Neural Correlates of Mental State Decoding in Human Adults: An Event-related Potential Study

Mark A. Sabbagh, Margaret C. Moulson, and Kate L. Harkness

Abstract
Successful negotiation of human social interactions rests on having a theory of mind—an understanding of how others' behaviors can be understood in terms of internal mental states, such as beliefs, desires, intentions, and emotions. A core theory-of-mind skill is the ability to decode others' mental states on the basis of observable information, such as facial expressions. Although several recent studies have focused on the neural correlates of reasoning about mental states, no research has addressed the question of what neural systems underlie mental state decoding. We used dense-array event-related potentials (ERP) to show that decoding mental states from pictures of eyes is associated with an N270–400 component over inferior frontal and anterior temporal regions of the right hemisphere. Source estimation procedures suggest that orbitofrontal and medial temporal regions may underlie this ERP effect. These findings suggest that different components of everyday theory-of-mind skills may rely on dissociable neural mechanisms.

INTRODUCTION
Successful negotiation of everyday human interactions rests on having a theory of mind—an understanding of how others' behaviors can be understood in terms of internal mental states, such as beliefs, desires, intentions, and emotions (Wellman, 1990). Research investigating the cognitive and conceptual bases of theory of mind got its start in the area of developmental and comparative psychology (e.g., Wimmer & Perner, 1983; Premack & Woodruff, 1978). Over the past 10 years, theory-of-mind research has become an interdisciplinary hub for developmental, clinical, social, cognitive, and comparative psychologists interested in understanding the foundations of everyday social-cognitive judgments and inferences (Malle, Moses, & Baldwin, 2001). Even more recently, cognitive neurologists and neuroscientists have begun to explore the neural bases of theory-of-mind skills. This latter brand of research is of basic importance for forming a picture of how the cognitive skills that enable high-level social understanding are organized in the human brain. Moreover, this research could be especially important for understanding the neuropathological bases of clinical disorders in which theory-of-mind reasoning may be especially impaired, such as autism (Baron-Cohen, 1995).

There have been a handful of neuroimaging (e.g., fMRI, PET), neurophysiological (e.g., ERP, MEG), and neuropsychological studies aimed at delineating the neural mechanisms that support theory-of-mind reasoning. The bulk of these studies have found evidence that medial frontal and prefrontal regions of the left hemisphere make important contributions to theory-of-mind reasoning (Siegal & Varley, 2002; Frith & Frith, 1999). For instance, Fletcher et al. (1995) developed a PET paradigm to reveal recruitment of left medial frontal areas specifically during tasks in which participants had to explain characters' behavior on the basis of false beliefs. The importance of left medial frontal areas was also shown in imaging studies that required making judgments about objects from another person's perspective (Goel, Grafman, Sadato, & Hallett, 1995). In a recent event-related potential (ERP) study, Sabbagh and Taylor (2000) showed that this left frontal activity was specifically associated with reasoning about beliefs and not other nonmental representations (i.e., photographs). The neuroimaging and neurophysiological findings regarding the contribution of left frontal regions to theory-of-mind reasoning have been replicated across a number of studies using both verbal and nonverbal techniques (e.g., Liu, Sabbagh, Wellman, & Gehring, 2002; Gallagher et al., 2000; Happé et al., 1996). Moreover, these findings converge with some recent neuropsychiatric evidence showing that individuals with acquired damage in left anterior frontal regions show impaired performance on tasks similar to those used in the neuroimaging studies (Channon & Crawford, 2000).

Although these findings provide compelling evidence regarding the role of the left medial frontal cortex in reasoning about others' mental states, there may be
some reason to believe that these regions do not represent the only neurocognitive contribution to everyday theory-of-mind reasoning. In addition to reasoning about mental states, theory-of-mind reasoning also requires the ability to “decode” mental states on the basis of immediately available information, such as facial expression or tone of voice (c.f., Perner, 1991; Wellman, 1990). Examples of decoding mental states might include identifying others’ focus of attention based on their direction of gaze, or identifying whether someone is being playful based on their facial expression. Ordinarily, mental state decoding skills work in concert with mental state reasoning skills in order to produce reliable inferences about others’ intentions in action. For instance, reliable detection of sarcasm (a high-level communicative intention) likely involves decoding relevant mental state information from prosodic contours of the voice, as well as integrating that information with particular facts about the speaker and his or her attitudes regarding the matter in question.

