The role of the pSTS in gaze following and joint attention

Dissertation

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

der Mathematisch-Naturwissenschaftlichen Fakultät

und

der Medizinischen Fakultät

der Eberhard-Karls-Universität Tübingen

vorgelegt

von

Hamidreza Ramezanpour

aus Mashhad, Iran

Oct-2019

Tag der mündlichen Prüfung:	25.11.2019
Dekan der Math.-Nat. Fakultät:	Prof. Dr. W. Rosenstiel
Dekan der Medizinischen Fakultät:	Prof. Dr. I. B. Autenrieth
1. Berichterstatter:	Prof. Dr. Hans-Peter Thier
2. Berichterstatter:	PD Dr. Steffen Hage
Prüfungskommission:	Prof. Dr. Hans-Peter Thier
	Prof. Dr. Uwe Ilg
	Prof. Dr. Martin A. Giese
	PD Dr. Steffen Hage

Erklärung / Declaration:

Ich erkläre, dass ich die zur Promotion eingereichte Arbeit mit dem Titel:

"**The role of the pSTS in gaze following and joint attention**"

selbständig verfasst, nur die angegebenen Quellen und Hilfsmittel benutzt und wörtlich oder inhaltlich übernommene Stellen als solche gekennzeichnet habe. Ich versichere an Eides statt, dass diese Angaben wahr sind und dass ich nichts verschwiegen habe. Mir ist bekannt, dass die falsche Abgabe einer Versicherung an Eides statt mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft wird.

I hereby declare that I have produced the work entitled: "**The role of the pSTS in gaze following and joint attention**" submitted for the award of a doctorate, on my own (without external help), have used only the sources and aids indicated and have marked passages included from other works, whether verbatim or in content, as such. I swear upon oath that these statements are true and that I have not concealed anything. I am aware that making a false declaration under oath is punishable by a term of imprisonment of up to three years or by a fine.

Tübingen,

______________________________	______________________________
Datum/Date	Unterschrift /Signature

Table of Contents

Synopsis

In humans, eye gaze of another person is a powerful stimulus, drawing the observer’s attention precisely to places and objects of interest to the other one. Monkeys also follow gaze, but rather than using eye-gaze, they rely more on peer head orientation to shift attention. Are the human and the monkey gaze following systems functionally related and do they depend on the same anatomical substrates, eventually sharing a common phylogenetic background? How can we voluntarily control our gaze following behavior and eventually suppress it when it is not suitable in a certain context? How does the brain decide which object to attend to in a situation in which the other’s gaze direction seems to point towards more than one object at the same time? What are the properties of single neurons in the putative gaze following region of the brain and which specific computations are they capable of? My thesis tried to address these intimately related questions.

In order to address the first question, we asked human subjects to make saccades to distinct spatial targets, either identified by the eye-gaze of a demonstrator portrait or, alternatively, by associating the demonstrators’ iris color to four spatial targets (gaze following vs. color matching). Using functional magnetic resonance imaging (fMRI) we identified a highly specific region, namely the “gaze following patch (GFP)”, in the posterior superior temporal sulcus (pSTS) activated by gaze following, well separated from those parts of the STS known to process visual information on faces and heads, a finding very similar to previous experiments on rhesus monkeys.

In order to answer the second question, we looked at the brain activity while subjects were required to suppress the gaze following response. In this experiment, we could show that the cognitive control of gaze following was based on the activation of two regions in the frontal cortex, the dorsolateral prefrontal cortex and the anterior cingulate cortex.

In a subsequent experiment, we integrated the need to extract the gaze vector direction and provided *a-priori* information in order to allow the viewer to select only one object out of several met by the same gaze vector. Using fMRI we demonstrated that the disambiguation of the potential object was mainly confined to a region in the inferior frontal junction. Finally, we used monkeys head gaze following as a model for human gaze following and recorded the activity of single neurons from the GFP of two rhesus monkeys

engaged in a battery of tasks in an attempt to understand how the other’s gaze might guide spatial attention to a target shared by the two agents.

We establish that the properties of neurons in the pSTS are indeed able to explain the monkeys’ ability to follow gaze. The fMRI work on the relationship of gaze following and face processing-related activity supports the notion that monkeys’ head gaze following might be homologous to the human eye-gaze following. The fMRI studies on cognitive control and the role of context reveal important features of the network centering on the GFP and the single neuron work on the monkey GFP is able to explain how gaze-related information is translated into shifts of attention by distinct sets of single neurons.

General introduction

Primates have sophisticated social interaction systems that rely primarily on the sense of vision. During social interactions, faces can provide rich information about an individual's age, gender, identity, emotional expression and mental state. In this regard, eyes are probably the most important components of faces, because they offer compelling information on the object and location of interest to the other, drawing the observer's attention to the same object and location, thereby allowing her or him to establish "joint attention". By associating our object-related intentions, expectations, and desires with the other one, joint attention allows us to develop a *theory of* (the other's) *mind* (TOM). Disposing of a viable TOM is a major basis of successful social interactions and arguably its absence is at the core of devastating neuropsychiatric diseases such as autism (Baron-Cohen 1994, 1995; Langton and Bruce 2000; Shimojo, Simion et al. 2003). Hence, in order to understand how the TOM may be dysfunctional in autism, it is very important to study how a healthy brain is able to establish its key building block, that is the gaze following.

