Neural bases of eye and gaze processing: The core of social cognition

Roxane J. Itiera,∗ and Magali Battyb,1

aPsychology Department, University of Waterloo, 200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada
bImagerie et Cerveau, Inserm U930, CNRS FRE 2448, Université François Rabelais de Tours, CHRU, Tours, France

Abstract

Eyes and gaze are very important stimuli for human social interactions. Recent studies suggest that impairments in recognizing face identity, facial emotions or in inferring attention and intentions of others could be linked to difficulties in extracting the relevant information from the eye region including gaze direction. In this review, we address the central role of eyes and gaze in social cognition. We start with behavioral data demonstrating the importance of the eye region and the impact of gaze on the most significant aspects of face processing. We review neuropsychological cases and data from various imaging techniques such as fMRI/PET and ERP/MEG, in an attempt to best describe the spatio-temporal networks underlying these processes. The existence of a neuronal eye detector mechanism is discussed as well as the links between eye gaze and social cognition impairments in autism. We suggest impairments in processing eyes and gaze may represent a core deficiency in several other brain pathologies and may be central to abnormal social cognition.

Keywords

Eyes; Gaze; Face; Social cognition; Theory of mind; Neuroimaging; ERPs; MEG

1. Introduction

The human face is arguably the most important visual stimulus we process everyday as it informs us how to behave socially: being able to discriminate whether the person coming at you is your friend or your boss and whether he looks angry or joyful will certainly make a difference in how you interact with him. The eye region of the face represents a special area due to the extensive amount of information that can be extracted from it. You can perceive your boss’s fake smile by the absence of wrinkles around the eyes while a friend’s averted gaze can inform you something is wrong. More than other facial features, the eyes are central to all aspects of social communication such as emotions, direction of attention and identity. The field of Cognitive and Behavioral Neuroscience has recently witnessed an explosion of studies investigating the processing of the eye region and gaze direction in various tasks and social situations but due to their extensive complexity, the underlying neural systems subtending these processes are far from being understood. The central role of
eyes and gaze in social cognition and the state of knowledge of the neural networks involved in perceiving these fundamental social cues are the topics of the present review.

The eye region is special because it plays a fundamental role in social and non-verbal communication: it is necessary for proper identity and emotional processing and indicates the direction of attention of others and their potential targets for intentions. The ocular muscles enable a very efficient mobility of the eyes which are constantly exploring our visual environment, focusing on regions and objects of interest in order to extract relevant information. This gaze movement is necessary for visual perception but also reveals to external observers where and what we are looking at. Someone else’s gaze thus informs us about his/her object of interest. The human brain has developed a very complex cognitive system of gaze direction analysis based on perceptual elements of faces and eyes. For instance, the contrast difference between the iris and the white sclera allows for the discrimination of gaze direction. If the dark iris is situated in the center of the sclera then the gaze is straight and there is eye contact (also called direct or mutual gaze). If it is turned to the left then the person is looking to the left, etc. Through evolution, morphological changes of the hominid face have occurred in parallel to brain changes linked to the emergence of social cognition (Emery, 2000). The reduction in face protrusion, the salience of cheekbones and the shape of the nose or eyebrows, all underline the position of the eyes within the face. Similarly, the great variety of facial muscles, especially those around the arches of the brows and those controlling the eye motility, allow for a large range of subtle facial expressions. The human eye also possesses the largest ratio of exposed sclera size in the eye outline of all species, allowing a better gaze direction discrimination even at a distance (Kobayashi and Kohshima, 1997). This characteristic is very pertinent to the detection of emotions such as fear (Whalen et al., 2004) and the potential imminent threat that it implies. It also represents a big advantage for survival. All these morphological characteristics give a major role to eyes and gaze in complex forms of social cognition.

Throughout this review, which focuses on humans (see Emery, 2000, for details in non-human primates and other species), we will distinguish the eye region, i.e. the facial features, from eye gaze, although both are intimately linked at a cognitive and neural level. We first review the importance of eyes and gaze in identity and emotion perception (Section 2) and then turn to fundamental aspects of social cognition linked to gaze (Section 3). Other recent reviews on gaze processing have been reported in the literature and the reader is referred to these excellent papers for more extensive details on some of the topics that we will only briefly describe here (Emery, 2000; Frischen et al., 2007; George and Conty, 2008; Kleinke, 1986). The emphasis of the present review is centered on the neural bases of all these processes detailed in Section 4 which is articulated around (i) reported neuropsychological cases of eye and gaze processing abnormalities, (ii) neuroimaging studies and (iii) temporal aspects of eyes and gaze processing as measured by electro- and magneto-encephalography techniques. We also briefly examine how the processing of eyes and gaze and their abnormal neural correlates seen in autism can be relevant to the understanding of this severe developmental pathology (Section 5). In these various sections, we will underline the inconsistencies reported in the literature and emphasize some of the exciting yet unanswered questions that should drive the field in the coming years.

2. Central role of eyes and gaze in face processing

2.1. Eyes are central to various aspects of face processing

Studies monitoring ocular movements during face perception have shown that people spend more time on internal features (eyes, nose, mouth) than external ones (hair, face contour, forehead, ears) (Althoff and Cohen, 1999; Walker-Smith et al., 1977; Yarbus, 1967). Such findings support the idea that internal features play a central role in face perception and
recognition, especially for familiar faces (Ellis et al., 1979). Numerous studies have shown that the eye region is the most attended of all facial features and the source of information the most utilized regardless of the task, whether it focuses on gaze, head orientation, identity, gender, facial expression or age (e.g. Henderson et al., 2005; Itier et al., 2007c; Janik et al., 1978; Laughery et al., 1971; Luria and Strauss, 1978; Schyns et al., 2002). This attraction to the eyes is even more pronounced for familiar faces (Althoff and Cohen, 1999).

Like faces, eyes vary greatly from one individual to another and the eye region may in fact be the facial zone that varies mostly between people. This inter-subject variability has been investigated by many morphological and biometrical studies, especially in the field of Anthropology (Farkas, 1994; Hall et al., 1989). Eye color and shape, and inter-ocular, inter-canthal and inter-pupillar distances are specific to each individual. Other elements of the eye region such as eyebrows, eyelids, eyelashes and their respective distances constitute numerous configural cues necessary for the recognition of face identity and for this reason are used in robot-portraits, the computerized drawing representations of individual faces obtained from descriptions of their various features. For instance, 13 measures of the eye region are utilized in forensic identification in which the face of missing people is reconstructed based on old photographs or even skulls (Farkas, 1994; Farkas et al., 1994). When the eye region is masked, face recognition performances drop while masking the nose or the mouth has little or no effect (McKelvie, 1976). Familiar face recognition performances drop even more so when eyebrows, rather than the eyes, are removed from the picture (Sadr et al., 2003). Similarly, face detection is disproportionately impaired when the eye region is occluded compared to when the nose, mouth or forehead are occluded (Lewis and Edmonds, 2003). Image classification techniques have also shown that the eye region is the diagnostic feature used to discriminate gender (Schyns et al., 2002; Vinette et al., 2004) and to recognize identity (Caldara et al., 2005), i.e. it is the principal element subjects use to decide whether a face is male or female or who it is. When noise is added to the picture, identity discrimination between two faces is performed using the eye region including eyebrows (Sekuler et al., 2004). The eye region is thus a key element of face recognition.

In addition to its important role in processing identity, the eye region carries information necessary for emotion recognition (Calder et al., 2000; Ekman and Friesen, 1978; Fox and Damjanovic, 2006; Smith et al., 2005) and is thus central to non-verbal communication. Fear and surprise are characterized by wide open eyes and by a larger white sclera size (Whalen et al., 2004) and masking the eye region results in a drop of fear recognition performances (Adolphs et al., 2005). The inferior eyelid is contracted when the person is expressing fear but relaxed when expressing surprise (Ekman and Friesen, 1978). In faces expressing disgust however, the eyes are squinted. Joy is mostly characterized by the inferior part of the face, especially by the mouth (smile) which is the diagnostic element enabling the recognition of this emotion (Schyns et al., 2002). However, a fake smile will be betrayed by the absence of expressive cues in the eye region, such as the wrinkles around the eye corners or the squinting of the eye opening by the ocular muscles (the so-called “Duchenne smile”, Duchenne, 1990; Ekman, 1992). A recent study showed that, although not affecting accuracy rates, ocular cues influence the speed of joy recognition but this depends on the task context (Leppänen and Hietanen, 2007). Anger is implied by the frowning of the eyebrows and other eye cues (Calder et al., 2000; Smith et al., 2005), and sadness by a down-looking gaze (Ekman and Friesen, 1978). Thus, all six basic emotions described by Ekman (joy, fear, anger, sadness, surprise and disgust, Ekman and Friesen, 1971) involve a specific change in one element of the eye region. Even though, as measured by one image classification technique, the eye region is the diagnostic element utilized to recognize fear (Schyns et al., 2007), the scanning of the face always starts with the eyes regardless of its emotion (Schyns et al., 2007; Vinette et al., 2004), again supporting a role of the eye region in processing all facial expressions. Finally, isolated eye regions are often sufficient to
recognize the six basic emotions but also more complex feelings such as jealousy, envy or guilt (Baron-Cohen et al., 1997b, 2001a). The eyes, a “window to the soul”, inform us on the emotion and the state of mind of others, a topic we return to in Section 3.2.

The eye region thus attracts attention and represents a special area of the face from which extensive amount of information can be extracted, such as identity and emotion cues. We now turn to another important piece of information derived from the eye region, gaze direction. Before dealing with the role of gaze direction in attention orienting (Section 3.1), we first describe below the influence of gaze on various processes such as the perception of gender, identity and facial emotions.

2.2. Gaze direction in face perception, gender discrimination, identity recognition and facial expression discrimination

The systematic attraction of attention towards the eyes of a face reviewed above, starts early in development (Maurer, 1985) and seems linked to the perception of gaze. Newborns prefer to look at a face with open rather than closed eyes (Batki et al., 2000) and look longer at faces whose gaze is directed at them compared to averted-gaze faces (Farroni et al., 2002). Three-month-old infants also smile less when an interacting person gazes away after having made eye contact (Hains and Muir, 1996). If gaze contact is often perceived as a threat in most species (Emery, 2000), its meaning has evolved in humans and the early attraction of newborns to direct gaze is linked to social communication (Farroni et al., 2002). In everyday life, a direct gaze signals a potential social interaction (positive or negative) while an averted gaze implies that the person is attending to something or someone else than us. For instance, a recent study showed that direct gaze is related to approach behavior while averted gaze is related to avoidance, but this is found only when real persons are part of the study design rather than pictures of faces (Hietanen et al., 2008b). In humans, the power of gaze is due to its social impact (Kleinke, 1986).

Because of the spontaneous attraction of attention to the eye region, it seems reasonable to believe that information in this area, and in particular gaze direction, could influence other processes related to faces. Several behavioral studies have shown that gaze direction can influence person perception and recognition but findings are inconsistent. For instance, gender categorization was faster for direct than averted-gaze faces in one study (Macrae et al., 2002) while the opposite was found in another study but only when the face was in 3/4-view (Vuilleumier et al., 2005). This gaze effect on gender categorization was also more pronounced when the face was of opposite gender to that of the subject (Vuilleumier et al., 2005). This difference in the two studies could be due to the stimuli used. Vuilleumier et al. (2005) used a fully counterbalanced design between head orientation (front- and 3/4-view) and gaze direction (direct or averted) while Macrae et al. (2002) used 3/4- and front-view faces with direct gaze but only 3/4-view-faces with averted gaze.