To date, very few studies have been designed specifically to investigate the neural bases of decoding mental states. As such, it is not known whether the same neural systems that have been characterized in tasks that require reasoning about mental states would also contribute to the ability to decode mental states. Some evidence in this regard, however, may come from studies investigating the neural bases of facial emotion recognition. After all, emotion is a mental state that can be reliably decoded from observable facial expressions.

There has been considerable work delineating the neural systems that are important for processing stimuli that have emotional content (e.g., Pizzagalli, Regard, & Lehmann, 1999; Carretié, Iglesias, & Garcia, 1997; Lang, Nelson, & Collins, 1990). In these studies, neural activity is recorded as participants are presented with stimuli (e.g., faces) of varying emotional valence, but participants are not typically asked to make judgments about the semantic content of the emotional stimulus (e.g., What emotion is displayed in this face?). Indeed, surprisingly few studies have investigated the neural systems that play an important role in decoding others’ emotional mental states (Haxby, Hoffman, & Gobbini, 2002). Some neuropsychological evidence suggests that amygdala damage impairs recognition of facial emotions, especially fear (e.g., Calder et al., 1996; Adolphs, Tranel, Damasio, & Damasio, 1994). These findings are supported by a small-scale (three subject) MEG study, in which medial temporal regions of the right hemisphere were activated in an emotion recognition task (Streit et al., 1999). In addition to medial temporal regions, there is some evidence that individuals with ventral frontal lesions, especially in the right hemisphere, are also impaired in their abilities to decode emotions from facial expressions (Hornak, Rolls, & Wade, 1996).

These studies have led to the intriguing hypothesis that medial temporal regions (including the amygdala) and ventral frontal cortical regions, such as the orbitofrontal cortex, may make an important contribution to decoding others’ emotional mental states (Kawashima et al., 1999; Rolls, 1996; Baron-Cohen, 1995). In addition, there is some evidence that these neural systems may be right-lateralized. However, interpretation of these studies is limited somewhat by their small scale and because the stimuli used for emotional judgments were different from ones used for control judgments (e.g., sex, identity). A notable exception, however, comes from Baron-Cohen et al. (1999), who used fMRI to index the neural systems important for making judgments about others’ complex emotional and intentional states based on pictures of eyes (i.e., pictures of faces cropped from the brow to the bridge of the nose). Importantly, in the control task, participants were asked to make judgments about the sex of the same eyes. Findings revealed activation of a number of areas, including the amygdala and medial frontal areas of the left hemisphere. At first blush, these findings appear to be at odds with previous suggestions that the right hemisphere may make a special contribution to decoding mental states. However, the authors themselves note that these findings should be interpreted cautiously given the fact that the pictures of eyes were presented along with complex emotion words. Thus, some of the neural activation indices may have reflected enhanced linguistic processing during the emotional judgments. This confound makes it difficult to isolate the aspects of neural activity that were associated with decoding others’ mental states, per se.

The goal of the present study is to use 128-channel dense-array ERP (see Figure 1) to delineate the spatial and temporal aspects of the neural systems that are important for decoding others’ emotional mental states. To do this, we have adapted the “Reading the Mind in the Eyes” task (Baron-Cohen, Wheelwright, Hill, Raste, & Plum, 2001) for use in an ERP paradigm. In addition to having been used in the fMRI study described above, the task has been used in a variety of studies designed to characterize the emotional mental state decoding abilities of adults (Baron-Cohen & Hammer, 1997; Baron-Cohen et al., 2001). Further, the task has been combined with the sex decoding task to index a special difficulty among autistic individuals in decoding emotional mental states (Baron-Cohen et al., 2001). Thus, a characterization of the neural correlates of this dissociation may help to provide evidence relevant to understanding the elusive neural correlates of autism (see Rumsey & Ernst, 2000, for a review).

