate ‘judge’ who would try to determine which were lies. This instruction was not only intended to make them really try to lie, but also to discourage bizarre responses (such as ‘purple’ or ‘heavy’ to a question like ‘what color are your mother’s eyes’, which would be easily spotted as a lie). We then asked whether they had any questions, and if so the investigator answered them.

Image Acquisition

fMRI acquisition was conducted on a 1.5 T scanner (General Electric Signa, Milwaukee, WI) with a standard quadrature head coil and echoplanar capability (Instascan, ANMR Systems, Wilmington, MA). T2*- weighted echoplanar images sensitive to blood oxygen level- dependent contrast (BOLD) were acquired during the functional scans (gradient echo; TE = 3000 ms; T1 = 40 ms; alpha = 75°; image matrix = 64 × 128; in-plane resolution = 3.125 × 3.125 mm; slice thickness = 6 mm). Sixteen to twenty axial slices per volume were acquired, depending on head size. Anatomical images for these slices were obtained with a T1-weighted sequence (T1 = 500 ms; T E = 11 ms). Whole-brain anatomical images (coronal) were acquired after the functional scans with a SPGR sequence (T1 = 35 ms; TE = 5 ms; FOV = 240 mm; slice thickness = 1.5 mm; imaging matrix = 256 × 192).

Analysis

Data were analyzed with AFNI (Cox, 1996). The data were first corrected for motion artifacts using AFNI program ‘3dvolreg’ (Cox and Jesmanowicz, 1999). Because this motion correction algorithm can only correct small motions, the threshold for the exclusion of a scan due to motion was a shift of more than 4 mm in any direction and a rotation of more than 1.5°. No scan exceeded this threshold. We estimated maps of percent BOLD signal change for the eight series of trials by using the correlation methods described in Cox (Cox, 1996). We then transformed these maps into Talairach space (Talairach and Tournoux, 1988), using the scheme provided by AFNI. Briefly, the brain is divided into 12 regions by means of user-placed markers (including the anterior and posterior commissures) and a continuous piecewise affine transformation is then used to transform the original brain into Talairach space. These maps were then resampled onto a 5 × 3 × 3 mm grid and smoothed with a Gaussian filter (full-width half-maximum = 7 mm, AFNI program ‘3dmerge’).

Following this, we submitted the maps to a one-way repeated-measures ANOVA to identify regions of interest (ROIs) that showed a main effect of condition. We only retained clusters of 40 or more contiguous voxels that were significant at P < 0.005, leading to an alpha of 0.05 for the entire 3D image. This minimum cluster size was determined using the Monte-Carlo approach described by Xiong and colleagues (Xiong et al., 1995) and implemented by programs ‘3dFWHM’ and ‘AlphaSim’ (with 1000 iterations) in AFNI. This method (i) estimates the smoothing present in the data based on a variant of the algorithm described by Forman et al. (Forman et al., 1995), and (ii) determines the number of clusters of a given size that would be significant at a particular threshold due to chance. The probability of a false positive detection across the entire image is then determined by the frequency counts of cluster sizes. Next, we performed planned contrasts on functionally defined ROIs, comparing the two lie conditions with the truth condition, and the two types of lies directly. The alpha for the planned contrasts within the ROIs was P = 0.005, corresponding to Z = 2.81. Preliminary analyses did not reveal reliable differences in the results from the two response modalities (yes/no and one-word response), and thus we combined the data over the two types of responses in order to increase statistical power.

Results and Discussion

The behavioral data from seven participants (data from the remaining three were not recorded due to equipment problems) indicate that the participants did in fact follow the instructions. In fact, the average error rates (defined as responses that were not appropriate for that condition) were less than 10% in every condition (51 of the 56 participant-by-condition cells had error rates below 10%). To obtain a stable baseline we administered the truth condition twice (with different questions) which resulted in a total of eight blocks (two for the MS lie condition, two for the SI lies condition, and four for the Truth condition, half the blocks requiring a manual response, and half requiring a verbal response for each condition).