Gaze following emerges very early during ontogeny (Shepherd 2010). Already babies at the age of three months show preferential responses to the presence of eyes (Emery 2000). Around the age of two, they follow gaze to objects within their view to establish joint attention. Finally, at an age of around four, they are capable of developing a full-fledged TOM (Emery 2000). At this age, the gaze following and joint attention system have been fully developed and children are even able to follow the gaze towards an object which is not in their visual field by extrapolating the gaze direction of the others (Butterworth and Jarrett 1991). This pattern of development is absent or at least significantly delayed in subjects suffering from autism spectrum disorder (ASD) (Baron-Cohen 1995; Baron-Cohen, Baldwin et al. 1997).

Several studies have suggested that the gaze following has two components with respect to its speed and controllability. One is an extremely fast and reflex-like component which is very hard to suppress (Friesen and Kingstone 1998; Marciniak, Dicke et al. 2015) and a second later component that is more goal-driven and controllable (Ricciardelli, Carcagno et al. 2013). The features of the early gaze following component and the early appearance during development qualify it as a domain-specific cognitive process according to the modularity criteria of Fodor (Fodor 1983). However, until very recently, it was unknown if also the third criterion of Fodor modularity requirements was met, namely if gaze following had a specific neural architecture, separate from the one for the processing of nonsocial stimuli for orienting attention or other, more general-purpose networks. Precise localization of the relevant machinery and its relationship to other forms of attentional orienting systems has recently been provided by a number of functional MRI (fMRI) studies. Those works have identified a circumscribed region in the posterior superior temporal sulcus (pSTS) of both hemispheres, adjacent to the middle and superior temporal gyri, often referred to as pSTS region or area or, more loosely, the gaze-following patch (GFP) (Puce, Allison et al. 1998; Allison, Puce et al. 2000; Hoffman and Haxby 2000; Pelphrey, Morris et al. 2004; Materna, Dicke et al. 2008; Laube, Kamphuis et al. 2011), the latter emphasizing a character reminiscent of the well-known face patches in the STS (Tsao, Moeller et al. 2008; Tsao, Schweers et al. 2008). The gaze following-associated neural responses found in the previous studies were dissociable from those evoked by shifts of attention based on nonsocial cues such as arrow (Callejas, Shulman et al. 2014). Furthermore, patients with a temporal cortex lesion including the putative GFP have been shown to be unable to use the other's eye gaze to shift their gaze while still being able to use directional information provided by an arrow (Akiyama, Kato et al. 2006a, 2006b). Altogether these studies confirm the notion that the GFP might be actually "the" domain-specific module which computes directional social information coming from faces to guide the observers' attention. This cognitive module is hypothesized to be evolutionary very old and shared between humans and their primate ancestors (Emery 2000). The notion that the GFP may be a homologous structure subserving as the basis of this cognitive module, shared within the primate order, is supported by observations that demonstrate that monkeys follow gaze in a similar way to humans and that monkey show gaze following-related BOLD activity in a similar region. Monkeys' gaze following has been shown to be able to operate very fast and automatic and develop early after birth, similar to humans (Shepherd 2010; Marciniak, Dicke et al. 2015). This general correspondence not withstanding, there may also be differences between old world monkeys and man. Probably the most obvious difference is the source of directional social cues used to follow gaze and to establish joint attention. While the clear contrast between the human iris and sclera allows detection even from a considerable distance and to pinpoint the other’s focus of attention, the majority of primates have very dark sclerae and pupils as well as less elongated eye shapes, making it almost impossible to detect their conspecifics’ eye gaze direction changes from more than a few meters (Kobayashi and Kohshima 1997, 2001). Hence, monkeys are thought to rely much more on the other’s head and eventually also body orientation. Nevertheless, comparable developmental time scales and very similar anatomical substrates suggest that key features of the underlying neural mechanisms might be similar in monkeys and humans if not identical.

The extraction of eye gaze orientation requires knowledge of the orientation of the eyes relative to the face and ultimately knowledge about the orientation of the other’s face relative to the observer and the world to establish a stable frame of reference. The need to care about particular aspects of faces might suggest that gaze following may build on the information provided by the parts of the cortex known to be devoted to the processing of faces, including their constitutive elements such as the eyes.

In fact, the human GFP, lighting up in gaze perception tasks, seems to be located in close vicinity to face-selective areas in the ventral visual cortex and eventually overlapping with neighboring face patches. This raises the possibility that the GFP may actually be one of the members of this face-processing network that involves distinct elements in the ventral visual cortex and the frontal cortex, namely the occipital face area (OFA), the fusiform face area (FFA), the STS face area (STS-FA), and the inferior frontal gyrus face area (IFG-FA) (Kanwisher, McDermott et al. 1997; Haxby, Hoffman et al. 2000; Tsao, Moeller et al. 2008). These areas are interconnected and seem to be devoted to processing particular aspects of faces. One important question which needs to be addressed here is: Could it be that the GFP is actually part of the machinery for face processing rather than being confined to converting directional information on face orientation to precise spatial coordinates?

Human gaze following is geometric. This means that we use the other´s gaze vector to identify the exact location of the object of interest. As said earlier, the unique morphology of human eyes allows us to determine the direction of the eye at high resolution. However, most of the time the knowledge of the direction alone is not sufficient to pinpoint an object in 3D. In principle, differences between the directions of the two eyes, i.e. knowledge of the vergence angle, could be exploited to triangulate the object position. Yet, this will work only for objects close to the beholder as the angle will become imperceptibly small if the objects are located far with respect to the beholder or too close with respect to each other. Hence, how is the brain able to disambiguate this situation and select only one object?