The most recent version of the task was modified and simplified to make it suitable for use in an ERP paradigm, and to circumvent some of the problems that may have dogged interpretation of results in the past. In this modified version (see Figure 2), participants first saw a word labeling either a mental/emotional state (e.g., embarrassed) or a sex (e.g., female). The word was then...
replaced with a fixation point, and then a picture of eyes. Participants’ task was to judge (yes/no) whether the eyes displayed the emotion or sex labeled by the preceding word. ERPs were recorded time-locked to the onset of the picture of the eyes as participants made their decisions. Because emotional mental state and sex judgments were made with respect to the same photographs, differences in ERP activity could not be attributed to general differences between stimuli. Moreover, because the words and the pictures were not presented simultaneously, activity associated with making judgments about the eyes could not be due to additional reading or linguistic processing.

Based on our review of the current literature regarding both theory-of-mind and emotion recognition skills, we predicted that judgments about emotion would be indexed by focal ERP component differences over frontal, and possibly temporal, regions of the right hemisphere. Based on previous studies that have used ERP to index the neural correlates of theory-of-mind reasoning, we predicted that these effects would be found on ERP components that emerge at, or just prior to, 300 msec following the onset of the stimulus for which mental state judgments are to be made (Sabbagh & Taylor, 2000).

Figure 1. (a) Channel map of the geodesic sensor net (GSN) with channel numbers labeled. (b) Average position of 128 channels in 3-D space projected on a normalized cortical surface from the front and side views.

Figure 2. Schematic representation of the experimental trial structure.
RESULTS

ERP Analysis Procedure

A representative sampling of the ERPs elicited by the eyes in both conditions is provided in Figure 3. The general pattern of ERPs we observed was very similar to that described in electrophysiological studies of face perception and recognition (e.g., Bentin, Allison, Puce, Perez, & McCarthy, 1996). Over posterior regions, a P100 was followed by an N170 that was then followed by an extended positive deflection, which peaked between 300 and 500 msec (P300–500). This general morphology was reversed over anterior regions; an N100 was followed by a P200 that was then followed quickly by an extended negativity that peaked between 270 and 400 msec (N270–400). This anterior negativity was extended beyond 400 msec at inferior anterior temporal sites.

Differences in the peak amplitudes of these ERP components were characterized in two stages. First, Wilcoxon signed-rank tests were used to identify local differences in the ERP at all time points throughout the recording epoch. To this end, the tests were performed for each sample on all 128 channels ($p < .05$, 2-tailed). To avoid false positives, strict criteria were adopted whereby only differences that were (1) maintained on a single channel for 10 continuous samples (20 msec) and (2) in a cluster of at least four other channels that showed a similar pattern of activity were considered significant. Second, the condition differences that were identified as significant in the nonparametric tests were then further characterized using ANOVA ($p < .05$). Only differences that were significant in both the nonparametric and ANOVA tests will be discussed here.

N270–400

The nonparametric tests revealed that over the right hemisphere, the N270–400 component was significantly more negative in the mental state condition relative to the sex condition. The spatial and temporal characteristics of this effect are illustrated in Figure 4. As is clear from this figure, the N270–400 difference was focal to right inferior frontal (9, 10, 3, 2) and anterior temporal sites (123, 122, 121, 117, 116, 115, 110, 109). The
nonparametric tests revealed no significant differences with respect to this component over the left hemisphere.

To further characterize this effect, we calculated the mean amplitude occurring between 270 and 400 msec and averaged across a cluster of right inferior frontal and right anterior temporal sites (shown in Figure 4). A $2 \times 2 \times 2$ (Condition $\times$ Lobe $\times$ Hemisphere) repeated measures ANOVA revealed a significant main effect of condition, $F(1,17) = 7.91, p = .01, MSE = 4.32$. However, this main effect was qualified by a significant Condition $\times$ Hemisphere interaction, $F(1,17) = 4.76, p = .04, MSE = 2.44$.