In contrast, using the same type of stimuli as Vuilleumier et al. (2005) but an explicit gaze direction judgment task, one study reported faster response times only for direct-gaze-front-view faces (Pageler et al., 2003). More recently, using an explicit gaze direction judgment and a head orientation judgment along with the same kind of stimuli, Itier and colleagues reported a true interaction between gaze direction and head orientation, with faster response times for congruent conditions and longer response times for incongruent conditions when face orientation and gaze direction did not match (Itier et al., 2007a,c). These behavioral results were found on two different subject groups and replicate previous findings (Langton, 2000). The reasons for these inconsistent results in the literature concerning gaze direction and head orientation interactions are still unclear but could be due, in addition to task effects, to the use of slightly different paradigms adapted to the different methodologies used (e.g. ERPs and eye tracking in Itier et al., 2007a,c; fMRI in Pageler et al., 2003; strictly
behavioral tests in Vuilleumier et al., 2005 and in Langton, 2000). In another recent gaze direction discrimination task with similar stimuli, subjects were faster to detect eyes moving towards the viewer (direct gaze motion) rather than moving away from the viewer (averted gaze motion) and this was found regardless of head orientation, although the effect was more pronounced for front-view faces (Conty et al., 2007). In contrast to previous studies which used static faces, this paradigm involved gaze motion and this may explain the difference in results.

More consistent findings have been reported concerning the influence of gaze direction on identity encoding and recognition. It has been shown that direct-gaze faces are better encoded (Mason et al., 2004) and better recognized than averted-gaze faces, especially for 3/4-view faces of opposite gender to the subject (Vuilleumier et al., 2005). This effect of direct gaze on face recognition has been reported in newborns (Hood et al., 2003), 4-month-old infants (Farroni et al., 2007) and children from 6 to 11 years of age (Smith et al., 2006) for front-view faces and will have to be tested with 3/4-view faces. In one study, subjects were also faster to categorize letter strings as words when they were primed by a direct-gaze compared to an averted-gaze face regardless of head orientation (Macrae et al., 2002). The authors concluded that direct gaze facilitates the access to semantic information concerning a person (Macrae et al., 2002), although that study used unfamiliar faces and common words that were not specifically descriptive of people. Conversely, familiarity can impact on gaze processing as suggested by one study in which a recently encountered face (that thus became recently familiar) biased the perception of downward gaze direction which was perceived as more directed towards the subject compared to when the face was unfamiliar (Teske, 1988). Overall, it seems that gaze direction influences face categorization and recognition processes (and gaze perception may also be influenced by factors such as familiarity) but no clear effect of direct or averted gaze can yet be drawn from all these studies given the variability in task and stimuli used.

Gaze direction also seems to modulate the perception and the understanding of someone’s emotion. A special role for direct gaze in communicating increased emotional intensity regardless of the emotion was suggested in earlier work (Kleinke, 1986) while more recent work argues in favor of a differential role of direct and averted gaze depending on the emotion. For instance, one study reported that angry and joyful faces with direct gaze were categorized more quickly and more accurately than angry and joyful faces with averted gaze while the opposite was found for fearful and sad faces which were categorized faster and better when gaze was averted rather than direct (Adams and Kleck, 2003). In another study by the same authors, anger and joy were also perceived as more intense in direct-gaze compared to averted-gaze faces while fear and sadness were perceived as more intense for averted-gaze faces (Adams and Kleck, 2005). The authors suggested that direct gaze enhances the perception of approach-oriented emotions (anger and joy) while averted gaze enhances the perception of avoidance-oriented emotions (fear and sadness). Alternatively, these faster responses and increased perceived emotional intensity may be explained in terms of increased threat-related levels: an angry face looking straight at you may imply a coming danger, as you may be the object of the anger while a fearful face looking to the side may signal a coming danger on that side which you need to detect fast. However, this threat-related account would not explain the results found for joy and sadness. An effect of direct gaze on the perception of anger has also been reported in 4-month-old infants while the perception of other facial emotions does not seem to be modulated by gaze direction that early in development (Striano et al., 2006).

However, other studies failed to reproduce the original results of Adams and Kleck (Bindemann et al., 2008; Graham and LaBar, 2007). Bindemann et al. (2008) showed that emotion categorization varied as a function of the number of facial expressions included in a
given paradigm and suggested that the results of Adams and Kleck likely reflected strategic task effects rather than real effects of gaze on facial expression categorization (Bindemann et al., 2008). Graham and LaBar (2007) also showed that gaze direction modulates expression processing only when facial expressions are difficult to discriminate (i.e. more ambiguous). These inconsistencies remain so far unclear but future studies will have to compare all facial expressions in each paradigm, including surprise which is rarely involved in gaze and emotion paradigms. We also want to point out that these studies used only front-view faces and that in addition to task context and baseline “discriminability” of facial emotions, face orientation could also influence facial expression categorization just like it influences other categorization tasks as reviewed above and this will also have to be investigated in the future.

Gaze also influences the perception of face attractiveness: faces with direct gaze are rated as more attractive than faces with averted gaze (Strick et al., 2008), and some have suggested a reward effect of direct gaze when the face is attractive but not when it is unattractive (Kampe et al., 2001). Furthermore, gaze direction influences the affective appraisals of objects in the environment as objects that are looked at by other people are liked more than objects that do not receive others’ attention (Bayliss et al., 2006). The affective appraisal of objects is further influenced by the combination of others’ gaze direction and their facial expression: objects are judged more pleasant when the person who is looking at them looks joyful rather than disgusted (Bayliss et al., 2007). However, these studies by Bayliss and colleagues used cuing paradigms (see Section 3) and these interaction effects between gaze, facial emotions and object appraisal should be verified in non-cuing studies. One study also reported that objects were evaluated more positively when they were associated with direct-gaze attractive faces but not with averted-gaze attractive faces (Strick et al., 2008). Finally, the association of others’ gaze direction with their facial expressions constitutes a social cue that seems to be utilized very early in development. As early as 1 year of age, infants use this information to judge whether they can approach and manipulate an object or whether it represents a danger and should be avoided (Mumme and Fernald, 2003).

The inconsistencies reported in the literature preclude any firm conclusion about the specific impact of direct and averted gaze on various cognitive processes related to faces. Factors like head orientation, gender of the subject and of the stimuli and task context seem to interact in complex ways with gaze direction. Finally, the use of photographs rather than real persons is another important factor that modulates the impact of gaze on face processing (Hietanen et al., 2008b) and the understanding of these complex interactions represents a challenge for future studies.

3. Basic aspects of social cognition linked to eye gaze processing

3.1. Orienting of attention by gaze

Another communicative function of eyes is to direct attention on specific places and objects of the environment through gaze. If someone is looking directly at us then we are the object of their attention. Direct or mutual gaze is a prerequisite to social interactions. In contrast, when the gaze of someone is averted to another direction than towards oneself, it informs us that we are not the object of interest and that the person is attending to something or someone else (Baron-Cohen, 1995; Baron-Cohen et al., 1997b; Emery, 2000) and we then usually turn our attention towards this object.

Averted gaze is frequently used in attention paradigms to demonstrate orienting of attention by gaze (Fig. 1). It has been shown that when a face is centrally presented prior to the onset of a lateral target, target detection is faster when the gaze of the face is directed towards the side where the target later appears and longer when the gaze is looking in the opposite
direction (Driver et al., 1999; Friesen and Kingstone, 1998; Hietanen, 1999; Langton and Bruce, 1999; Vuilleumier, 2002; for a very detailed review, see Frischen et al., 2007). This robust effect is seen as early as 3 months of age (Farroni et al., 2000; Hood et al., 1998) and is present even for simple schematic drawings of faces (Friesen and Kingstone, 1998). The fact that this effect is fast (within 200 ms of gaze shift) and occurs when gaze direction is not predictive or even counter-predictive of target location, has been interpreted as reflecting an automatic, reflexive and stimulus-driven (exogenous) orienting of attention mechanism which is impossible to suppress (Driver et al., 1999; Friesen and Kingstone, 1998; Hietanen, 1999; Langton and Bruce, 1999; Mansfield et al., 2003; Ricciardelli et al., 2002).

The magnitude of this orienting effect is similar regardless of the identity of the cue, whether it is a human face, an animal face (e.g. ape or tiger), an object such as an apple or a glove with eyes (Quadflieg et al., 2004), or an inverted face (Tipples, 2005). This suggests that any stimulus possessing eye-like attributes can trigger spatial orienting of attention (but see Langton and Bruce, 1999, for different results with face inversion). Although some studies have reported a similar orientation of attention by simple arrows (Ristic et al., 2002; Tipples, 2002), arguing against the uniqueness of gaze in triggering the orienting effect, a direct comparison of a glove with eyes or with arrows in place of the eyes showed overall faster reaction times to the target for the glove-with-eyes, suggesting gaze cues may exert quantitative rather than qualitative effects on spatial attention, and may reflect different underlying neural networks (Friesen et al., 2004; Quadflieg et al., 2004; but see Hietanen et al., 2006 for faster response times for arrows than gaze cues). Friesen and colleagues further showed that arrows produced only a volitional orienting effect rather than an automatic shift of attention as seen with gaze, further supporting a different neural mechanism for orienting attention by gaze or arrow cues (Friesen et al., 2004). Gaze orienting is also accompanied by small ocular saccades executed in the direction signaled by gaze rather than by the target (Mansfield et al., 2003; Ricciardelli et al., 2002). In an ecological and evolutionist perspective, such an automatic mechanism allows for the fast detection of a potential danger and is thus crucial for species survival.

Some have suggested that the apparent reflexive nature of this effect could be due to the paradigm used, which is always a Posner-like attentional cuing task (Posner, 1987). For instance, one study measured subjects’ eye movements in addition to their response times in two different non-cuing tasks involving the same front-view and 3/4-view faces with averted and direct gaze (Itier et al., 2007c). A purely reflexive orienting mechanism to gaze would predict that, even in these non-cuing tasks, subjects’ gaze would always be attracted by the eye region of the face and would go in the direction signaled by the perceived gaze as shown by eye movement monitoring studies (Mansfield et al., 2003; Ricciardelli et al., 2002). In contrast, the first saccade made by subjects after face presentation was directed to the eye region in 90% of trials when the task required an explicit gaze direction judgment, but in only 50% of trials when the task was a face orientation discrimination, reflecting task modulations of the attraction of attention to the eye region (Itier et al., 2007c). Moreover, these first saccades were performed in the direction signaled by gaze only in the gaze judgment, while they were performed in the direction signaled by head in the head orientation task. This study suggests that orientation of attention towards gaze direction can be modulated by task demands and hence may not be a truly reflexive mechanism (Itier et al., 2007c). The importance of these top-down modulations has recently been demonstrated in a study involving Western Caucasian and East Asian subjects. While the Caucasians fixated more on the internal features of faces and especially on the eyes (Section 2), the East Asian participants tended to look in the center of the faces (Blais et al., 2008). This study demonstrates the effect of culture on face perception and shows that the eye region does not always attract attention. It will be necessary to compare the gaze orienting effects between various populations.
Furthermore, most gaze orienting studies used faces in front-view (e.g. Driver et al., 1999; Friesen and Kingstone, 1998; Friesen et al., 2005; Mansfield et al., 2003; Ricciardelli et al., 2002). However, head orientation may modulate the gaze cuing effect just like it modulates the influence of gaze in various categorization judgments (see Section 2). To our knowledge, only one study so far involved 3/4-view faces in a gaze cuing paradigm (Hietanen, 1999) and found no cuing effect for 3/4-view faces gazing at the side target, reinforcing the idea of an interaction between head orientation and gaze direction. The author interpreted these data as reflecting that the attention orienting system was not using information of others’ gaze direction in reference to the observer but rather in reference to the others’ head orientation.