High saliency of gaze cues let us always feel an urge to follow the other's gaze. However, we are able to control gaze following at least to some extent if alternative behaviors may be more pertinent in a given moment. What is the source of the control signals? Is gaze following under cognitive control exerted by prefrontal signals similar to many other functions (Miller and Cohen 2001; Aron, Robbins et al. 2004; Ridderinkhof, van den Wildenberg et al. 2004)?

As said earlier, nonhuman primates are social animals and communicate based on nonverbal cues much like humans. The question of whether or not nonhuman primates, dispose of a full-fledged TOM similar to humans is debatable. But as said before, there is a substantial evidence that the key building block for the development of a TOM, the ability to follow the other's gaze in a very similar way to humans is available (Tomasello, Hare et al. 1999; Emery 2000; Tomasello and Carpenter 2005; Tomasello, Hare et al. 2007). I already alluded to the demonstration of a cortical area revealed by blood-oxygen-level-dependent (BOLD) imaging of the monkey brain and activated by head gaze-following, the monkey GFP, having a position in the STS reminiscent of the location of the human GFP and eventually homologous to it (Marciniak, Atabaki et al. 2014). However, knowing the location of a module and knowing that it is activated in a task-dependent manner tells one very little about the underlying neuronal information principles. The most important questions here are: first, how can GFP neurons dissociate a simple perception of the other's gaze direction from a truly geometrical gaze following response towards certain spatial targets identified by an eye movement of the observer? Second, are GFP neurons able to show modulation by cognitive signals such as rules to enable the GFP machinery to undergo executive control in case gaze following-responses need to be suppressed? Third, is there any clear topographical distinction between the cluster of gaze following-neurons and those involved in the passive perception of faces located nearby? And finally, and most importantly, how specific are the neuronal responses to shifts of attention guided by gaze cues in comparison to other sources of information such as other social cues and learned arbitrary associations between cues and spatial targets? Answers to all the questions presented in this introduction will nominate the GFP as a domain-specific cortical module implicated in the control of social interactions based on other’s gaze signals and allow us to shed light on the processes allowing this module to convert gaze information into joint attention. Finding answer will eventually help to pave the road to a better understanding of the pathophysiology of disturbances of social interactions.

Aims of this dissertation

This dissertation addresses a key aspect of our ability to interact with others non-verbally, our ability to establish ‘joint attention’ with the other by following the other’s gaze. Joint attention is a key step towards developing a TOM. In order to establish joint attention, we rely on the other’s gaze to identify her/his object of attention, a capacity which we share with nonhuman primates. Previous work has delineated a network of cortical “patches” in the primate cortex, processing faces, eventually also extracting information on the other’s gaze direction. Yet, the neural mechanism that links information on gaze direction, guiding the observer’s attention to the relevant object has remained elusive.

In order to reveal the neural mechanisms affording gaze following and joint attention and its relationship to the previously described ‘face patch’ system, I and the colleagues, who joined forces with me, ran several fMRI experiments on humans and an electrophysiological investigation of rhesus monkeys.

First, we localized the main brain region implicated in eye-gaze following in humans and its anatomical relationship to the previously known ‘face patch’ system in the temporal cortex (1st study, Chapter 1).

Second, we tried to reveal the neural substrates allowing disambiguation of the object of joint attention when the gaze vector points towards several potential objects, deploying fMRI (2nd study, Chapter 2).

Third, in an event-related fMRI experiment, we identified the neural network that is important for volitional control of gaze following responses (3rd study, Chapter 3).

Finally, we studied the neural underpinnings of monkey head-gaze following as a model of the human eye-gaze following at the level of single neurons in a distinct region in the STS, the “gaze following patch” (GFP) (4th study, Chapter 4).

Statement of contributions

Study 1 (Chapter 1):

K. Marquardt*, **H. Ramezanpour***, P. W. Dicke, P. Thier

Following eye gaze activates a patch in the posterior temporal cortex that is not part of the human “face patch” system. eNeuro, 2017, doi:10.1523/ENEURO.0317-16.2017 *(*equal contribution)*

HR, KM, PWD, and PT designed research; KM and PWD performed research; KM and **HR** analyzed data; KM, **HR**, and PT wrote the paper.

Study 2 (Chapter 2):

P. Kraemer*, M. Görner*, **H. Ramezanpour***, P. W. Dicke, P. Thier
A fronto-temporo-parietal network disambiguates potential objects of joint attention. Under Review in Cerebral Cortex and on BioRxiv, **doi:** https://doi.org/10.1101/542555, *2019*. *(*equal contribution)*

HR, PT and PK designed the experiments. PK and PWD performed the experiments. PK and MG analyzed the data. All authors contributed to the interpretation of results and the writing.

Study 3 (Chapter 3):

M. S. Breu*, **H. Ramezanpour***, P. W. Dicke, P. Thier
A neural substrate for volitional control of gaze following. Under Review in Neuropsychologia, 2019. *(*equal contribution)*

HR and PT designed the study; MSB and PWD collected the data. MSB and **HR** analyzed the data. MSB, **HR** and PT wrote the manuscript.

Study 4 (Chapter 4):

H. Ramezanpour, P. Thier

Decoding of the others’ focus of attention by a temporal cortex module. Under Review in PNAS and on BioRxiv, **doi:** https://doi.org/10.1101/681957, 2019.

HR and PT designed the experiments, interpreted the results and wrote the paper. **HR** performed the experiments and analyzed the data.