Follow-up means contrasts revealed that in the right hemisphere, the N270–400 amplitude was more negative in the mental state condition ($M = -1.91 \mu V, SE = .53$) than in the sex condition ($M = - .37 \mu V, SE = .38$), $F(1,17) = 15.54, p < .001$, but not at left hemisphere sites, $F(1,17) = 1.22, ns$. Thus, the N270–400 was prominent at sites in the right hemisphere only when participants were decoding mental states from the pictures of eyes, and not when making sex judgments.

We used low-resolution electromagnetic tomography to identify the regions of the brain that were the most probable contributors of the N270–400 condition difference (LORETA-KEY: Pascual-Marqui, 1999; Pascual-Marqui, Michel, & Lehmann, 1994). Briefly, this technique parcels the brain into equal sized units (i.e., voxels) and then employs a least squares method to assign an activation value to each unit based on the distribution of electrical activity measured at the scalp. The activation values are assessed in a manner similar to assessing eigenvalues in a factor analysis. Here, the interpolated difference wave (mental states $\times$ sex) was sampled at 300 msec and submitted to the LORETA-KEY analysis. This period was selected because the N270–400 component elicited in the mental state condition peaked within 20 msec of this point for all participants. The analysis showed that the units with the highest activation values were located within (1) the right orbitofrontal cortex (approximate Talairach coordinates: $x = 4, y = 45, z = -27 \text{ mm}$) and (2) the right anterior medial temporal

**Figure 4.** (a) ERPs from inferior frontal and anterior temporal sites from the left and right hemisphere. Gray-shaded regions on the ERP indicate the time period of the N270–400. Electrode sites filled in black showed significant condition differences during the N270–400. Circled and shaded regions on the GSN map demarcate electrode groups that were combined for ANOVA analyses. (b) Three-dimensional spherical spline interpolations of the difference in scalp electrical activity (mental states $\times$ sex) measured at 300 msec poststimulus.
cortex (approximate Talairach coordinates: $x = 25, y = -4, z = -34$ mm) (see Figure 5).

**P300–500**

Nonparametric tests also showed that a condition effect emerged at posterior parietal sites on the P300–500 component of the ERP. Specifically, in contrast to directionality of the N270–400 effect, the P300–500 component was significantly greater in the mental state condition relative to the sex condition. These condition differences were centrally distributed over a broad region encompassing central and parietal sites (see Figure 6).

Because the nonparametric tests revealed P300–500 condition differences with a broad spatial distribution, we elected to explore these effects using a $2 \times 2 \times 2$ (Condition $\times$ Superiority $\times$ Hemisphere) repeated measures ANOVA. The superiority variable was introduced to characterize whether the effect was stronger towards more superior (or more central) sites relative to more inferior (or more temporal) sites. We extracted the mean amplitude occurring between 300 and 500 msec, then averaged across two groups of left parietal channels (inferior: 53, 60, 66; superior: 54, 61, 67) and homologous sites over the right hemisphere. There was a significant main effect of condition, $F(1,17) = 11.34, p = .004, MSE = 4.81$, indicating that the P300–500 component was larger in the mental state condition ($M = 4.33 \mu V, SE = .26$) relative to sex ($M = 3.10 \mu V, SE = .25$). Also, there was a significant main effect of superiority, $F(1,17) = 28.98, p < .001, MSE = 1.44$, indicating that both conditions elicited a P300–500 that was greater over superior (more central) sites ($M = 4.25 \mu V, SE = .26$) than over inferior sites ($M = 3.18 \mu V, SE = .25$). The ANOVA did not, however, reveal any significant interactions with the condition variable.

The similar timing of the N270–400 and P300–500 differences raised the possibility that the two effects were correlated, and thus possibly reflecting the same underlying neural generation. To explore this possibility, we conducted a principle components analysis (with varimax rotation) on the dependent measures that were used in the two previous ANOVAs, including the N270–400 amplitude at inferior frontal and anterior...
temporal sites and P300–500 amplitude at superior and inferior parietal sites of both hemispheres. One analysis was conducted for these dependent measures in the mental state condition, and one for the sex condition (see Table 1). For the mental state condition, the first factor included all of the parietal sites and the left inferior frontal and left anterior temporal sites. The second factor included the right inferior frontal and right anterior temporal sites only. In contrast, the factors for the sex task were more clearly distinguished by hemisphere. Two aspects of these findings are notable. First, in both conditions, the N270–400 effect was statistically dissociable from the P300–500 effect. Second, in the mental state condition, the N270–400 activity specifically associated with the right hemisphere made a distinct contribution to the overall scalp electrical activity.