The fact that gaze cuing is modulated by head orientation (Hietanen, 1999) and that the direction of saccades landing in the eye region is modulated by task (Itier et al., 2007c), suggests top-down influences on the gaze orienting effect which may not be truly reflexive. In a natural environment, eyes alone are rarely the unique source of information on others’ direction of attention. Head orientation, body position and pointing behavior are other important directional cues that are widely used to identify the source of interest of a peer when the eyes are not visible (Emery, 2000; Langton et al., 2000). It would make sense that we could orient towards these cues in a voluntary way depending on the information we need to extract in a given situation rather than in a reflexive way. This idea has been recently supported by a series of attention cuing studies using conflicting simultaneous gaze and arrow cues while subjects attended to one cue and ignored the other (distractor). Nummenmaa and Hietanen (in press) hence showed that gaze and arrows produced similar distracting effects. In other words, gaze distractors did not exert a greater influence on orienting of attention than arrow distractors, arguing against a real automatic mechanism for gaze. Rather, the authors suggested that the attentional systems processed conflicting directional information in a flexible manner. Altogether, these studies argue against the purely reflexive nature of attention orienting by eye gaze, a topic we return to in Section 4.1 on Neuropsychology.

Gaze cuing paradigms have also used direct-gaze rather than averted-gaze stimuli. When the gaze of the face is direct, longer reaction times for target detection have been reported (Senju and Hasegawa, 2005), which are even longer than when gaze is directed to the opposite direction as the target (Vuilleumier, 2002). By attracting attention, direct gaze would increase the disengagement time from the central face before orienting attention to the target location. This attention grabbing effect of direct gaze is also supported by visual search paradigms where staring eyes embedded in an array of averted-eyes stimuli are detected better and faster than averted eyes in arrays of staring eyes (Doi and Ueda, 2007; Senju, 2005; von Grünau and Anston, 1995). This “stare-in-a-crowd-effect” was confirmed in a recent study using more naturalistic stimuli, but only when the faces from which the eye regions were extracted were in 3/4-view (Conty et al., 2006). The use of photographs rather than schematic drawings of eye regions (as in von Grünau and Anston, 1995) is likely the main reason for this difference with previous studies. Overall, these studies support the idea of a specific role of mutual gaze in human cognition.

Finally, these cuing paradigms have also been used with emotional faces. In anxious individuals the orienting-to-gaze behavior is even more pronounced for stimuli representing potential threat such as angry (Holmes et al., 2006) or fearful faces (Fox et al., 2007; Mathews et al., 2003; Putman et al., 2006; Tipples, 2006) compared to neutral ones. In contrast to previous studies which manipulated only two expressions at a time, Fox et al. (2007) used a design where neutral, angry, fearful and happy expressions were intermixed and showed that, in anxious individuals only, the gaze cuing effect was greatly enhanced for fearful faces but also reduced for angry faces compared to neutral and happy expressions. Moreover, the magnitude of the effect was correlated with the anxiety level, positively for
fear expression and negatively for angry expression. This study suggests a specific interaction of averted gaze and fear expression rather than with negative emotions in general, and supports the threat-related ecological interpretation of this effect (Fox et al., 2007). For non-anxious individuals, most studies failed to report modulations of the gaze orienting effect with facial expressions (Fox et al., 2007; Hietanen and Leppänen, 2003; Holmes et al., 2006; Mathews et al., 2003; Tipples, 2006). Using dynamic emotional faces (short movies rather than static pictures), one study reported an increased gaze cuing effect for fearful expressions in non-anxious individuals (Putman et al., 2006), and the difference with previous studies may be due to the use of more ecologically valid (dynamic) stimuli. Whether increased cuing effects could also be seen in normal individuals with anger or surprise dynamic expressions, which were not tested in Putman et al. (2006), will need to be addressed in a paradigm including all facial expressions, as used by Fox et al. (2007). Finally, given the important role of the eye region in facial expressions, future studies will have to determine whether the effects of facial expression on target detection in gaze cuing paradigms are due to the expression \textit{per se} or rather to low-level ocular cues such as the amount of visible white sclera or the size of the eye openness which vary with a given expression (Section 2).

### 3.2. Gaze perception, joint attention and theory of mind

As we already mentioned, the direction of someone else’s gaze typically signifies where his or her attention is being directed. In a triadic relationship involving two persons (A and B) and one object, the gaze direction of B will inform A of his attention onto the object and A will also attend to it. This is called \textit{joint attention} as both persons attend to the same object (Fig. 2). However, only one of them uses the other’s gaze direction to orient to the same target (A sees that B looks at the object and A then looks at the object). \textit{Shared attention} in contrast, implies that both individuals are aware of each other’s object of attention and each of them will use the other’s gaze direction to check that both attend to the same target (A sees that B looks at the object and will attend to the object; B notices that A attends to the object too and A and B look at each other’s eyes – mutual gaze – to make sure they both attend to the same object). Shared attention is thus more complex than joint attention and both play fundamental roles in social cognition.

Gaze following has been reported as early as 3–6 months of age (D’Entremont et al., 1997), although the exact age at which this capacity emerges is controversial (Emery, 2000). Before 9 months, infants can follow their mother’s gaze but are not capable of directing their attention towards the object of her interest. The joint attention capacity, which includes not only gaze monitoring but also pointing gestures, emerges around 9–14 months (Baron-Cohen et al., 1997a) but it’s only around 18 months of age that infants can attend to the same object of interest as their mother if the object is situated outside of their own visual field such as behind them (Butterworth, 1991). Joint attention is very important for the acquisition of language, which starts with the association between a word and the object it represents. Being able to orient one’s attention in the direction of gaze of the person naming the object is thus crucial (Baldwin, 1993; Baron-Cohen et al., 1997a). If someone says “dog” while looking at a dog, a young child listening for the first time to this word will orient his attention in the gaze direction of that person and will associate the word to its meaning. This learning strategy based on the use of people’s gaze direction emerges between 12 and 19 months of age (Baron-Cohen et al., 1997a) and positive correlations between gaze-following at 10–11 months of age and subsequent vocabulary scores at 18 months have been shown (Brooks and Meltzoff, 2005). A recent modeling study showed that gaze-following behavior at 10–11 months of age significantly predicted accelerated vocabulary growth until 2 years of age, even after controlling for the effects of age and maternal education (Brooks and Meltzoff, 2008).
Because of the important role of gaze early in development, the existence of an innate module specialized in the detection of gaze direction (eye direction detector—EDD) has been proposed (Baron-Cohen, 1995). The first function of EDD would be to detect any eye-like stimulus while its second function would be to determine whether the observed gaze is directed towards oneself or elsewhere. This module would play an essential role in the development of shared attention and in theory of mind (ToM). The term ToM refers to the capacity to explain others’ behaviors in terms of mental states, i.e. intentions, desires and beliefs and was originally introduced by primatologists to describe the possibility that chimpanzees understood certain mental states in other chimpanzees (Premack and Woodruff, 1978). This capacity was called a ‘theory’ because it is impossible to directly access others’ minds; we are simply guessing and inferring their mental states. ToM skills emerge around 4–5 years of age (Mitchell and Lacohée, 1991) and can be underlined by cartoon tests portraying social situations in which understanding false belief is essential, as in the classic Sally-and-Ann test (Wimmer and Perner, 1983; for more details on ToM, see Brune and Brune-Cohrs, 2006).

Baron-Cohen proposed the existence of another module called the intentionality detector (ID) which would understand any movement in the environment in terms of volitional movement, i.e. the goal-directed movement of an external agent (Baron-Cohen, 1995). Both EDD and ID would contribute to the development of shared attention, itself necessary for the development of ToM. Perrett and Emery (1994) proposed a direction-of-attention-detector (DAD) module that could process not only gaze cues but any attentional cue including head and body orientation (Perrett and Emery, 1994). They also proposed a mutual attention mechanism and suggested that the activation of EDD or DAD would be necessary for joint attention while shared attention would require the activation of the mutual attention mechanism in addition to EDD or DAD. The fact that theory of mind exists in congenitally blind individuals suggests that vision is not necessary for ToM to develop. However, in the course of normal development, the face and especially the eyes remain one of the richest sources of social information for the attribution of mental states to others. For instance, a 4-year-old child is capable of inferring that someone is thinking about something when their eyes are directed upward and to nothing in particular (Baron-Cohen, 1995). In Baron-Cohen’s theory, gaze direction is thus an important and privileged stimulus for the attribution of mental states. A test based on photographs of isolated eye regions has even been developed for the evaluation of ToM capacities in adults (Baron-Cohen et al., 1997b, 2001a) and in children (Baron-Cohen et al., 2001b). In this force-choice test, subjects need to designate, amongst four words evoking a mental state (e.g. preoccupied, puzzled, reassuring, jealous), which one best describes the eye region presented. This test was called the ‘reading the mind in the eyes’ test and was found appropriate in revealing ToM impairments in special clinical populations such as autistic spectrum disorders (see Section 5). The processing of gaze is thus an extremely important step in developing a social cognition and a theory of mind and relies on a very rich neural network that we describe in the following section.

4. Neural bases of eye and gaze processing

A few neuropsychological cases of brain lesioned patients and numerous brain studies using neuroimaging, electrophysiology and magneto-encephalography techniques, have suggested the existence of specialized neural circuits involved in eye and gaze processing. We review these fields in turn.

4.1. Neuropsychology evidence of eye and gaze processing impairments

In humans, a few rare lesion cases have revealed the involvement of cortical and subcortical brain areas in processing the eye region and gaze direction. One patient (MJ), whose
extended lesion involved almost the entire right superior temporal gyrus (STG), presented
important difficulties in gaze contact (Akiyama et al., 2006a). Her perception of others’ gaze
direction was also altered and she perceived left averted gaze as direct, and to a lesser
extent, direct gaze as averted to the right. Furthermore, she did not present the normal gaze
cuing effect seen in controls (Section 3.1) whereas, like them, she could normally orient to
the direction signaled by arrows (Akiyama et al., 2006b). Similarly as what is seen in
monkey studies in which the bilateral ablation of some parts of the superior temporal sulcus
(STS) considerably impairs the perception of gaze direction (Campbell et al., 1990;
Heywood and Cowey, 1992), this neuropsychological case suggests involvement of the right
STG in gaze perception and spatial orientation of attention by gaze. MJ presented a sort of
deficit in discriminating contra-lesional gaze direction only, suggesting a directionality role
of the STG in gaze processing. However, the extent of MJ’s lesion spanning the entire gyrus
length precludes any conclusion as to the existence of possible sub-regions of the STG
involved more specifically in one or the other aspect of gaze processing. The directionality
role of the STG also needs to be confirmed by left lateralized STG lesions (see Akiyama et
al., 2006a for a detailed discussion).

A few patients with amygdala lesions also present significant deficits in gaze processing
(Young et al., 1995) and in attending to the eye region (Adolphs et al., 2005; Spezio et al.,
2007). However, this type of patient usually presents other impairments especially in
emotional recognition of fear (Adolphs et al., 1994; Calder et al., 1996) and more generally
in the perception of social communication (Adolphs et al., 1998). Interestingly, in the
monkey, amygdala or temporal damage also lead to the disruption of social reactions, known
as the Klüver-Bucy Syndrome (Klüver and Bucy, 1939). Single cell studies in monkeys have
shown an important response of the amygdala to the presentation of faces and isolated eyes
(Leonard et al., 1985) and specific nodes of this structure contain cells sensitive to gaze
direction (Brothers, 1990; Brothers et al., 1990). A recent monkey fMRI study showed a
different activation within the amygdala for facial expressions and for gaze direction that
seems linked to attention and arousal (Hoffman et al., 2007). A similar functional and spatial
division in the human amygdala could explain the deficits in both facial expressions and
gaze processing seen in some amygdala patients. The precise role of the amygdala in gaze
processing in humans is still unclear but a recent study showed that the complete ablation of
this structure lead to a significant reduction in direct gaze contact during normal
conversations with others, along with an abnormal increase of eye movements directed at
the mouth rather than the eyes (Spezio et al., 2007). This suggests that the amygdala is
involved in directing attention to the eyes, an idea supported by the study of patient SM.
Following bilateral amygdala lesions, SM could not recognize the expression of fear
(Adolphs et al., 1994) and recently was found to fixate less on the eye region than control
participants (Adolphs et al., 2005). Using an image classification technique, the authors
showed that, unlike controls, SM was incapable of extracting the relevant information from
the eye region necessary to the recognition of fear expression, but remained capable of
recognizing fearful expressions when she was explicitly asked to pay attention to the eyes.
Although SM is impaired at recognizing fear emotion from other sources such as music
(Gosselin et al., 2007), Adolph et al.’s data suggest that her impairment in recognizing fear
from a face comes from a problem in spontaneously orienting attention to the eye region and
from an incapacity to extract relevant information from that region, rather than from the
incapacity to recognize the facial expression per se (Adolphs et al., 2005). This lack of
spontaneous exploration of the eye region is consequential to the amygdala lesions but not
necessarily to the impairment in recognizing fear itself. However, it is not clear whether SM
is also impaired at processing gaze or identity and whether her abnormal use of the eye
region could also be seen in other tasks than fear recognition. The role of the amygdala in
detecting and orienting attention towards relevant social stimuli like gaze has recently been
supported by the case study of five unilateral amygdala-damaged patients who did not
present the normal gaze-orienting effect but oriented normally to arrow cues in cuing paradigms (Akiyama et al., 2007).