Each block lasted 210 s, beginning with 30 s of rest followed by three cycles during which 30 s of an experimental condition (five trials, lasting 6 s each; the auditory question lasted 3 s, on average, and the length did not differ between conditions) alternated with 30 s of rest. Therefore, there were 15 trials for each block, regardless of the response modality. We used a blocked design instead of an event-related paradigm because the study was designed to detect differences between conditions (given the relatively limited number of stimuli), not to estimate the timecourse of the hemodynamic response (Birn et al., 2002). We counterbalanced the order of the conditions across participants with the constraint that both lie conditions always came first to avoid potential short-term interference from the truth condition (such as actively having to inhibit primed responses).
framework assumes that the generation of various types of lies we have described in the introduction, our neurocognitive signatures for various types of lies. These results suggest that during the generation of a SI lie, one may need to access semantic and episodic knowledge. These functional demands were reflected by activation of (bilateral superior) BA 10 (Grady, 1999), the precuneus (Krause et al., 1999) and the cerebellum (Andreasen et al., 1999). In addition, this process appears to be accompanied by visual imagery (right cuneus (Kosslyn et al., 2001)). These results also suggest that, while constructing a viable lie, the retrieved information is maintained in working memory, which was reflected by activation in BA 8/9 and posterior visual cortex [fusiform gyrus and cuneus (Grady, 1999)]. Moreover, they suggest that, while formulating the lie one may need to check that it is not the truth but nevertheless is plausible, which may in part be responsible for activation in the anterior cingulate (Ruff et al., 2001). Finally, they suggest that the generated lie is encoded into episodic memory, which would explain why the parahippocampal cortex (Wagner et al., 1998; Epstein et al., 1999) was activated.

In contrast, to generate a MS lie, the participants first would have needed to retrieve the false scenario from episodic memory, which relied on bilateral superior BA 10 and right inferior BA 10 (Grady, 1999), the precuneus (Krause et al., 1999), and the cerebellum (Andreasen et al., 1999). They then may have generated a lie according to the memorized scenario, and subsequently encoded the reconstructed lie into episodic memory [which again relied on parahippocampal cortex (Epstein et al., 1999; Wagner et al., 1998)]. Note that, in general, activation of right BA 10 tends to be larger for items that are repeated (old) than for new items (Wagner et al., 1998). Since half of the T questions were repeated from the SI and MS blocks, this could introduce a confound. However, if anything, item repetition should decrease the size of the difference between the MS and T conditions. Since none of the questions in the lie conditions were repeated, this factor cannot affect the direct comparison between the MS and SI conditions.

The only region that was significantly more active when participants told MS lies than when they told SI lies was the right inferior BA 10, which has been implicated in episodic retrieval operations (Fletcher and Henson, 2001; Grady, 1999); see (MacLeod et al., 1998) for other tasks that activate BA 10). Although MS lies were both coherent and memorized, the task did not require subjects to cross-check or make use of the processes which, as an ensemble, may provide reliable neural signatures for various types of lies. These results suggest that during the generation of a SI lie, one may need to access semantic and episodic knowledge. These functional demands were reflected by activation of (bilateral superior) BA 10 (Grady, 1999), the precuneus (Krause et al., 1999) and the cerebellum (Andreasen et al., 1999). In addition, this process appears to be accompanied by visual imagery (right cuneus (Kosslyn et al., 2001)). These results also suggest that, while constructing a viable lie, the retrieved information is maintained in working memory, which was reflected by activation in BA 8/9 and posterior visual cortex [fusiform gyrus and cuneus (Grady, 1999)]. Moreover, they suggest that, while formulating the lie one may need to check that it is not the truth but nevertheless is plausible, which may in part be responsible for activation in the anterior cingulate (Ruff et al., 2001). Finally, they suggest that the generated lie is encoded into episodic memory, which would explain why the parahippocampal cortex (Wagner et al., 1998; Epstein et al., 1999) was activated.