**Control Analysis**

Although the sex and mental state decoding tasks are matched in terms of the ERP eliciting stimulus and the response demands of the task, the two tasks are not matched on the dimension of semantic complexity. The mental state terms were more semantically diverse and complex than the sex terms. This task difference may have made the mental state task more difficult than the sex task. Indeed, behavioral accuracy data indicated that participants were more accurate in the sex task ($M = 87.8\%$ correct) than in the mental state task ($M = 77.2\%$), paired $t (17) = 7.09, p < .05$, thereby supporting the fact that the mental state task may have been more difficult.

There are several factors that may potentially make mental state decoding more difficult than sex decoding. We will address these in the discussion. However, it is important to first rule out the possibility that the difficulty imbalance may have artifactualy contributed to our ERP condition differences by adding more error trials to the mental state relative to the sex conditions. To ensure that our main effects were not carried by error trials alone, we performed a series of analyses identical to those listed above considering only trials in which participants responded accurately. The pattern of results obtained in this analysis was identical to those described above, and thus gave confidence that the differences could not be attributed to performance on error trials. This analysis also had the benefit of equating the two conditions for accuracy. That is, after removing error trials, participants’ ERPs were based on 100% correct performance in each condition. The fact that the patterns of findings did not differ in these conditions

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mental State</th>
<th>Sex</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Component 1</td>
<td>Component 2</td>
</tr>
<tr>
<td>Left superior parietal</td>
<td>.829</td>
<td>-.378</td>
</tr>
<tr>
<td>Left inferior parietal</td>
<td>.855</td>
<td>-.071</td>
</tr>
<tr>
<td>Left anterior temporal</td>
<td>.910</td>
<td>.015</td>
</tr>
<tr>
<td>Left inferior frontal</td>
<td>.884</td>
<td>.020</td>
</tr>
<tr>
<td>Right superior parietal</td>
<td>-.817</td>
<td>-.454</td>
</tr>
<tr>
<td>Right inferior parietal</td>
<td>-.909</td>
<td>-.155</td>
</tr>
<tr>
<td>Right anterior temporal</td>
<td>.061</td>
<td>.981</td>
</tr>
<tr>
<td>Right inferior frontal</td>
<td>-.256</td>
<td>.897</td>
</tr>
</tbody>
</table>

Note: Bold values represent highest loading for variable per condition.
suggests also that our findings cannot be attributed to experimental artifacts induced by task difficulty.

These analyses provide some confidence that our ERP differences can be attributed to whatever differences characterize the cognitive processes associated with decoding mental state versus sex. We turn now to a consideration of how we can currently best interpret these differences.

**DISCUSSION**

We found two aspects of the ERP that distinguished mental state decoding from sex decoding: a frontal N270–400 and a posterior P300–500. Specifically, we found that at right hemisphere sites, the N270–400 was more negative than in the sex decoding condition. Using low-resolution electromagnetic tomography (LORETA-KEY) we found that the right orbitofrontal cortex and the right anterior medial temporal cortex were strong candidates for generators of this condition difference. In contrast, the P300–500 elicited in the mental state condition was more positive than in the sex condition.

To interpret these condition differences, it is important to consider several factors that may distinguish the mental state decoding task from the sex decoding task. First, and perhaps most relevant to the goals of our study, the two tasks differ in terms of their content. Specifically, the emotion-decoding task is designed to tap specific social-cognitive skills related to theory-of-mind reasoning. Given this fundamental difference between tasks, it is possible that some aspects of our ERP differences may be an index of the neural systems associated with mental state decoding and theory of mind per se.