The abnormal exploration of the eye region has also been reported in a case of prosopagnosia, the inability to recognize familiar faces (Bodamer, 1947). Patient PS is a case of acquired prosopagnosia with no visual agnosia, resulting from bilateral lesions to parts of the occipital and temporal lobes (Rossion et al., 2003a). Contrary to normal participants, PS used the mouth more than the eyes in a face recognition task and this impairment in exploring the eye region was not seen in real social interactions during which she could make normal eye contact (Caldara et al., 2005). PS is also not impaired at recognizing facial expressions or at discriminating gaze direction. The authors suggested that one possible cause of PS' prosopagnosia could be an impairment in extracting from the eye region the relevant configural information necessary to create accurate face representations in memory, rather than being a general deficiency in attending to the eyes (Caldara et al., 2005). As noted previously, the eyes seem to be used even more so in the processing of familiar than unfamiliar faces (Althoff and Cohen, 1999) and the role of eyes in identity recognition may increase with the familiarity the subject has with the face/person, a topic that will need to be explored in the future.

Impairments in gaze processing have also been reported in a few prosopagnosic patients (Campbell et al., 1990; Perrett et al., 1988), but the exploration of the eye region was not analyzed in these cases. Patient RB was seriously impaired at discriminating whether a face was looking at him or away, even at 20° of gaze deviation (Perrett et al., 1988). Campbell and collaborators reported the cases of an acquired prosopagnosic patient (AP) with a right posterior lesion and a developmental prosopagnosic patient (DP) with no known lesion, both impaired in a task requiring to point at the face whose gaze was directed at them. While the AP was slightly impaired, the DP performed at chance level and a detailed analysis revealed that she was using only face orientation to respond although she was clearly instructed to focus on gaze direction. Interestingly, both patients also presented impairments in processing facial expressions. However, the majority of reported cases of congenital prosopagnosia (i.e. with no lesions) do not present impairments in gaze direction processing (Dobel et al., 2007) and such impairments may not be a characteristic feature of this condition.

Abnormalities in processing both the eye region and its gaze have also been reported in Capgras delusion, an extremely rare neurological syndrome in which the patient believes that very familiar people (e.g. parents, siblings, spouse) have been replaced by identical-looking impostors or robots (Capgras and Reboul-Lachaux, 1923). Patient DS showed spared facial expression discrimination and identity recognition but poor accuracy in judging gaze direction regardless of the degree of gaze deviation used, answering that the face was looking at him even when it was not (Hirstein and Ramachandran, 1997). The authors suggested that his delusion was the result of a disconnection between face sensitive areas (such as inferior-temporal and STS areas) and the limbic system, likely the amygdala (Hirstein and Ramachandran, 1997). This could explain his gaze processing abnormalities given the implication of the amygdala in gaze processing and the influence of face familiarity on gaze perception (Teske, 1988, Section 2.2). However, no MRI data could confirm this disconnection and DS originally suffered a right parietal fracture sustained in a traffic accident (Hirstein and Ramachandran, 1997). The gaze abnormalities could thus be due to this parietal injury as well as, or in place of, the hypothetical disconnection between the limbic system and the face sensitive regions. However, the study of patients with unilateral neglect, a neurological disorder generally arising after right parietal lesions in which patients fail to detect or respond to contralesional stimuli (see Danckert and Ferber, 2006 for a review), does not support the involvement of parietal regions in gaze processing.

Neurosci Biobehav Rev. Author manuscript; available in PMC 2014 February 14.
Indeed, in these patients gaze cues could orient attention to the contralesional field and alleviate extinction, whereas endogenous orienting by arrow cues failed to produce similar effects (Vuilleumier, 2002). If anything, lesions to the right parietal areas were beneficial to gaze processing in these patients. In addition to supporting a specific role of gaze in orienting attention, these data on neglect syndrome support the hypothesis of a disconnection between the limbic system and the face sensitive regions in Capgras patient DS rather than an effect of the parietal system.

More recently, a case of Capgras delusion with no known psychiatric cause or brain lesion, was reported to spend significantly less time than controls fixating on the eye region of faces, but gaze direction was not assessed in this study (Brighetti et al., 2007). The Skin conductance response (SCR), a measure of emotional reaction and arousal, was also positively correlated with the time spent fixating the eye region of familiar faces in healthy controls (Brighetti et al., 2007). It thus seems that the eye region of very familiar persons such as family members triggers an emotional reaction that seems to be part of the entire familiarity experience. In parallel to the classic ventral route involved in identity processing, some authors have suggested a dorsal route for the affective response to faces (e.g. Bauer, 1984). Eyes could be at the crossroads of these two routes, conveying both configural cues and emotional cues necessary for accurate recognition of highly familiar faces. As reviewed in Section 2, the elements extracted from the eye region are different depending on the task performed: the ocular cues extracted for identity are mainly configural (e.g. distance between the eyes or eyebrows, etc.) and are not necessarily the same as the ones extracted for emotional processing (e.g. size of the white sclera, amount of eye openness, wrinkles around the eyes, etc.). It is possible that the ventral route is used to extract configural cues while the dorsal one is used to extract the information necessary for emotional processing. Future studies will have to address these questions more precisely.

Finally, frontal lesions have also been linked to impairments in gaze attention orienting. A recent study reported the case of patient VCR with bilateral orbito-frontal (OFC) lesions who did not present the normal target detection facilitation seen in gaze cuing paradigms (Vecera and Rizzo, 2006). VCR could normally orient attention towards peripheral cues but was unable to orient attention from word or gaze cues regardless of whether they were predictive (Vecera and Rizzo, 2004) or non-predictive (Vecera and Rizzo, 2006) of target location. According to dual-process theories of attention (Corbetta and Shulman, 2002), peripheral cues tap into exogenous, stimulus-driven attention orienting which is reflexive while symbolic cues such as words and arrows, tap into endogenous or goal-directed attention, which is voluntary. The fact that VCR presented intact exogenous orienting yet could not orient to gaze questions the reflexive nature of gaze orienting that we already mentioned in Section 3.2. Vecera and Rizzo (2006) suggested that, like arrows, gaze orienting was an endogenous, goal-oriented rather than exogenous and reflexive mechanism. In their “associative hypothesis”, they proposed that the apparent reflexive nature of gaze orienting is due to an over-learned association between gaze direction and the location the eyes refer to. The OFC would be responsible for this association and when damaged, no more gaze orienting would be seen. In a developmental perspective, this theory implies that early on, children would learn to associate others’ gaze direction with their object of attention and that this mechanism would involve the development of frontal areas. This idea agrees with the current state of knowledge concerning the role of the joint and shared attention mechanisms in cognitive development (Section 3), and with the hypothesis that the development of ToM skills is related to the maturation of the frontal lobes (Stuss and Anderson, 2004).

The review of the neuropsychological literature suggests that gaze perception involves a large brain network including at least the amygdala, the STG/STS, some ventro-temporal...
regions involved in face recognition such as the fusiform gyrus (the lesion of which is involved in most types of acquired prosopagnosia), and some frontal areas. In addition, most patients as described above present a specific deficit in the perception of direct/mutual gaze rather than general gaze direction impairments. We now turn to the neuroimaging literature on eye and gaze processing.

4.2. Neuroimaging data of eye and gaze processing

4.2.1. Brain areas involved in processing faces and eyes—Studies of single cell recordings in monkeys have revealed the existence of cells selective to faces and cells selective to eyes situated mainly in the inferior temporal cortex (IT) and in the STS (Perrett et al., 1982, 1984, 1985). These cells seem to be part of a larger neural network specialized in social interactions (Hasselmo et al., 1989; Perrett et al., 1985). It is important to note that face selective cells can respond to isolated face parts such as eyes (Perrett et al., 1982), while eye selective cells do not usually respond (or very little) to the eyes presented in the context of a face (Perrett et al., 1982, 1985). This peculiarity reflects the direct impact of face configuration on the neuronal response. The majority of face selective cells (about 60%) are also selective to head orientation (Perrett et al., 1985, 1991). Some cells respond preferentially to front-views of faces while others respond more for profile views (Desimone et al., 1984; Perrett et al., 1985). The majority of head orientation selective cells are also selective to gaze direction (Perrett et al., 1985). This double selectivity allows cells to respond when only one cue is available (the gaze or the head). Head orientation thus seems to be the default cue used to extract the direction of attention of others when gaze direction is not visible, for instance when the individual is far away (Emery, 2000; Perrett et al., 1992).

In humans, most neuroimaging studies used only face stimuli and the possible separate processing of eyes by a different neural network than the one involved in face processing (assuming there could be eye selective cells in the human brain just like in the monkey brain) remains to be established. The fusiform gyrus (FG) is the most studied brain region involved in face perception and recognition (Haxby et al., 2000; Kanwisher et al., 1997; McCarthy et al., 1997; Puce et al., 1995; Sergent et al., 1992). One traditional model based on neuroimaging data, suggests that facial features are processed within the inferior and medial occipital gyri (IOG/MOG), and then integrated within the FG where identity processing takes place (Haxby et al., 2000). Few studies have focused on the response obtained for isolated eyes and their results differ. One PET study involved the judgment of intentions inferred from the presentation of photographs of eye regions. In addition to brain areas responsive to emotions, an important activation was found in the inferior part of the STG while no activations within the FG, IOG or MOG were reported (Wicker et al., 2003). Similarly, a mental state judgment fMRI study using the same eye stimuli as used in the mind reading test of Baron-Cohen et al. (Section 3), reported the involvement of the amygdala, the FG and the STG (Baron-Cohen et al., 1999). In contrast, another study reported a decrease of the BOLD (blood oxygen-level dependant) response in the FG for isolated eyes compared to faces (Tong et al., 2000). However, this response was still much larger for eyes than for objects, suggesting a sensitivity of the FG to eyes. The sensitivity of the left (but not the right) FG to face parts was also demonstrated in a study using whole faces in a part-based judgment focusing on the eye regions (Rossion et al., 2000a). In a recent study, isolated eye regions were used in a match-to-sample task and activations in the amygdala, FG, STS, inferior parietal sulcus (IPS) and inferior OFC but not in occipital areas, were reported (Hardee et al., 2008). All these studies suggest that the task in which subjects are involved is as important as the stimulus itself and that classic models of face processing (e.g. Haxby et al., 2000) should be revised as eyes do not seem to be integrated in the rest of the face in the FG but rather activate their own large neural network with...
variations depending on the task. The traditional view that facial features are processed first and then integrated into a face percept in more anterior areas is also at odds with most electrophysiology data showing that face parts are processed after full faces, a point we come back to in Section 4.4. Finally, let’s note that in all these studies contrasting eye regions and faces, gaze is a confounding factor as the eyes are always open with a direct gaze, making it difficult to tease apart the effects of direct gaze from those of the eye region itself.