Chapter 1

Following eye gaze activates a patch in the posterior temporal cortex that is not part of the human “face patch” system

Kira Marquardt*, Hamidreza Ramezanpour*, Peter Dicke, Peter Thier (*equal contribution)

eNeuro, 2017, doi:10.1523/ENEURO.0317-16.2017

We compared the pattern of blood oxygenation level-dependent (BOLD) imaging contrasts reflecting the passive vision of static faces with the one evoked by shifts of attention guided by the eye gaze of others in the same set of subjects. The viewing of static faces revealed the face patch system. On the other hand, eye gaze-following activated a cortical patch (the GFP) with its activation maximum separated by more than 24 mm in the right and 19 mm in the left hemisphere from the nearest face patch, the STS face area (FA). This clear segregation indicates that the GFP accommodates a functionality not found in the face-selective areas, although most probably building on pertinent information contributed by the latter.

Cognition and Behavior

Following Eye Gaze Activates a Patch in the Posterior Temporal Cortex That Is not Part of the Human "Face Patch" System

Kira Marquardt,[1,*] Hamidreza Ramezanpour,[1,2,3,*] Peter W. Dicke,[1] and Peter Thier[1,4]

DOI:http://dx.doi.org/10.1523/ENEURO.0317-16.2017

[1]Department of Cognitive Neurology, Hertie Institute for Clinical Brain Research, 72076 Tübingen, Germany, [2]Graduate School of Neural and Behavioural Sciences, University of Tübingen, 72074 Tübingen, Germany, [3]International Max Planck Research School for Cognitive and Systems Neuroscience, University of Tübingen, 72074 Tübingen, Germany, [4]Werner Reichardt Centre for Integrative Neuroscience (CIN), University of Tübingen, 72076 Tübingen, Germany

Abstract

Humans follow another person's eye gaze to objects of interest to the other, thereby establishing joint attention, a first step toward developing a theory of the other's mind. Previous functional MRI studies agree that a "gaze-following patch" (GFP) of cortex close to the posterior superior temporal sulcus (STS) is specifically implicated in eye gaze-following. The location of the GFP is in the vicinity of the posterior members of the core face-processing system that consists of distinct patches in ventral visual cortex, the STS, and frontal cortex, also involved in processing information on the eyes. To test whether the GFP might correspond to one of the posterior face patches, we compared the pattern of blood oxygenation level–dependent (BOLD) imaging contrasts reflecting the passive vision of static faces with the one evoked by shifts of attention guided by the eye gaze of others. The viewing of static faces revealed the face patch system. On the other hand, eye gaze-following activated a cortical patch (the GFP) with its activation maximum separated by more than 24 mm in the right and 19 mm in the left hemisphere from the nearest face patch, the STS face area (FA). This segregation supports a distinct function of the GFP, different from the elementary processing of facial information.

Key words: face patch; gaze-following patch; joint attention; posterior superior temporal sulcus

Significance Statement

Human observers follow another person's eye gaze to objects and locations of interest to the other, thereby establishing joint attention, a major step toward developing a theory of the other's mind. Previous functional MRI (fMRI) studies agree that a patch of cortex around the posterior superior temporal sulcus is specifically implicated in eye gaze-following. This gaze-following patch is located in the same region as the posterior elements of the face patch system, also extracting information on the eyes. Using fMRI, we show that the gaze-following patch is distinct from the face patch system, supporting a role beyond the elementary processing of facial information accommodated by the face patch system.

Introduction

Eye gaze, head, shoulder, and trunk orientation are important examples of body cues that offer compelling information on the object and location of interest to the other, drawing the observer's attention to the same object and location, thereby establishing "joint attention," a first

Received October 21, 2016; accepted February 20, 2017; First published March 09, 2017.
Authors report no conflict of interest.

Author Contributions: KM, HR, PWD, and PT designed research; KM and PWD performed research; KM and HR analyzed data; KM, HR, and PT wrote the paper.

and major step toward developing a theory of the other's mind (Baron-Cohen, 1994, 1995; Emery, 2000; Langton and Bruce, 2000; Shimojo et al., 2003). In humans, eye gaze is arguably the most important social cue guiding the observer's attention (Emery, 2000). A precise localization of the relevant machinery has recently been provided by a number of functional MRI (fMRI) studies. This work has identified a circumscribed region in the posterior superior temporal sulcus (pSTS) of both hemispheres, adjacent to the middle and superior temporal gyri, often referred to as pSTS region or area or, more loosely, the gaze-following patch (GFP; Puce et al., 1998; Allison et al., 2000; Hoffman and Haxby, 2000; Pelphrey et al., 2003, 2004; Materna et al., 2008a; Laube et al., 2011). A cortical area involved in macaque monkeys' head gaze-following, the monkey GFP, has recently been described in a comparable cortical region that may eventually turn out to be homologous with the human GFP in the pSTS (Marciniak et al., 2014).

The extraction of eye gaze orientation requires knowledge of the orientation of the eyes relative to the face and ultimately also knowledge about the orientation of the other's face relative to the observer and the world. This need to care about particular aspects of faces might suggest that eye gaze-following may build on information provided by the parts of cortex known to be devoted to the processing of faces, including their constitutive elements such as the eyes. Actually, this influence of the eyes is suggested by a number of studies that have demonstrated that, for instance, information on identity and emotional expression, provided by the eye region, influences not only perception but also the activity in distinct face patches (Fox and Damjanovic, 2006; Chan and Downing, 2011).