A “theory-of-mind” interpretation may be appropriate for the N270–400 component. Our finding that this component was enhanced in the mental state relative to the sex condition is consistent with a recent ERP study that has shown that processing emotional faces (both happy and fearful) elicited an enhanced N270 component (Sato, Kochiyama, Yoshikawa, & Matsuura, 2001) that was focal to right-hemisphere sites. While it is difficult to directly compare the results of their study with the present one due to differences in methodology and recording techniques, the extent to which they converge provides a beginning basis for understanding the electrophysiological characteristics of the neural correlates of mental state decoding.

The spatial distribution of the N270–400 and the supporting LORETA analysis suggested that this effect may be associated with the orbitofrontal and medial temporal cortex. This possibility converges well with the emerging literature on the neural bases of everyday social understanding. Several researchers have hypothesized that each of these areas plays a crucial role in the perceptual processing of a variety of mental states including emotional states (e.g., Haxby et al., 2002; Adolphs, 1999; Brothers, 1996; Baron-Cohen et al., 1994; Baron-Cohen, 1995). Our findings extend these in an intriguing way by suggesting that activity in orbitofrontal and medial temporal areas might be sensitive to the processing goals of the individual. Specifically, we indexed N270–400 activity differences over orbitofrontal and medial temporal regions when participants were cued to make judgments about mental states as compared with judgments about sex—even though the stimuli in both conditions were identical. These findings suggest that there may be an interesting relation between emotion perception and emotion decoding that warrants further research. Nonetheless, when considered in light of previous research, it seems reasonable to suggest that the N270–400 reflects the neural processes associated with the emotional decoding differences between conditions.

Although the present study was designed to index the dissociation of neural activity responsible for decoding mental states versus sex, our behavioral data revealed that items in the mental state condition were more difficult than in the sex condition. This behavioral difference points to the fact that there are likely additional ways in which the tasks differed that may have contributed to our ERP effects. One way in which the mental state and sex conditions might be distinguished concerns their relative visual information processing requirements. Although little is known about the specific kinds of information that are required for making judgments about emotional states relative to sex, it seems possible that emotional state judgments may require somewhat more subtle or variable information that is harder to obtain from the eye stimulus.

Intriguingly, this possibility may provide a compelling account for the P300–500 differences obtained in the present study. Recent research has shown that P300–500 amplitude is enhanced when the same stimulus is subject to increasingly difficult visual discrimination tasks (Senkowski & Hermann, 2002). In the present context, the enhanced P300–500 elicited by the eyes in the mental state condition may reflect the increased difficulty of the visual discrimination required. Of course, the present study was not designed to test this hypothesis. Thus, future research is required to gain confidence about the kinds of visual information that are necessary to decode different kinds of information from the eyes (including mental state and sex).

An additional way in which the mental state and sex conditions might be imbalanced is with respect to the relative complexity of the semantic judgments required. The mental state terms were more diverse and are likely to possess more semantic features than the sex terms, of which there were only two (i.e., male and female). Interestingly, this hypothesis may provide another reasonable account of the P300–500 condition differences. That is, the P300–500 may reflect the additional semantic complexity associated with mental state relative to sex concepts. Some indirect support for this possibility...
comes from our factor analysis showing that P300–500 amplitude was related to activity at left frontal and temporal sites—sites known to index ERP correlates of language processing (e.g., Neville, Mills, & Lawson, 1992). Again, however, the present study was not designed to test this speculation and, thus, future research is required to better understand the role of semantic processing generally in mental state decoding.

We have argued that the difficulty differences between experimental conditions may be related to cognitive factors that may well be reflected in the P300–500 condition differences in the ERP. A lingering question, however, is whether the N270–400 may also reflect some of these difficulty imbalances. Unfortunately, this is not a possibility that we can rule out entirely. However, it is noteworthy that the P300–500 is a common index of task difficulty in the ERP literature (see Castro & Diaz, 2001), and our factor analysis showed that the right frontal and anterior temporal regions that showed the N270–400 effect were statistically separable from the P300–500 activity seen at parietal sites. These arguments, along with the supporting literature mentioned above, lead us to suggest that the N270–400 effects reflect the neural correlates of mental state versus sex decoding per se, while the P300–500 effects may reflect peripheral cognitive processing that supports mental state decoding in the present task.