4.2.2. Brain areas involved in gaze processing—In agreement with the monkey literature, numerous PET and fMRI studies have shown that gaze processing, usually studied with faces rather than isolated-eye stimuli, involves the STS region (Allison et al., 2000; Bristow et al., 2007; Hoffman and Haxby, 2000; Hooker et al., 2003; Pelphrey et al., 2003; Puce et al., 1998; Wicker et al., 1998, 2003). Some have proposed a more general role of the STS in processing biological motion, with gaze being a specific type of biological motion (for a review, see Puce and Perrett, 2003). The STS is also involved in the gaze orienting effect but not in attention-orienting to arrows (Kingstone et al., 2004). However, discrepancies are found in the comparison between averted and direct gaze conditions. Some studies have found larger activations of the STS for direct gaze (Calder et al., 2002; Pelphrey et al., 2004; Wicker et al., 2003) while others have found larger activation of that region for averted gaze (for the left STS, Hoffman and Haxby, 2000). Some other studies have not found STS differences between averted and direct gaze (Pageler et al., 2003; Wicker et al., 1998). In a recent adaptation paradigm, a dissociation between left and right averted gaze was reported within the right anterior STS but the comparison between averted and direct gaze was not performed (Calder et al., 2007). Non-selectivity of the STS for gaze direction or emotion from eye regions is supported by a recent study (Hardee et al., 2008) and agrees with recent monkey data (Hoffman et al., 2007). In other gaze processing studies however, the STS is not even activated (George et al., 2001; Kawashima et al., 1999). It has to be emphasized that the term STS region is general and some studies report anterior areas (e.g. Calder et al., 2007; Kingstone et al., 2004) while others report posterior ones (sometimes referred to as pSTS, see Allison et al., 2000). Discrepancies in the reported results may thus also come from a difference in the actual localization of the so-called STS region in addition to differences in the paradigms and stimuli used.

Another key brain structure involved in processing the eyes and their gaze is the amygdala. In addition to explaining the impairments of some amygdala patients in both gaze and facial expression recognition, the possible division of the human amygdala into gaze and emotion neural nodes hypothesized in Section 4.1 could also explain the response of this structure to the combination of facial emotion and gaze direction, although conflicting results have been reported in the literature. In one fMRI study, the left amygdala responded more to an angry face with averted rather than direct gaze and more to a fearful face with direct rather than averted gaze (Adams et al., 2003). The authors interpreted these results as reflecting the ambiguous source of threat expressed by these faces (Section 2.2). In another study, the opposite was found, with larger left amygdala response to direct-than averted-gaze angry faces (Sato et al., 2004). However, in contrast to Adams et al. (2003) who used only front-view faces, in this study head orientation always matched gaze direction so that the averted gaze condition was in fact a 3/4-view face with averted gaze. Given this imbalance in the design, it is not possible to disambiguate the effect of head orientation and gaze direction in combination with facial expression. However, interestingly, Sato et al.’s study showed a positive correlation between the activation of the left amygdala and the negative emotion experienced by the subjects in viewing these faces, but not with the perceived negative emotion of the faces which did not differ between the two gaze/face orientations. The authors suggested that the amygdala activation reflects the emotional significance of the facial expression and the viewer’s emotional reaction towards the expression (Sato et al.,
In both the Adams et al. (2003) and the Sato et al. (2004) studies, the left but not the right amygdala was involved. A recent study also suggests a hemispheric difference in amygdala response linked to the type of eye stimulus used. In this implicit processing of eye regions, the left amygdala activated only for fearful eyes but not for gaze shifts even though the eye white area had been equated between gaze and fear conditions, while the right amygdala responded to all conditions equally, including joyful and control eyes (Hardee et al., 2008). These results contrast with the idea that the amygdala responds only to the eyes’ white area (Whalen et al., 2004) and rather suggest a hemispheric difference in stimulus selectivity for this structure. Hardee et al. (2008) suggested that the lack of selectivity of the right amygdala could reflect a mechanism tuned to the fast and coarse detection of potential dangers, while the left amygdala could reflect a mechanism tuned to details enabling the verification of whether the threat is real. This hypothesis could explain the results of Sato et al. (2004) of a larger left amygdala response to angry faces looking straight at the viewer given this condition elicited the most negative feelings in subjects who may have perceived this stimulus as a threat.

Thus, in addition to its likely role in orienting attention towards the eye region (Section 4.1), results from the neuroimaging literature suggests the amygdala response may not be modulated by gaze direction per se but rather by the emotional implication of a given stimulus for the subject in a particular task context. This interpretation may explain why, like the STS region, discrepancies between direct and averted gaze have been found for the amygdala. Some studies have reported that the amygdala was more active for direct than averted gaze (George et al., 2001; Kawashima et al., 1999) while others have found the opposite (Hooker et al., 2003; Wicker et al., 2003), or no amygdala activation was reported (Pageler et al., 2003). Again, the difference in stimuli used (especially the head orientation) and/or in the emotion experienced by the subject in viewing these stimuli, could explain these inconsistencies that future studies will have to address by systematically correlating the brain activations to psychological measures.

Modulations of the FG activity by gaze direction have also been reported and again, results are inconsistent. Some studies have reported larger activities for direct than averted gaze (Calder et al., 2002; George et al., 2001; Pageler et al., 2003), a finding interpreted as reflecting an increased processing of faces due to the social significance of direct gaze (George and Conty, 2008). However, other studies failed to find gaze modulations in this region (Hooker et al., 2003; Pelphrey et al., 2004). In addition to the stimuli and possible personal emotional involvement of subjects, these inconsistent results (including for the STS and amygdala) could be due to a different sensitivity of all these areas to a specific gaze direction depending on the subject’s task. The great variability of tasks used in the gaze literature could contribute, at least in part, to these discrepancies: active (Hoffman and Haxby, 2000) or passive tasks (Wicker et al., 1998), gender discrimination judgment (Adams et al., 2003; George et al., 2001; Sato et al., 2004), attribution of intentions (Wicker et al., 2003), eye brow size judgment (Calder et al., 2002), explicit gaze direction judgment (Hooker et al., 2003; Kawashima et al., 1999; Pageler et al., 2003), perception of eye movement (Pelphrey et al., 2004; Puce et al., 1998), delayed-match-to-sample task (Hardee et al., 2008), etc. This effect of task is supported by Itier et al.’s findings that the FG, some areas of the STS and especially the medial frontal areas (BA10) presented different activations to direct compared to averted gaze depending on whether subjects were involved in an implicit or explicit gaze judgment task, with larger activations seen for direct gaze in the explicit judgment but for averted gaze in the implicit task (Itier et al., 2005). More
studies directly manipulating tasks are needed to sort out the inconsistent results reviewed above.

In agreement with the recent neuropsychological study mentioned above (Vecera and Rizzo, 2006), a few neuroimaging studies also reported the involvement of frontal areas during gaze processing. In a recent study, direction of gaze was manipulated implicitly while subjects had to judge the size of the eyebrows of face pictures. In addition to the STS region, a larger activation was found in medial frontal areas (BA8/9 and BA10) for averted compared to direct gaze or even faces with closed eyes (Calder et al., 2002). This bilateral activation of superior frontal regions (BA8) was also reported in an explicit gaze direction judgment (Hooker et al., 2003). The involvement of the superior and medial frontal gyri (BA6) was reported in another gaze study (Wicker et al., 1998) but as the task was passive, it is difficult to know whether this region was involved in gaze processing per se or was responding to other factors. Similarly, the isolated eyes of Hardee et al. (2008) involved bilateral activation of the OFC that was not selective to gaze shift or emotion but it is difficult to interpret the involvement of this region as the task was implicit. Importantly, these frontal regions, especially the medial prefrontal and orbito-frontal cortices, are found in numerous ToM and joint attention studies (Section 3), just like the STS and the amygdala (Adolphs, 1999; Amodio and Frith, 2006; Baron-Cohen et al., 1999; Frith and Frith, 1999; Stone et al., 2003; Williams et al., 2005). This suggests that gaze processing recruits a large network of brain areas involved in ToM and social cognition and that the various degrees of involvement of each of these regions depends on the specific task utilized.

Finally, a few studies have also reported the activation of some parietal areas in gaze perception. The intra-parietal sulcus (IPS) was reported for the viewing of averted eye movements (Bristow et al., 2007; Hardee et al., 2008; Pelphrey et al., 2003; Puce et al., 1998) and perception of averted gaze in static faces (Hoffman and Haxby, 2000). Other studies have reported the activation of the inferior parietal and superior parietal lobules for the movement of eyes within faces (Calder et al., 2007; Wicker et al., 1998). As the parietal cortex including the IPS is involved in covert shift of spatial attention (Corbetta and Shulman, 2002; Grosbras et al., 2005), gaze-related activity in these regions is usually thought to reflect the engagement of the attentional system for encoding the spatial direction of another’s gaze and orienting attention in that direction. This general role is supported by the non-selectivity of that region for gaze as reported by Hardee et al. (2008).

The review of the neuroimaging literature on gaze perception confirms the results from the neuropsychological literature regarding the involvement of the amygdala, the STS/STS, some ventro-temporal regions involved in face recognition such as the FG, some parietal and frontal areas, in agreement with a fronto-parietal circuit for gaze as found in a meta-analysis involving 59 neuroimaging studies (Grosbras et al., 2005). This analysis also found that gaze perception shared common neural substrates with visually triggered saccades and visually triggered shifts of attention (Grosbras et al., 2005), especially the temporo-parietal junction (TPJ) area, which is adjacent to the pSTS region. However, the precise role of these regions in processing direct or averted gaze and the influences of task demands remain to be better understood. In addition, future studies will have to investigate the possible left/right gaze processing asymmetries and their underlying neural bases.

4.3. Temporal aspect of eye and gaze processing: electrophysiology and magnetoencephalography evidence

4.3.1. Early processing of faces and eyes—Most current electrophysiological data in humans concern the negative event-related potential (ERP) component N170 which is maximally recorded over lateral posterior sites of the scalp between 130 and 200 ms after face onset (Bentin et al., 1996; George et al., 1996; Rossion et al., 1999b). This component
responds more to faces than other object categories and is thus qualified of “face-sensitive component”. The N170 is also sensitive to eyes presented in isolation while it responds very little to other face parts such as mouths or noses (Bentin et al., 1996). Although delayed, the N170 is even larger for eyes than faces (Bentin et al., 1996; Itier et al., 2006b; Jemel et al., 1999; Shibata et al., 2002; Taylor et al., 2001a,c) and this is seen as early as 4 years of age (Taylor et al., 2001a). For this reason, it was initially suggested that it may reflect the processing of eyes rather than faces. However, when eyes are erased from the face, the N170 is slightly delayed but of the same amplitude as for normal faces (Eimer, 1998; Itier et al., 2007b), which suggests that this component reflects the configural (or holistic, global) aspect of face processing and not the activity of an eye detector (Eimer, 2000b; Rossion et al., 1999a). Many studies using the well known “face inversion effect” (FIE) have supported the idea that the N170 reflects the configural processing of faces. Faces presented upside-down (inverted by 180°) are harder to perceive, memorize and recognize than upright faces and this effect is disproportionately larger than for common objects (reviewed in Rossion and Gauthier, 2002). Inversion mainly disrupts the configural processing of faces and, like objects, inverted faces are processed analytically, in a feature-based manner (Maurer et al., 2002; Rhodes et al., 1993; Rossion and Gauthier, 2002). The FIE is also seen on the N170 which is larger and delayed compared to upright faces (Bentin et al., 1996; Eimer, 2000a; Itier et al., 2004, 2006b; Itier and Taylor, 2002, 2004a; Latinus and Taylor, 2006; Rossion et al., 1999b, 2000b; Sagiv and Bentin, 2001) while inverted objects only induce an N170 delay but no amplitude changes (Itier et al., 2006b; Rossion et al., 2000b). The N170 recorded to faces and eyes present with different developmental trajectories (Taylor et al., 2001a) and their topographies are different even in adults (Itier et al., 2006b, 2007b), reflecting different underlying brain generators for the two types of stimuli (Fig. 3).