In contrast, to generate a MS lie, the participants first would have needed to retrieve the false scenario from episodic memory, which relied on bilateral superior BA 10 and right inferior BA 10 (Grady, 1999), the precuneus (Krause et al., 1999), and the cerebellum (Andreasen et al., 1999). They then may have generated a lie according to the memorized scenario, and subsequently encoded the reconstructed lie into episodic memory [which again relied on parahippocampal cortex (Epstein et al., 1999; Wagner et al., 1998)]. Note that, in general, activation of right BA 10 tends to be larger for items that are repeated (old) than for new items (Wagner et al., 1998). Since half of the T questions were repeated from the SI and MS blocks, this could introduce a confound. However, if anything, item repetition should decrease the size of the difference between the MS and T conditions. Since none of the questions in the lie conditions were repeated, this factor cannot affect the direct comparison between the MS and SI conditions.

The only region that was significantly more active when participants told MS lies than when they told SI lies was the right inferior BA 10, which has been implicated in episodic retrieval operations (Fletcher and Henson, 2001; Grady, 1999); see (MacLeod et al., 1998) for other tasks that activate BA 10). Although MS lies were both coherent and memorized, the task did not require subjects to cross-check or make use of the
coherent characteristics of the memorized scenario. Thus, we interpret the findings in terms of the fact that MS lies were based on memorized scenarios *per se*. Specifically, the right inferior BA 10 may have been more active during MS lying because SI lies do not require one to retrieve information only from episodic memory; instead, SI lies may rely on a mixture of episodic and semantic information. The fact that the MS lies were learned as a scenario, which included relations among the lies, may have also been a factor contributing to this effect in BA 10 (Christoff *et al*., 2001).

In addition, the fact that BA 10 was more strongly activated during MS lying than during telling the truth could indicate that MS lies are more difficult to retrieve than truthful knowledge. That is, unlike MS lies, truthful knowledge is acquired via extensive and multimodal interactions with the real world. Thus, truthful memories are more redundant and have many more
retrieval cues than MS lies, which are acquired during a brief verbal exchange that provided only limited details to be encoded. One neural correlate of this difference may be that truthful memories are encoded in a larger network of areas than rehearsed lies, as suggested by modality-specific areas activated during episodic retrieval (Wheeler et al., 2000).

A set of regions was significantly more active during the SI than the MS lie condition. The stronger anterior cingulate activation is consistent with the notion that this region is involved in conflict monitoring (Ruff et al., 2001) and in the inhibition of competing responses (Braver et al., 2001). Although competing responses may also be present when people tell MS lies because of incidental recall of truthful knowledge, the response in the MS lie condition is unique and is entirely determined by the alternative scenario memorized prior to the scanning session. Anterior cingulate activation has also been associated, among other factors, with working memory load (Bunge et al., 2002), and with arousal (Lane et al., 1997), which could also contribute to the present finding of stronger anterior cingulate activation in the SI condition. Stronger activation in visual cortex is consistent with the idea that visual imagery may be used to generate SI lies. Visual imagery may not be used when one tells MC lies because the participants memorized verbal responses when constructing the alternative scenario. Kosslyn and Jolicouer (Kosslyn and Jolicouer, 1980) found that imagery typically is not used spontaneously when people have either memorized the response or can infer it easily from associated information (such as superordinate categories); the SI condition had neither of these characteristics, whereas the MS condition did.

Lastly, we note that some of the activated regions were entirely unexpected. Specifically, we did not predict modulation of activation in the primary motor cortex (close to the hand and mouth representations). We can speculate that these activations may be due to some differences in the motor response (the response times were not different, but one could press the button harder while generating an SI lie), or could be related to the presence of competing potential lies, because they were largest in the SI condition (DeSoto et al., 2001).