**Links with Autism**

At the outset, we noted that the use of the present task was motivated, in part, by the fact that it has been used to index a relatively specific deficit in decoding mental states in individuals with autism. That is, these individuals are fine at decoding sex and poor at decoding emotional states (Baron-Cohen et al., 2001). We hope, then, that the electrophysiological indices of the mental state/sex decoding dissociation may bring a unique kind of evidence to bear on our understanding of the neuropsychological functioning of individuals with autism.

Intriguingly, our findings regarding the N270–400 and their associated neural generators dovetails very nicely with recent research conducted by Dawson, Osterling, Rinaldi, Carver, and McPartland (2001) indicating that ventral–medial prefrontal functioning may be impaired in autism. Our findings suggesting that these regions may be important for decoding emotional states provide a link between this known cognitive impairment in autism and known neuropsychological characteristics.

Our findings may also raise a warning flag for interpreting the “specificity” of the deficit shown in autism. Our findings show that the N270–400 and its associated neural generators (including orbitofrontal regions) was not alone in distinguishing mental state decoding from sex decoding in the current context. Although when taken together with previous findings, it seems reasonable to suggest that the N270–400 is associated with the emotion decoding differences between tasks, it seems likely that these two tasks differ on additional dimensions (e.g., visual-information processing, linguistic complexity). These additional differences limit the extent to which we can conclude that the task indexes a specific emotional decoding deficit in autism. Future research attempting to characterize the neural bases of everyday mental state reasoning in both typically developing and autistic individuals must address these concerns directly.

**Implications for Understanding the Neural Bases of Theory of Mind**

Our results may be helpful in integrating disparate theoretical positions regarding the neural bases of theory of mind. In particular, Frith and Frith (1999) performed an integrative review of several findings that were broadly related to theory-of-mind reasoning. In this work, they suggested that medial frontal areas just anterior to the anterior cingulate cortex are the most consistently activated in theory-of-mind tasks. In contrast, Brothers (1996) and Baron-Cohen (1995) have argued that medial temporal and orbitofrontal regions of the right hemisphere are particularly important for understanding others’ mental states. Our contention is that both of these regions are crucial, although their contributions to the task of everyday theory-of-mind reasoning differ in important ways. Specifically, orbitofrontal and medial temporal regions may contribute specifically to mental state decoding while left medial frontal regions are important for reasoning about mental states. By better understanding these relative contributions, we will gain better understanding of the neural underpinnings of social cognition.

**METHODS**

**Participants**

Eighteen adults (5 men, 13 women) volunteered to participate in this study either for pay (5 people) or for credit towards their mark in an undergraduate psychology class. Participants were all of European–Canadian descent. All participants were right-handed, had normal or corrected-to-normal vision, and reported no history of organic or acquired brain damage. An additional 3 individuals participated but their data were excluded due to excessive electrophysiological artifact during recording (2) or to equipment failure (1).

**Materials**

**Stimuli**

The stimuli sets were created based on the “Reading the Mind in the Eyes” test (revised: Baron-Cohen et al., 2001) that was designed to provide a sensitive test of emotion recognition abilities in high-functioning autistic...
individuals and populations that share aspects of the broader autistic phenotype (see Baron-Cohen & Hammer, 1997). The test consists of gray-scale photographs of the eye region of human faces (brow to the bridge of the nose). In all photographs, the eye region was clearly visible. Upon viewing these photographs, participants are asked to select one of four presented adjectives that best describes the emotion portrayed in the face. The target adjectives are summarized in Table 1, along with the results from a pilot study in which participants rated the emotional valence of each eye picture and target adjective.

To adapt this test for use in an ERP paradigm, the photos were digitized using a flatbed scanner, and resized to a standard screen size of 14.5 (width) × 5.5 (height) cm. The stimuli were presented on a 15-in. flat-screen LCD panel monitor using a PC running E-prime (Psychology Software Tools, Pittsburgh, PA). The stimuli were balanced for luminance, presented foveally at a distance of approximately 60 cm, and subtended a visual angle of approximately 13.78°.