In magneto-encephalography, a similar face sensitive component appears at the same latency as the N170 and is called the M170 or M2 (Halgren et al., 2000; Lu et al., 1991; Sams et al., 1997). The M170 is also delayed for eyes compared to faces (Taylor et al., 2001b; Watanabe et al., 1999b) but in contrast to the N170, it is of similar amplitude for both categories (Taylor et al., 2001b). Source analyses have systematically found the FG as the main source of the M170 component (Halgren et al., 2000; Itier et al., 2006a; Linkenkaer-Hansen et al., 1998; Liu et al., 2000; Lu et al., 1991; Sams et al., 1997; Sato et al., 1999; Swithenby et al., 1998; Taylor et al., 2001b; Watanabe et al., 1999a, 2003). In contrast, source analyses of the N170 are still controversial. Some have reported the STS region as the main source of this component (Batty and Taylor, 2003; Henson et al., 2003; Itier et al., 2006a,b; Itier and Taylor, 2004b; Watanabe et al., 2003) while others have reported the FG (Itier and Taylor, 2002; Rossion et al., 2003b; Schweinberger et al., 2002; Watanabe et al., 2003) (Fig. 4). This source difference could be explained by the different sensitivity of the two techniques. While EEG is sensitive to both tangential and radial sources, the MEG is mostly sensitive to tangential sources. Any source in the FG that is not uniquely tangential would have a tangential component and a radial component and would thus be caught by both techniques. In contrast, if the STS source was oriented radially as hypothesized by Watanabe et al. (2003), it would be only caught by the EEG technique. EEG would thus catch both the FG and STS regions as the main sources of the N170 while the MEG would catch only the FG as a source of the M170. The reason why one or the other source has been reported as the main source for the N170 is unclear but several factors may be involved. The first one could be the source analysis technique used (dipole source modeling versus 3D current density sources) and the fact that most studies performed source analyses on grand averages rather than on individual data. This argument is as tenable as using another technique, namely a beam-former analysis and an individual subject approach, some have found that the M170 recorded to faces originated from the FG but also from the IOG (Itier et al., 2006a). A second factor is that the involvement of the two (or more) sources likely varies with task demands. For instance, based on neuroimaging data (Haxby

Neurosci Biobehav Rev. Author manuscript; available in PMC 2014 February 14.
et al., 2000), it could be possible that the FG was more involved than the STS in a face recognition task, while the opposite could be found in a gaze discrimination task. Future studies will need to perform source analyses on a single subject basis, combining simultaneous ERP and MEG recordings and directly comparing task effects on these sources.

In pharmaco-resistant epileptic patients implanted with sub-dural electrodes for surgery, the N200, a negative wave directly recorded on the cortical surface, is selective to faces but also to face parts such as the eye region (Allison et al., 1994a,b, 1999; McCarthy et al., 1999; Puce et al., 1999). The N200 has been recorded on two main areas of the cortex, ventrally on the surface of the FG, ITG and IOG (centroïd within the FG) and laterally, on the surface of the STS region (which included the MTG and STG). In both of these areas, face-specific and part-specific N200s have been recorded in adjacent patches of cortex of approximately 12–16 mm wide and 15–35 mm long (Allison et al., 1999) (see Fig. 5A). Importantly, just like the N170 and M170, the N200 is always delayed for eyes compared to faces, whether it is recorded from the face-specific or part-specific areas (McCarthy et al., 1999). These intracranial data suggest that, like monkeys, humans may possess face and eye selective cells in the FG and STS regions and that both regions contribute to the generation of the N170 and M170 components recorded on the scalp after the presentation of faces and eyes. Based on this assumption, Itier et al. recently proposed a new alternative model of early face processing (Itier et al., 2007b), using the FIE described above. The N170 FIE (i.e. N170 amplitude increase for inverted faces) is classically thought to reflect the recruitment of object selective neurons needed to process inverted faces, due to the disruption of the face configuration (Rossion and Gauthier, 2002). However, this interpretation is difficult to reconcile with the absence of N170 FIE when eyes are removed from the face yet a clear inverted face, whose configuration has been disrupted, can still be seen (Itier et al., 2007b). Itier et al. showed that, in contrast to normal faces, isolated eye regions and faces-without-eyes did not present the N170 FIE (Fig. 5B). Moreover, the N170 to inverted faces was not different from that to eyes (upright or inverted) and the same was seen for its topographies. The authors suggested that the FIE on the N170 was driven by the presence of the eyes as this effect was no longer seen when eyes were removed from the face.

Instead of the STS being specialized in processing solely the eyes while the FG processes whole faces (Sagiv and Bentin, 2001), Itier et al. (2007b) proposed that both face-selective and eye-selective neuronal populations coexist in the STS region whose response (hence the magnitude of the N170) is modulated by the facial context—Fig. 5C. By disrupting the relationships between the eye region and the rest of the face, inversion somewhat isolates the eyes so that they would be processed as if the rest of the face were not present. This model suggests that processing inverted faces is more or less equivalent to processing the eye region and it accounts for (i) the absence of N170 modulations when the eyes are eliminated from the face (face-without-eyes), (ii) the absence of inversion effects for faces where eyes are not clearly visible (e.g. schematic or Mooney faces, Latinus and Taylor, 2005; Sagiv and Bentin, 2001) and (iii) the larger amplitude of the N170 recorded to isolated eyes than to faces. Finally, the model suggests that eyes and faces are processed by different neuronal populations whose response depends on the facial context. This agrees with monkey electrophysiological data and with recent ERP data in humans suggesting that information in the eye region is consequential only if the stimulus has a face configuration (Bentin et al., 2006). Here we propose some modifications to this original model. Itier et al. (2007b) assumed that the N170 mainly came from the STS region and that the eye neurons were simply not responding to full faces. However, as explained above, it is likely that both the STS and the FG contribute to the N170 and both face and eye neurons seem to coexist in these two cortical regions (Puce et al., 1999). Moreover, it is possible that eye neurons are activated by the presentation of faces but because of the facial context, they are inhibited by

Neurosci Biobehav Rev. Author manuscript; available in PMC 2014 February 14.
the nearby face neurons when faces are presented upright (Fig. 5C). Although neural inhibition mechanisms (active suppression) have been proposed before (Allison et al., 2002), they are speculative and more cell recordings in monkeys are needed. The respective contributions of the STS and FG sources to the scalp-recorded components also likely depend on the task and more source analyses studies manipulating tasks are needed to clarify these issues and refine the present model.

4.3.2. Temporal dynamics of gaze perception—Given the sensitivity of the N170 to eyes and the likely involvement of the STS and FG regions in generating this component and in processing gaze (Section 4.2), it would seem logical that the N170 was sensitive to gaze direction. However, most studies did not find gaze modulations on the N170 (Grice et al., 2005; Klucharev and Sams, 2004; Schweinberger et al., 2007; Taylor et al., 2001c) and the few that did, reported different effects. One study reported a weak amplitude increase of the N170 for averted-gaze faces restricted to the right hemisphere with no latency modulations (Watanabe et al., 2002). Another one reported an interaction between gaze and head orientation, with slightly larger N170 amplitudes for averted- than direct-gaze front-view faces but not for 3/4-view faces (Itier et al., 2007a). Interestingly, this study also reported a sensitivity to head orientation as early as 100 ms (P1 component), suggesting head orientation processing begins before that of gaze direction (Itier et al., 2007b). This agrees with the idea mentioned before that head orientation is the default cue used to extract the direction of attention of others when eye gaze is not visible (Emery, 2000; Perrett et al., 1992). A third study reported a delayed N170 for closed eyes but no difference between averted and direct gaze (Taylor et al., 2001c). Finally, one MEG study found that the M170 was earlier and smaller for averted than direct gaze for isolated eye stimuli but not for full faces (Taylor et al., 2001b). Here again, the discrepancies between studies and between the M170 and N170 components are unclear and will require further investigations.

All these studies compared gaze directions in static stimuli. The presentation of a face or eye region with straight gaze immediately followed (i.e. no delay) by the same face or eye region with averted gaze (or vice versa) induces a perceived gaze motion. In these conditions, both N170 and M170 are greatly sensitive to the direction of the apparent gaze motion. An earlier and larger N170 was reported for a perceived gaze movement going from direct to averted gaze compared to when the gaze was first averted and then direct (Puce et al., 2000). This contrasts with the results of a recent explicit gaze direction discrimination study in which averted and direct gaze motions were compared using the same intermediate gaze position as a starting point in order to avoid a bias in one or the other gaze direction (Conty et al., 2007). In these conditions, the N170 was larger when the eyes moved towards the viewer (direct gaze motion) compared to when they moved away (averted gaze motion), and this effect was more pronounced for front-view faces. The authors suggested these modulations may represent biological motion processing added onto the face sensitive encoding processes reflected by the N170 and/or emotional effect triggered by gaze contact (Conty et al., 2007). This greater sensitivity to direct gaze motion has also been reported in MEG with the equivalent current dipole moment of the M170 being stronger for direct gaze motion (from averted to direct position) compared to averted gaze motion (Watanabe et al., 2006). Other studies will need to confirm this special neural sensitivity to a gaze moving towards the observer but these results support a special effect of mutual gaze as already mentioned previously (Section 2).

Few gaze studies have focused on later components than the N170. A large increase of slow wave potentials for direct-gaze compared to averted-gaze static faces was reported between 400 and 600 ms at centro-parietal sites, regardless of face orientation or task (Itier et al., 2007a), similar to the larger P3 component for direct than averted gaze motions reported by Conty et al. (2007). Again, this electrophysiological sensitivity for direct gaze independent...
of task suggests a specific role for direct/mutual gaze in human cognition. However, at this long latency, these modulations are likely related to more cognitive steps following gaze direction discrimination rather than gaze discrimination itself. This idea agrees with two other studies. The first one reported a modulation of later ERPs (250–650 ms) by the social context, differentiating between gaze avoidance, mutual gaze and joint attention conditions (Carrick et al., 2007). In the second study, late ERPs (270–400 ms) distinguished between a ToM condition where a mental state judgment was inferred from the presentation of isolated eye regions and a simple gender categorization control condition (Sabbagh et al., 2004).

Thus, the exact latency at which gaze direction discrimination occurs is unclear. In static stimuli, it may start at the level of the N170 or soon after, around 250 ms (Schweinberger et al., 2007) while the social interpretation or mental state judgment derived from the perceived gaze likely occurs between 300 and 600 ms. In dynamic stimuli, gaze starts to be processed as early as 150 ms (Conty et al., 2007). In this last study, source analyses of the gaze motion differences found between 150 and 200 ms after gaze motion onset reported the involvement of frontal regions (BA8/9) as early as 150 ms, followed by orbito-frontal areas starting around 180 ms and by the right STS around 190 ms. These areas are the same as reported by neuroimaging studies (Section 4.2) with an interesting temporal sequence of activation starting with frontal cortices. Judging gaze direction in isolated faces is also impaired when the right superior temporal cortex is stimulated between 200 and 300 ms by transcranial magnetic stimulation (TMS) (Pourtois et al., 2004), confirming the involvement of this region around 200 ms and likely after.

In developmental studies, electrophysiological data on gaze processing are scarce. It was shown that as early as 4 months of age, a difference between direct and averted gaze faces (static stimuli) could be seen as an increase of the N290 component for direct gaze (Farroni et al., 2002), the infant face-sensitive component that may be the precursor of the N170 (de Haan et al., 2002). The results were also found for 3/4-view faces (Farroni et al., 2004). Localization methods and source analyses have suggested that the N290 originates from the FG (cited in Grossmann and Johnson, 2007), which would suggest that gaze processing involves mainly this brain structure early on. The processing of face and gaze direction would thus recruit common networks within the first few months of life that would then dissociate to be more specialized in adults, although adults would also retain a sensitivity for gaze in the FG as we saw earlier (Section 4.2). Other studies have focused on the infant slow wave potential (SWP) at fronto-central sites and found a larger amplitude for direct compared to averted gaze only when faces expressed anger, but not for joyful or neutral faces (Striano et al., 2006). However, the general lack of developmental data, in particular in children and adolescents, urges caution in interpreting the meaning of these infant results on gaze processing and more studies are needed before we can conclude that gaze processing occurs that early in the infant brain.