To our knowledge, there have been only three published MRI studies related to deception (Spence et al., 2001; Langleben et al., 2002; Lee et al., 2002). In the study by Spence et al. (Spence et al., 2001), the deception condition consisted of asking people to lie in response to yes/no questions by pressing one of two buttons. During any given series of trials the participants alternated between lying and telling the truth, depending on the color of a probe. The main finding of the study was bilateral activation in ventrolateral prefrontal cortex (BA 47) in the lie compared with the truth condition, a finding that was interpreted to reflect motor response inhibition. According to our taxonomy, this deception condition can be characterized as spontaneous and not fitting into a coherent scenario, and thus is an SI lie. Although we found activation in the anterior cingulate for SI lies (close to the medial frontal regions reported by Spence et al. (2001)), which we interpreted as related to response inhibition, we did not find significant activation in BA 47 per se in the SI lie condition. This could be due to our random-effect analysis, which, while giving us more confidence in the generalizability of the results, was more conservative than the fixed-effect analysis reported by Spence et al. (Spence et al., 2002). Consistent with this hypothesis, we found a cluster of nine voxels for which the SI > T contrast was significant at P < 0.005 in right BA 47 when we compared SI lies with Truth in a spherical ROI (8 mm radius) centered at the coordinates reported in Spence et al. (Spence et al., 2001). It is also possible that alternating between lying and telling the truth in the same series of trials changes the strategies participants use to perform the task (Dove et al., 2000), and that BA 47 would not be more active if telling the truth and lying were carried out in separate sets of trials.

A comparison with the results reported by Lee et al. (Lee et al., 2002) and by Langleben et al. (Langleben et al., 2002) is more difficult because of major differences between the paradigms. However, the prefrontal-parietal network reported by Lee et al. (Lee et al., 2002) when people deliberately performed poorly on a set of arithmetic problems, or faked poor memory on a set of simple autobiographic questions, is generally consistent with our findings. The anterior cingulate activation found by Langleben et al. (Langleben et al., 2002) when people lied compared with when they told the truth in a simplified version of the GKT is also consistent with our findings.

All the fMRI studies of deception conducted so far, including ours, have used group analyses to detect difference between telling lies and telling the truth. This is a reasonable first step, but whether fMRI can become a useful tool for the detection of deception (setting aside for now important practical issues such as its cost) depends on whether reliable neural signatures of deception can be identified in single participants and in single trials. Thus, a substantial amount of research both on the deception paradigms and on the analysis methods remains to be conducted before we can fully assess the potential of fMRI as a lie detection device.

In summary, this study is a first attempt to demonstrate that different types of lies are associated with different patterns of brain activation. The results suggest that future neuroimaging studies of deception in more realistic settings should not collapse qualitatively different types of lies into a single category. This inappropriate pooling of results would increase variability in the data, and would obscure the ability to identify signals associated with specific types of deception.

To conclude, we must note that our two dimensions for characterizing types of lies are just the beginning. For example, lies are associated with a greater or lesser emotional response. A major limitation of our study compared to real settings is that the participants were not as emotionally involved in lying as they would be in a non-laboratory situation. Although we tried to use interesting scenarios about memorable events of their lives, the participants probably did not feel bad or guilty about lying (in fact, quite the reverse — they were cooperating by following the instructions — but were still engaged in prevarication). In addition, one can lie about one’s own actions or about something one merely observed. We suspect that this dimension of self-involvement would also affect the types of processes that underlie the generation of lies. Thus, we are faced with at least four dimensions along which lies may vary, and there are undoubtedly more. An accurate and precise lie detection system will likely have to account for these dimensions and exploit the variations that arise from the different types of processing involved in the different types of lies.

Notes

We are grateful to Dr Haline E. Schendan for insightful comments and to Anne Smith at the McLean Hospital for technical help. This work was supported by a contract from the Office for Research and Development (Contract no. 98-F134600-000).

Address correspondence to Giorgio Ganis, Department of Psychology, Harvard University, 35 Kirkland Street, Cambridge, MA 02138, USA. Email: ganis@wjh.harvard.edu.
References