In fact, the human GFP, lighting up in gaze-perception tasks, is located in close vicinity to face-selective areas in the ventral visual cortex. This raises the possibility that the GFP may actually be one of the members of this face-processing network that involves distinct elements in the ventral visual cortex and frontal cortex, namely the occipital face area (OFA), the fusiform face area (FFA), the STS face area (STS-FA), and the inferior frontal gyrus face area (IFG-FA; Kanwisher et al., 1997; Haxby et al., 2000; Tsao et al., 2008). These areas are interconnected and seem to be devoted to particular aspects of faces. For instance, the FFA emphasizes the encoding of constant aspects of the face underlying identity decisions (Grill-Spector et al., 2004). On the other hand, the STS-FA, the face-selective area closest to the known location of the GFP, has been shown to contribute to encoding changeable aspects of faces such as facial expression and face orientation, the latter an aspect obviously important for gaze-following (Puce et al., 1998; Wicker et al., 1998). Could it be that the STS-FA is actually part of the machinery for gaze-following, rather than being confined to providing information on face orientation? In this case, we would expect at least partial overlap between the GFP and the STS-FA. In view of the interindividual variability in the location of the GFP and also the STS-FA, the question whether the two overlap requires testing the same subjects in gaze-following and face-perception tasks. Using well-controlled fMRI paradigms in the same set of subjects, we show that the two systems are actually well segregated, a finding that clearly indicates that the GFP accommodates a functionality not found in the face-selective areas, although most probably building on pertinent information contributed by the latter.

Material and Methods

Subjects and instrumentation

Eleven adult male and nine adult female subjects, age range 21–46 years (mean 26, SEM 5.5 years) participated in the current study. All participants were right-handed and healthy and had normal or corrected-to-normal vision. Subjects were provided with transparent and comprehensible information about the study goals and the procedures involved and gave their written consent. Participants ran a training behavioral session before an imaging session to minimize errors inside the MRI scanner caused by potential misunderstanding of task rules or a lack of practice. The study was approved by the Ethics Review Board of Tübingen Medical School and was conducted in accordance with the principles of the 1964 Declaration of Helsinki.

In the training session, subjects' eye movements were recorded deploying a commercial Eye Tracker (Chronos Vision C-ETD). During the imaging session, subjects' heads were stabilized by foam rubber to minimize residual head movements. The visual stimuli (32° × 24° visual angle) were presented on a translucent screen using an LCD projector (NEC GT 950, 1024 × 768 pixels) viewed by the subjects via a two-mirror system with 60-cm distance between the translucent screen and the subjects' eyes. During the imaging procedure, a certified, MRI-compatible Eye Tracker (SMI iView X MRI-LR) was used to record the subjects' eye movements. The recorded eye movements were evaluated offline after the experiments.

Visual stimuli and experimental tasks

The participants had to perform three tasks. The first task required the observer to extract the portrait's eye-gaze direction and make a saccade toward one of a set of five spatial targets that the portrayed demonstrator looked at (gaze-following task). The second task also required an indicative saccade to targets singled out by

This work was supported by the Werner Reichardt Centre for Integrative Neuroscience (CIN) at the Eberhard Karls University of Tübingen. The CIN is an Excellence Cluster funded by the Deutsche Forschungsgemeinschaft (DFG) within the framework of the Excellence Initiative (EXC 307). Further support was granted by the German Research Foundation (DFG) project TH 425/12-2.

*KM and HR contributed equally to this work.

Correspondence should be addressed to either of the following: Peter Thier, Department of Cognitive Neurology, Hertie Institute for Clinical Brain Research, Hoppe-Seyler-Str. 3, 72076 Tübingen, Germany. E-mail: thier@uni-tuebingen.de; or Hamidreza Ramezanpour, Department of Cognitive Neurology, Hertie Institute for Clinical Brain Research, Hoppe-Seyler-Str. 3, 72076 Tübingen, Germany. E-mail: hamidreza.ramezanpour@gmail.com.

DOI:http://dx.doi.org/10.1523/ENEURO.0317-16.2017

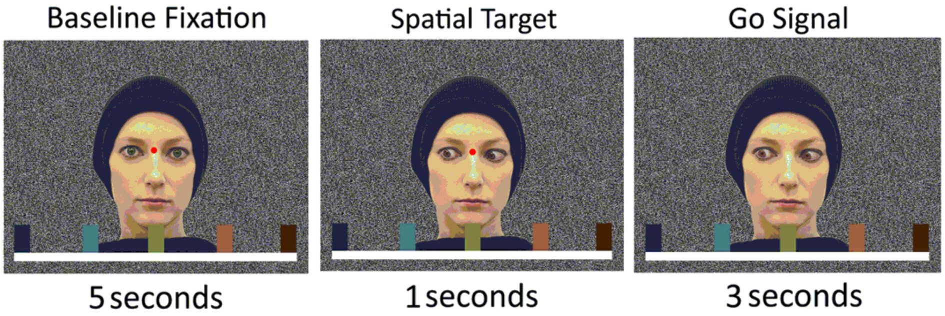

Fig. 1. Sequence of visual stimuli in the active task. At the beginning of each block of trials, a written instruction (either gaze-following or color-matching) was presented on the screen for 5 s. Each trial started with a baseline fixation picture with direct gaze (lasting for 5 s), immediately followed by one of five possible portraits (target portraits), present for 4 s, with the demonstrator's gaze directed at a specific target and exhibiting a distinct iris color. Subjects were not allowed to make an eye movement until the disappearance of the fixation target. Afterward, alternately 10 fixations (each 5-s duration) and 10 trial pictures (each 4-s duration) were presented. The demonstrator has agreed for her portrait to be published.

information provided by the same demonstrator portraits; however, in contrast to the first task, a different rule applied. Rather than following the demonstrator's gaze, the observer was required to make a saccade to the target that had the same color as the portrayed demonstrator's iris (color-matching task). Note that the visual information provided in the two tasks was the same, i.e., in both tasks the iris color varied from trial to trial, adopting the distinctive color of one of the five targets, arranged on a horizontal line met by the demonstrator's gaze axis. Using the same visual stimuli for the gaze-following task and the control task and requiring the same behavioral responses, any differences in the associated blood oxygenation level–dependent (BOLD) imaging responses would have to be caused by differences between the cognitive operations induced by the two sets of cues. Finally, participants were subjected to a third experiment that required fixation on a small dot while passively viewing images of faces and nonface stimuli, centered on the fixation dot, not requiring any behavioral response (passive face perception task).