Finally, as mentioned in Section 3, gaze direction seems to facilitate target detection, and understanding the relationships between gaze direction and the object of interest being gazed at represents a fundamental step in the proper development of social cognition. In adults, the early visual ERP components (P1 and N1) recorded after the presentation of the target are earlier and larger in amplitude when the gaze direction of the preceding cuing face is congruent rather than incongruent with the target localization, and this has been found for static (Schuller and Rossion, 2004) and dynamic (Schuller and Rossion, 2001) gaze stimuli. These results directly demonstrate an attentional effect onto the processing of targets cued by eye gaze. The time course of attention orienting by arrows is also known (e.g. Hopf and Mangun, 2000; Nobre et al., 2000) and a recent study showed that the orienting effect by gaze and arrow cues were different at parietal and frontal sites between 220 and 260 ms after cue...
onset, reflecting different underlying generators involved in these attentional effects (Hietanen et al., 2008a), in agreement with neuroimaging studies (Hietanen et al., 2006).

4.4. Is there an eye detector in the human brain?

The hypothesis of the eye direction detector (EDD) proposed by Baron-Cohen (Section 3.2) assumes that a mechanism in the human brain is specialized for the detection of eyes in the environment, and this “module” would likely be present at birth. Its primary goal would be to detect the presence of eyes or eye-like stimuli in the environment and then to determine whether these eyes are looking towards or away from oneself. However, its existence is not yet supported empirically at the brain level and it is still unknown whether eyes are processed independently in specialized areas or as part of the face in the same face responsive areas. Only a few imaging studies have explored the brain response to isolated eyes and reported conflicting results (Section 4.2.1). On the other hand, ERP data suggest an earlier sensitivity to eyes than faces in young children as well as the existence of different neuronal populations responding to eyes and faces as seen in monkeys, all of which support the possible existence of EDD. The eye detector hypothesis is also supported by behavioral data in infants and by the fact that the scanning of the face always starts with the eyes early on in visual processing (Section 2). If EDD exists, how does it work at the neural level?

In contrast to the classic view derived from the imaging literature, according to which face components are processed first and subsequently integrated into a face percept (Haxby et al., 2000), the EEG/MEG literatures suggest faces are processed first as a whole, and facial features processed later. However, recent studies suggest that holistic processing of the face and featural processing of the eyes occur in parallel and compete depending on the task, attention and context bias (Bentin et al., 2006; Itier et al., 2007b). In most instances holistic processing would prevail over featural processing except in some situations, for instance when viewing isolated eyes (a face with a hood or hat and a scarf up to the nose) or when voluntarily focusing on the eyes for a specific judgment. One recent study combining one image classification technique with EEG recordings suggested that within whole faces, the eyes were attended and processed first, between 100 and 150 ms, i.e. before the N170 peak at which point the whole face is supposedly processed (Schyns et al., 2007). According to the authors “the N170 reflects a process that integrates facial features over time”. However, the use of only one task (emotion discrimination) and one specific image classification technique in that study, precludes any generalization to other tasks and situations. Nevertheless, it agrees with another recent study showing that face discrimination based on face parts is impaired by TMS pulses delivered over the right OFA between 60 and 100 ms after stimulus onset (Pitcher et al., 2007). Although full faces were used, subjects’ judgment was biased towards the eyes and the mouth, suggesting the involvement of the rOFA in facial feature processing before the N170 (and even before, or around, the P1 component). These new results raise once again the empirical question which will have to be addressed by future studies: what comes first, the features or the whole?

Whether a neuronal system specialized in processing the eyes exists in the human brain and is different from the face processing system is thus not yet established. If EDD exists then the traditional models describing the way we process faces (e.g. Bruce and Young, 1986; Haxby et al., 2000) should be revised in order to explain the dynamics of the two systems, because in this view face perception would not be the result of the integration of equally important features within a whole but would rather result from the integration of a system dedicated to the eyes together with a system devoted to processing the rest of the face. If this eye detector mechanism exists, it could (i) recruit one or several brain regions that are also part of the face system, (ii) recruit the entire face system or (iii) recruit another brain network that is different from the face one. Based on the few imaging studies available so
far (see Section 4.2), this last possibility is unlikely. It is also possible that eyes recruit the same areas as faces but that these areas are interconnected differently depending on whether the subject focuses on the local or global aspects of the face, or depending on what information is extracted (emotion, identity, intentions, etc.). The EEG/MEG literature suggests the involvement of several brain sources as early as 170 ms after stimulus onset but the exact areas and their differential engagement as a function of task are still debated. In contrast, imaging studies identified several brain nodes involved in the processing of faces and eyes but their temporal dynamics are unclear. The spatio-temporal neural networks involved in these various face and eye processes and their functional connectivity depending on stimulus category and subjects’ tasks remain to be established. Future studies combining EGG/MEG and fMRI techniques, along with network analyses such as structural equation modeling, may help resolve these questions.

We now turn to a pathology in which impairments in processing the eye region and gaze direction may be central to the social interaction deficits observed.

5. Abnormal gaze processing: the case of Autism Spectrum Disorders (ASDs)

5.1. Impairments in gaze and social cognitive processes in ASDs

Autism is a severe developmental pathology presenting with great impairments in communication and reciprocal social interactions. Autistic individuals have difficulties extracting social cues from faces. Even in Asperger Syndrome (AS) individuals whose IQs remain normal or superior to average, deficits in the recognition of face identity, gender, age and expressions have been reported (Boucher and Lewis, 1992; Celani et al., 1999; Gepner et al., 2001; Tantam et al., 1989; Teunisse and De Gelder, 1994). A recent review however suggests these face processing impairments may not be as important as previously thought (Jemel et al., 2006).

One explanation for these deficits is a lack of interest for the human face (Jemel et al., 2006). Attraction for faces normally starts early during development except for infants later diagnosed with ASDs (Baron-Cohen et al., 1996; Osterling and Dawson, 1994). Autistic children look less at faces than age-matched control children and when they do so, perceptual processes and exploratory ocular movements seem abnormal. In contrast to controls who process faces mainly configurally, ASD individuals rely preferentially on features to process faces (Frith, 1989; Mottron and Burack, 2001; Mottron et al., 2006; Senju et al., 2008). Logically, this should favor the processing of eyes and gaze but eye movement monitoring studies have shown that ocular exploration of faces is disorganized, incoherent and quite variable, and that ASD individuals fixate less on the eye region than controls and more on other parts such as the mouth, the chin, the hair line-forehead limit or the ears (Dalton et al., 2005; Klin et al., 2002; Langdell, 1978; Pelphrey et al., 2002). This is seen during free viewing or when the task requires facial emotion recognition (Pelphrey et al., 2002). A recent study reported that autistic subjects performed ocular saccades away from the eyes, which were correlated with the amount of social information contained in that facial zone (Spezio et al., 2007), supporting the idea of an eye avoidance in autism. These data are in agreement with the clinical diagnosis of the pathology (abnormality in eye contact is one diagnostic criteria of ASDs according to DSM IV, APA, 1994) and suggest an aversion for direct gaze that could explain a lot of non-verbal communication and social interaction impairments these individuals present. A recent study also reported a stronger skin conductance response to direct than averted gaze in children with autism but not in age-matched controls (Kylliäinen and Hietanen, 2006), supporting the idea that mutual contact is too emotionally arousing for autistic individuals who thus need to avoid it.
Furthermore, the spontaneous orienting of attention towards the object of interest indicated by others’ direction of gaze has been reported deficient in autistic children who are insensitive to gaze direction as an index of the speaker’s intention to infer (Baron-Cohen et al., 1997a). This deficit could result from a more general impairment in visual attention orienting but attention cuing studies (Section 3.1) have reported both age matched controls and autistic children presented orienting-to-gaze effects (Kylläinen and Hietanen, 2004; Ristic et al., 2005; Senju et al., 2004). In one study typically developed children located targets cued by eye gaze more quickly than children with ASDs and targets cued by arrows did not trigger reflexive orienting. In contrast, both social and non-social cues shifted attention to the cued location in children with autism (Senju et al., 2004; see also Vlamings et al., 2005 for similar findings). This abnormal spontaneous orienting towards arrow cues could reveal a lack of preferential social sensitivity to eye cues in autism (Senju et al., 2004). However, caution is required as spontaneous orienting towards the direction of arrows has been found in other studies in normal adults (Ristic et al., 2005; Tipples, 2002) and may thus not be abnormal. Spontaneous orienting towards arrow cues in autism could also reveal a lack of inhibition in autistic children, i.e. a difficulty in disengaging attention (Nation and Penny, 2008). Finally, it has been suggested that rather than being deficient, gaze-orienting might simply be delayed during development in ASDs (see Nation and Penny, 2008 for a more detailed developmental perspective).

When told explicitly to pay attention to gaze, autistic individuals are capable of general knowledge about the eyes and the information they give. For instance, they know that the eyes are necessary for sight. They can discriminate gaze direction (Ristic et al., 2005; Wallace et al., 2006) and identify whether someone is looking at them or what someone is looking at. One study has reported a deficit in detecting direct gaze compared to control children matched for non-verbal development, and while performances in controls were better for direct than averted gaze detection, autistic children did not present any difference between the two types of gaze (Senju et al., 2005). However faces were in 3/4-view in this study and these deficits may thus be due to a processing conflict between head and gaze orientations rather than a real impairment at detecting gaze direction. It was also shown that ASD individuals used specific strategies in gaze orienting: rather than treating the eyes as a social cue like controls, they seem to use mainly low-level information such as pupil direction and the contrast between iris and white sclera (Ristic et al., 2005).

According to some researchers, the social interaction deficits characteristic of ASDs mainly result from ToM deficits (see Section 3) stemming from impairments in processing social information derived from gaze (Baron-Cohen, 1995; Baron-Cohen et al., 1997b, 1999). While normal people understand and infer others’ thoughts effortlessly, ASD individuals present tremendous difficulties in perceiving and understanding others’ mental states (Baron-Cohen, 1995; Baron-Cohen et al., 2001a) and this could be due to abnormal EDD and/or ToM mechanisms. In fact, a complete absence of joint attention (precursor of ToM) at 18 months of age is predictive of a later diagnosis of autism (Baron-Cohen et al., 1996). A recent study reported that the normal development of ToM, measured with animated geometric shapes and language descriptions of mental states, is correlated with the development of the capacity to spontaneously infer the locus of attention of a face using gaze cues in control but not in ASD children (Campbell et al., 2006). This result underscores the links between gaze and ToM and supports the idea of their common neuronal basis as suggested by neuroimaging data (Section 4).

Taken together, studies in ASDs suggest no general impairment in visual attention or specific impairment in gaze orienting per se but rather a deficit in the perception of social cues and a possible impairment in direct/mutual gaze discrimination. The main deficits seem
to lie in the incapacity to extract relevant information from the eye region necessary for social communication (Nation and Penny, 2008).

5.2. Neural bases of gaze and social cognition impairments seen in ASDs

Numerous studies have tried to understand the neural basis underlying social deficits in autism and atypical brain activations have been reported in ASD individuals for face perception (Pierce et al., 2001; Schultz et al., 2000), face identity recognition (Dawson et al., 2002) and facial expression discrimination tasks (Critchley et al., 2000). Fewer studies have focused on the neural bases of gaze processing in this pathology. Because of the socio-emotional difficulties characteristic of this disorder, the limbic system has been focused on and anatomical and functional abnormalities of the amygdala have been reported (Abell et al., 1999; Bauman and Kemper, 1985). In a task requiring the attribution of mental states to photographs of isolated eyes, ASD individuals activated prefrontal and superior temporal cortices like controls, but not the amygdala (Baron-Cohen et al., 1999). The amygdala activation is also strongly correlated with the time spent fixating the eyes in ASD individuals (Dalton et al., 2005), supporting the eye avoidance hypothesis.