Response Collection

Participants’ responses were made via a response pad connected to the PC that presented the stimuli. The response pad had a dedicated timing mechanism that ensured accuracy within 2 msec. The speed and accuracy of participants’ responses were recorded digitally by the E-prime software.

Electrophysiological Recording

Electrophysiological data were recorded from the scalp using a 128-channel geodesic sensor net (Tucker, 1993), a network of 128 Ag/AgCl sponge sensors knitted into an elastic geodesic tension structure (see Figure 1). The sensor net has an even interelectrode distance of 2.7 cm. The EEG was recorded continuously throughout the experimental session. During the session, all recordings were referenced to the vertex electrode (Cz), sampled at 500 Hz, and digitally filtered between 0.01 and 200 Hz (time constant = 1 sec.). The continuous data were then segmented to the onset of the computerized presentation of the eye pictures. These single-trial data were edited with a software algorithmic artifact rejection program that combed the data for evidence of lateral eye movement, eye blink, and muscle artifacts. Trials that included artifact were not included in further analyses. Artifact-free trials were averaged to create the ERP for each subject. For analysis, ERPs transformed to average reference, corrected to a 100-msec baseline, and digitally filtered (low-pass Butterworth 30 Hz) to reduce environmental noise.

Participants were tested alone in a sound-attenuated room while seated in a comfortable chair. An adjustable chin rest ensured that participants’ distance from the flat-screen monitor remained constant and minimized head movement.

Procedure and Design

Net Placement

Upon providing consent, a researcher measured the circumference and located the vertex (Cz) of participants’ heads to determine the size of the sensor net to be used. The sensor net was then applied and adjusted to ensure correct placement and good contact with the scalp. The impedances of all 128 electrodes were under 30 kΩ at the beginning of the recording session (Ferree, Luu, Russell, & Tucker, 2001). Placement and adjustment of the sensor net took approximately 10–15 min for each participant.

Experimental Procedure

Once the sensor net was satisfactorily applied and adjusted, participants began the procedure. Instructions that outlined the task and the structure of each trial were presented to participants at the outset. Participants received no further coaching or feedback about their responses in the course of the experiment.

Experimental trials were presented for two within-subject conditions: mental/emotional state and sex. Trials from each of these conditions had the same structure and were mixed together randomly for the entire experiment. The trial began with a visually presented word referring either to an emotion (e.g., embarrassed) or a sex (e.g., male). The word remained on the screen for 2 sec, and was replaced with a fixation point for 1 sec. Participants were then shown a picture of a pair of eyes that remained on the screen for 2 sec. Participants’ task was to determine whether the eyes accurately depicted the emotion or sex referred to by the previously presented word. The eyes were replaced by another fixation point (1 sec), which was followed by a prompt to respond via the response pad (yes/no). Thus, participants’ responses were delayed by 3 sec following the presentation of the eyes.

After making their response, participants were asked to press a key when they were ready to continue to the next trial. Participants were instructed beforehand that they could use this prompt as an opportunity to take a break from the procedure.

In total, participants completed 108 trials. There were 72 trials in the emotion condition, two for each of the 36 pictures of the eyes. On half of the emotion trials, the eyes matched the previously presented word, and so the correct response from participants was yes. There were 36 sex trials, one for each of the 36 pictures of eyes. Again, on half of the trials the eyes matched the previously presented word. Trials were presented randomly with no constraints. Preliminary analyses revealed that there were no differences in the ERPs elicited by eyes.
that matched versus did not match preceding words (i.e., “yes” vs. “no” trials). Thus, analyses were conducted collapsing across trial type. Completing all 108 trials took approximately 25–30 min.

Acknowledgments

We thank Simon Baron-Cohen for generously providing us with stimulus materials. This research is supported by a grant from the National Sciences and Engineering Research Council of Canada, and by the Canada Foundation for Innovation.

Reprint requests should be sent to M. Sabbagh, Queen’s University, Psychology Department, Kingston ON, Canada K7L 3N6, or via e-mail: sabbagh@psyc.queensu.ca.

REFERENCES