The portraits used in the gaze-following/color-matching tasks (collectively referred to as the "active tasks") were photographs of a female in front of a white background. She was looking either straight into the camera (baseline fixation picture) or to one of five dot targets arranged on a horizontal board, 25° below the straight-ahead axis in the fronto-orthogonal plane, with a visual angle of 12.5° between targets. The digital photographs were processed using Adobe Photoshop CS5 to replace the original background with a black-and-white random dot pattern and color the portrait's iris and the five targets with five different colors (dark blue, light blue, green, light brown, and dark brown).

The tasks were run in separate blocks. Each block started with a written task instruction on the projection screen (either gaze-following or color-matching) present for 5 s. The whole block lasted for 95 s and contained 10 trials. Each trial started with a baseline fixation picture with direct gaze (lasting for 5 s), immediately followed by one of five possible portraits (target portraits), present for 4 s, with the demonstrator's gaze directed at a specific target and exhibiting a distinct iris color. Within one block, these 10 trials were sorted randomly. The whole experiment contained four sessions, each involving two blocks of gaze-following and two of color-matching, one after another.

During the presentation of the baseline fixation picture, the subjects were asked to fixate on a small dot with 0.3° visual angle radius presented between the demonstrator's eyes oriented straight ahead. This fixation dot was also present in the target portraits for the first 1 s and then turned off. The disappearance of the fixation point served as the "GO" signal for the participants to perform their saccade to the target singled out by the prevailing rule (gaze-following vs. color-matching). The subjects had to stay with their eye-gaze on the chosen target until the baseline fixation picture, now serving as "GO" signal, appeared again (see Fig. 1 and Fig. 2). Implementing this "GO" signal seemed to be necessary to allow us to reveal differences in BOLD signals between gaze-following and color-matching. Otherwise, possibly dominating BOLD signals associated with undelayed saccades might have concealed the differential BOLD activity associated with the preceding processes.

The stimuli deployed in the passive face perception task (in short, "passive task") were photographs of human faces (females and males), hands, and bodies plus man-made objects of daily life as well as food, each subtending a 12° visual angle. Facial stimuli were taken from the Radboud Face Database (Langner et al., 2010), showing females and males with averted gazes. Adobe Photoshop CS5 was used to create scrambled versions of all photographs and replace the backgrounds with the same black-and-white random dot background used in the gaze-following paradigm. Stimuli were presented in four sessions, each containing 10 blocks of 16 photographs. The sequence of blocks was the same in each session, but photographs were randomly distributed within a block. Each block lasted 38 s and started with the presentation of a fixation dot in front of a black-and-white random dot background for 5 s, followed by 16 photographs (presentation time, 1 s each) with black screens present for 0.2 s in between. During presentation, subjects were asked to maintain fixation on a small dot in the middle of the screen while viewing the photographs.

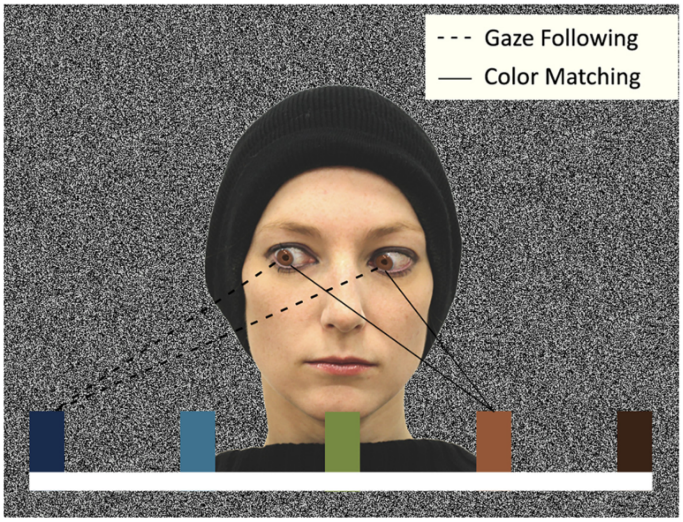

Fig. 2. Illustration of the first experiment's stimulus. The eyes of the person are directed to the dark-blue target (gaze cue), but the person's iris color corresponds to the light-brown target (color cue). According to the introduced condition at the beginning of the block, the subject would have to make a saccade toward the dark-blue target (gaze-following condition) or toward the light-brown target (color-matching condition). The demonstrator has agreed for her portrait to be published.

MRI imaging and preprocessing

A 3 Tesla MR-Scanner (Siemens Magnetom Trio Tim syngo MR B17) was used to scan subjects' brains. We used a T2*-weighted echo-planar sequence (TE, 35 ms, TR, 3000 ms; flip angle, 90°) covering the whole brain (44 transverse slices; matrix 64 × 64; slice thickness, 2.5 mm; in-plane resolution, 3 × 3) for image acquisition during the experiments. We used a T1-weighted, magnetization-prepared, rapid-acquisition gradient-echo sequence (MP-RAGE with TE, 2.92 ms; TR, 2300 ms; TI, 1100; flip angle, 8°; 176 × 256 × 256 voxel; voxel size, 1.0 × 1.0 × 1.0 mm) for the structural, anatomic scans. A total of 945 images were taken from each subject.