Recent fMRI studies on autism have also reported neuroanatomical and functional abnormalities in other brain areas, particularly in the temporal lobes (Abell et al., 1999; Carper et al., 2002; Zilbovicius et al., 2006). Using cuing paradigms, two studies showed that whether the gaze of the centrally presented face was directed at a target (congruent gaze) or in the opposite direction (incongruent gaze), a similar STS activation was found in ASD individuals while normal controls presented a larger activation for the incongruent gaze condition (Pelphrey et al., 2003, 2005). This lack of STS modulation to social context confirmed that although a change in gaze direction was detected, its communicative and social value remained impaired in ASDs (see also Mosconi et al., 2005). Hypo-activation of the FG in ASDs has also been consistently reported (Critchley et al., 2000; Dalton et al., 2005; Ogai et al., 2003; Schultz et al., 2000) and could result from a smaller fixation time onto the eyes given the correlation between FG activation and time spent fixating on the eye region (Dalton et al., 2005). Finally, abnormalities in the OFC of autistic individuals have been reported (Schmitz et al., 2007). Although the overall volume of this structure was not different between controls and ASD individuals, it was positively correlated with circumscribed interests in the latter group, a classic symptom of the disorder. In addition, the right lateral OFC was smaller for autistic subjects compared to age-matched controls (Schmitz et al., 2007). Taken together, these structural and functional abnormalities in the amygdala, STS, FG and OFC suggest an abnormal processing of social cues by the entire social brain network in autism.

Just like in brain imaging, abnormalities in ERPs have also been found. Some studies reported a delayed N170 to faces (McPartland et al., 2004; O’Connor et al., 2005) with or without a smaller amplitude of that component in autistic individuals compared to normal controls (Grice et al., 2001; O’Connor et al., 2005). This latency delay was also found for mouths or eye regions presented in isolation (O’Connor et al., 2007). However, the classic latency delay of the N170 for isolated eyes compared to faces (Section 4.3) is found in ASD individuals just like in controls, which suggests a delayed rather than abnormal, early processing of faces and facial features in autism. Recently, normal behavior and ERP components (P1, N170 and P2) were reported in autistic children during both implicit and explicit processing of emotional faces (Wong et al., 2008). However, source analysis suggested the strength and dipole orientations of these components which arose from the visual cortex, the FG and the medial prefrontal cortex, were weaker and/or slower in autism than in typically developed children suggesting abnormal functioning of the underlying neural generators in autism despite similar ERPs on surface. Interestingly, parietal responses
were stronger in children with ASDs in that study and could reflect a compensatory analytical strategy to process facial information (Wong et al., 2008).

Neural processing of gaze is also impaired in ASDs. A larger N170 for direct compared to averted gaze has been reported in young autistic children while normal age-matched controls showed no N170 modulation with gaze, just like control adults (Grice et al., 2005). This result could reflect an increased sensitivity to direct gaze very early during visual processing, in agreement with the gaze avoidance hypothesis. Note however, that the N170 in this study was measured at posterior medial sites. At posterior lateral sites where this component is usually maximal, no gaze modulation was seen in any group. This study also involved static stimuli and a passive task. In an explicit gaze direction discrimination task, the change detection in gaze direction was associated with a negative occipito-temporal component (Senju et al., 2005). In normal controls this component was more right lateralized and showed a sensitivity to gaze direction while in ASD children, it showed neither gaze sensitivity nor hemispheric asymmetry. These data suggest an abnormal neural processing of gaze in autism, which is linked to their abnormal perception of social cues and their impairments in ToM.

We have tried to review briefly the impairments in the processing of eyes and gaze in ASDs that seem to be linked to social deficits. We would like to emphasize, however, that most studies focused on high functioning autistic individuals including Asperger Syndrome subjects, rather than on the entire autism spectrum population. One reason for this is that low functioning and severely impaired autistic individuals are extremely hard to test, especially in an MRI scanner or using ERPs. However, scientists should remain careful in the generalization of their findings with these individuals to the entire ASD population. Are face and eye gaze perception related deficits the same in low and high functioning autistic individuals? Or is there a continuum in these impairments across the autism spectrum both behaviorally and at the brain level? For instance, it could be the case that severely impaired subjects simply cannot discriminate gaze direction while AS can but process its social meaning abnormally. An important challenge for future studies will be to bridge the gap between severe autism and Asperger Syndrome using neuroimaging.

6. Conclusions

In this review we hope we have convinced the reader of the importance of eyes and gaze in all aspects of face processing and visual social cognition, including identity and emotion recognition. Eye gaze provides information regarding the attention and objects of interest of others and the ability to process its direction starts very early in life and is fundamental for the development of normal social cognition. Gaze processing is subtended by a large network of non-selective brain areas including the superior temporal sulcus region, the amygdala, the fusiform gyrus, some frontal and some parietal areas. The specific functional connections between these areas is poorly understood but seem to vary as a function of task to allow influences of facial and social context on gaze processing. Future studies will need to better understand these various brain connectivity patterns in order to illuminate gaze impairments in various disorders.

The studies reviewed suggest that impairments in recognizing face identity, facial emotions and understanding others’ mental states could be linked to impairments in extracting the relevant information from the eye region including gaze direction, and this could be a generalized impairment in many clinical populations. ASDs, some cases of prosopagnosia, Capgras delusion and amygdala lesioned patients all have in common an abnormal processing of the eye region which they explore less than normal controls (Section 4.1). The various face processing problems seen in these disorders could thus be due to an abnormal
processing of the eyes and the relevant information necessary to each particular aspect of social interactions (identity cues in prosopagnosia, emotional and gaze cues in amygdala lesions and Capgras delusion, social information and ToM cues in ASD, etc.). This may be due to a dysfunction of the eye detector or to a problem in integrating the eyes into the rest of the face. Finally we would like to emphasize that impairments in face processing in general, and in eye and gaze processing in particular, have also been reported in other pathologies and psychiatric conditions. Schizophrenic individuals for instance present abnormal gaze processing and gaze orienting and have difficulties in extracting emotions and intentions of others from their gaze direction, deficits that seem to be different depending on whether the symptoms of the disease are negative or positive (Hooker and Park, 2005; Langton et al., 2006; Shamay-Tsoory et al., 2007). Individuals with Turner Syndrome (Elgar et al., 2002), Fragile X Syndrome (Garrett et al., 2004) and Fetal Alcohol Spectrum Disorders (Bishop et al., 2007) also present impairments in processing direct gaze.

The extensive review of all the pathologies involving eye and gaze processing abnormalities is beyond the scope of this paper but we suggest impairments in processing these fundamental social cues may represent a core deficiency in many pathologies and a key aspect of social cognition deficits. Determining whether there exists an eye detector mechanism in the human brain and characterizing the spatio-temporal dynamics of its network is one of the future challenges of Cognitive Neuroscience that will help understand better many of these disorders.

Acknowledgments

We wish to warmly thank Dr. Claude Alain from the Rotman Research Institute for pertinent comments and suggestions on previous versions of the manuscript, and Tanya Brown for careful proof-readings of this paper. This work was funded by the Canadian Institute of health Research (CIHR) to RJI and by the French Institut National des Sciences et de la Recherche Médicale (INSERM) to MB.

References


Fox E, Damjanovic L. The eyes are sufficient to produce a threat superiority effect. Emotion. 2006; 6 (3):534–539. [PubMed: 16938095]


Neurosci Biobehav Rev. Author manuscript; available in PMC 2014 February 14.


Itier RJ, Alain C, Kovacevic N, McIntosh AR. Explicit versus implicit gaze processing assessed by ERPs. Brain Res. 2007a; 1177:79–89. [PubMed: 17916340]


Itier RJ, Herdman AT, George N, Cheyne D, Taylor MJ. Inversion and contrast-reversal effects on face processing assessed by MEG. Brain Res. 2006a; 1115 (1):108–120. [PubMed: 16930564]


Fig. 1.
Typical orienting-to-gaze paradigm. A central face cue with averted gaze is presented prior to target onset. Although the cue does not predict the location of the target, subjects respond faster to targets when gaze direction and target location match (congruent trials) and slower when they do not match (incongruent trials).
Fig. 2.
(A–E) Schematic descriptions of the various social situations involving the use of gaze direction. The approximate ages at which the various capabilities emerge are in parenthesis. Adapted from Emery (2000), with permission.
Fig. 3.
The ERP component N170 recorded at a right cerebellar electrode (CB2) for faces and isolated eyes (adapted from Itier et al., 2006b). The N170 is larger and delayed for eyes compared to faces. The respective topographies, representing the voltage distribution on the scalp at the peak of the N170 for each category, are also different, reflecting different underlying generators.
Fig. 4.
The N170 and M170 components obtained after presentation of a face. Topographies are shown at the peak of the components. In one study (Itier et al., 2006a), the M170 recorded with MEG generated a right fusiform gyrus (FG) source and a bilateral source within the inferior and medial occipital gyri (IOG/MOG) when analyzed with the beam former technique event-related Synthetic Aperture Magnetometry (er-SAM). Using the 3D current density method LAURA, Batty and Taylor (2003) and Itier and Taylor (2004a) found that the N170 recorded with ERPs was best modeled by a bilateral source within the STS region. In contrast, in most dipole source analysis such as the one performed by Itier and Taylor (2002) using brain evoked source analysis (BESA), the N170 is often best modeled by dipoles within the FG. These findings led to the hypothesis that approximately 170 ms after a face onset, three different sources are active: the FG, the STS and the IOG/MOG. However, in some cases that remain to be determined, the STS would be best recorded with ERPs due to the proximity of the source to the scalp (underneath temporo-parietal sites) and to a possible radial orientation. The FG and IOG/MOG sources would be best captured with MEG if the sources are tangential. However the sources in the FG may be composed of both tangential and radial components, explaining why sometimes the N170 is modeled by sources in the FG.
Fig. 5.
(A) Intracranial data in humans showed that patches of cortex within the fusiform gyrus (FG, axial view) and superior temporal sulcus (STS, lateral view) are selective for whole faces and for facial components such as eyes (Courtesy of Dr. A. Puce). (B) Effects of inversion on the N170 (shown here at a right cerebellar electrode—CB2) in an orientation discrimination task for isolated eyes, faces, face-without-eyes and houses from Itier et al. (2007b). Horizontal lines show that the N170 amplitude for faces-without-eyes, presented upright or inverted, did not differ from that to upright normal faces. In contrast, inverting full faces increased the N170, which amplitude was not different from that recorded to upright or inverted eyes. This suggests that inverted faces are processed like isolated eyes and that the inversion effect on the N170 is driven by the presence of the eyes. (C) Simplified neural model of early face processing adapted from Itier et al. (2007b). It is assumed that the N170 arises from the FG and the STS region, in which face-selective and eye-selective neurons likely co-exist. The ‘+’ signs signify the neurons are active. Both face- and eye-selective neurons respond to isolated eyes. However, only face-selective neurons respond to faces. The ‘−’ sign followed by a question mark indicates a possible inhibition mechanism from the face neurons onto the eye neurons which would thus not respond to the eyes within an upright face configuration. Regardless of whether they are simply not activated or inhibited by the face neurons, the eye-selective neurons do not respond to the eyes of the face because of the facial context (configuration). When the face is inverted, the facial configuration is disrupted and the eye-selective neurons now respond, just like for isolated eyes, producing the N170 amplitude increase. Note that “face-selective” neurons are sensitive to the face configuration: although they respond to inverted faces and faces-without-eyes, their response is delayed compared to normal full upright faces (see text and Itier et al., 2007b for a complete description).