The preprocessing and analysis of the images was done with the statistical parametric mapping program package SPM8 (Wellcome Department of Cognitive Neurology, London, UK, http://www.fil.ion.ucl.ac.uk/spm) running on Matlab 2013. Images of each subject were reoriented by setting the origin to the anterior commissure and correcting for slice time (number of slices, 44; TR, 3 s; TA, 2.93; slice order, interleaved descending; reference slice, 22). Functional scans were spatially realigned (registered to first and mean images resliced). The anatomic scan was coregistered to the mean volume of the functional images and was normalized to the Montreal Neurologic Institute space (Friston et al., 1995). Functional images were normalized to the anatomic scan and then smoothed using a 7-mm full-width half-maximum Gaussian filter. Time series in each voxel were high pass–filtered with a cutoff frequency of 1/128 Hz.

MRI data analysis

To estimate the BOLD activation patterns associated with the experimental tasks, we assumed a standard hemodynamic response function, reflecting the task variables according to a general linear model (GLM). In the active task, the onset of the portrait defined time 0 of the ensuing event trace. We distinguished three different event types: fixation, gaze-following, and color-matching. In the passive task, the appearance of the first image in each block determined time 0 of an event trace spreading across the whole block.

The estimated head movements of the subjects during the sessions were considered as regressors of no interest in the GLM in addition to covariates of interest (experimental conditions: fixation, gaze-following, color-matching, faces, and nonfaces). For the active tasks, the following contrasts were calculated for each subject: response to gaze-following and color-matching versus baseline fixation and response to gaze-following versus color-matching and vice versa. For the passive task, contrasts between responses to faces and all nonface stimuli including the scrambled faces were calculated. *t*-statistics were used to identify significant changes ($p < 0.0001$ for the active task and a more conservative threshold of $p < 0.001$ for the passive task, taking into account its lower statistical power) in the BOLD signal at the level of individual subjects. To test whether results obtained for individual subjects are valid at the population level, we conducted a second-level analysis, deploying a random-effects model, comparing the average activation for a given voxel with the variability of that activation over the examined population (Friston et al., 1999). The average activation for a given voxel was taken as significant if the probability *p* provided by *t*-statistics fell below 0.0001 (uncorrected) for that voxel and in at least six neighboring ones. To optimally visualize and measure the cortical representations, statistical *t*-maps were projected onto inflated and flattened reconstructions of cortical surface gray matter using Caret (http://brainvis.wustl.edu/wiki/index.php/caret).

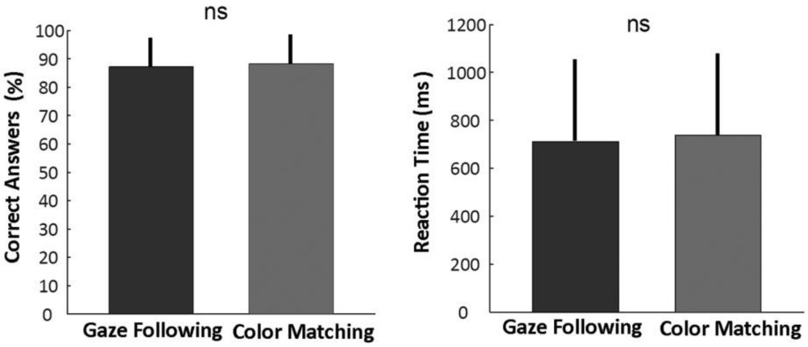

Fig. 3. Behavioral data for gaze-following (dark gray) and color-matching (light gray) showing no significant difference in either mean accuracy or mean reaction time (time between the go signal and the start of the eye movement (n = 20 sessions, 160 correct trials). Error bars represent SE.

Results

Behavioral findings

In the active experiment, participants were instructed to identify the target either by following the portrait's eye-gaze (gaze-following) or, alternatively, to identify it based on a color match with the iris of the portrayed demonstrator and execute a saccade to the target. In the first case, eye color, and in the second case, eye gaze direction, had to be discounted. The two variants of the active task did not differ with respect to the visual information available or the oculomotor behavior prompted but with regard to the cognitive strategy required to solve the task. One might argue that the two different strategies to be pursued might have been associated with different levels of difficulty and, consecutively, also different subjective task loads. This did not seem to be the case, as task performance was very similar. Participants performed the task in the scanner with high accuracy well above chance level (20%) in the gaze-following condition (correct responses: mean, 87%; SEM, 11%) as well as in the color-matching condition (correct responses: mean, 88%; SEM, 11%). Kolmogorov–Smirnov test showed that reaction times and correct responses showed a normal distribution. A paired t-test showed no significant difference in the number of correct responses (p = 0.61) or reaction times (p = 0.32) between the two conditions (gaze-following reaction time: mean, 711 ms; SEM, 366 ms; color-matching reaction time: mean, 736 ms; SEM, 341 ms; Fig. 3).

BOLD responses to gaze-following and color-matching

To identify brain areas activated during gaze-following, we looked at the contrast of gaze-following versus baseline fixation in a second-level analysis of the group data. This comparison delineated several brain areas in both hemispheres that had a significantly higher BOLD signal ($p < 0.0001$, in a cluster of six connected voxels each; see Fig. 4), among them dorsolateral prefrontal cortex, pre-

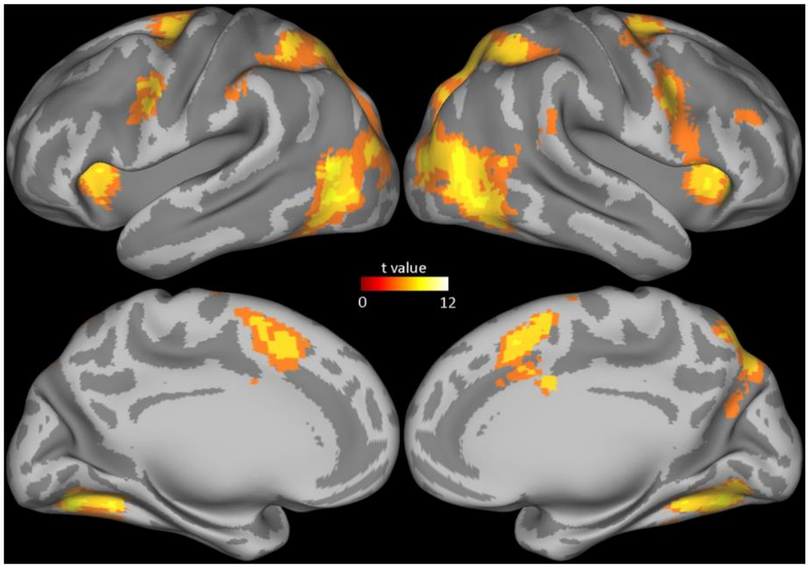

Fig. 4. MRI group data showing the BOLD response for the contrast gaze-following versus baseline fixation.

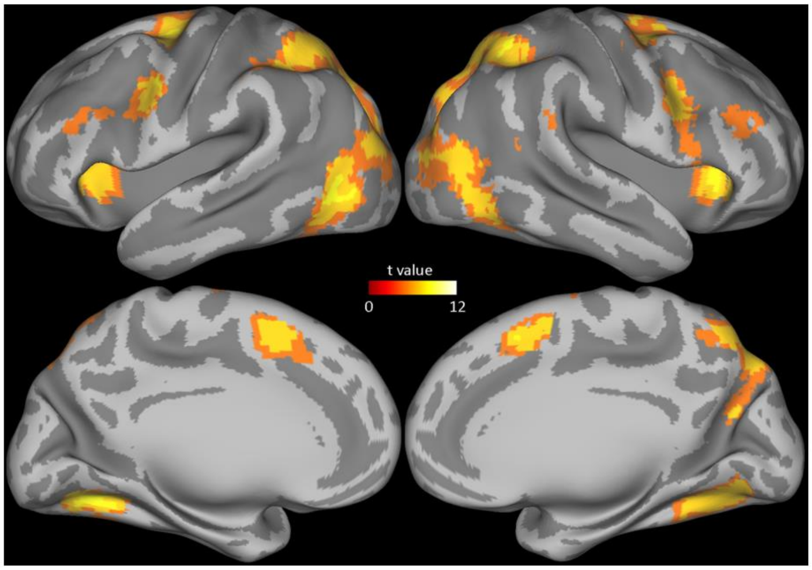

Fig. 5. MRI group data showing the BOLD response for the contrast color-matching versus baseline fixation.

motor cortex, supplementary motor area, cuneus, precuneus, fusiform gyrus, posterior middle temporal gyrus, inferior temporal gyrus, middle occipital gyrus, clustrom, middle frontal gyrus, inferior parietal lobule, superior parietal lobule, supramarginal gyrus, precentral gyrus, cingulate gyrus, superior frontal gyrus, lingual gyrus, superior occipital gyrus, parahippocampal gyrus, and cerebellum. This pattern was very similar to the one obtained when calculating the color-matching versus baseline fixation contrast (Fig. 5). The close, qualitative match between the patterns associated with the two tasks is not unexpected, given that both require the extraction of specific cues from faces to localize distinct objects to shift one's attention to them.

To identify cortical regions specifically or more strongly activated by the need to exploit gaze direction, we calculated the BOLD contrast between gaze-following and color-matching. A significant contrast (statistical criteria as before) was found in a patch of cortex bilaterally in the posterior part of the middle and inferior temporal gyrus, specifically with the peak contrast at Talaraich coordinates right (50, −64, 2) and left (–54, −67, 6; see Fig. 6). This location of activity is similar to gaze-following– and gaze-processing–related activity found in previous fMRI studies (Hoffman and Haxby, 2000; Hooker et al., 2003; Pelphrey et al., 2005b; Materna et al., 2008a). We refer to the activated patch as the gaze-following patch (GFP) and the cortical region in which it is located as the pSTS.

BOLD responses to the passive vision of faces

We identified cortex activated by the passive vision of static faces by delineating regions for which the contrast faces versus nonface objects (biological as well as nonbiological objects and scrambled faces were pooled) was significant in the second-level analysis ($p < 0.001$, uncorrected, six connected voxels). In accordance with previous studies (Ishai et al., 2005; Gobbini and Haxby, 2006; Fox et al., 2009), we found significant BOLD contrasts in the mid-fusiform gyrus bilaterally (these voxels are the FFA), the right inferior occipital gyrus (these voxels form the OFA), the posterior superior temporal sulcus bilaterally (these voxels correspond to the STS-FA), and the right inferior frontal gyrus (these voxels make up the IFG-FA). The highest BOLD contrast to faces was identified in the

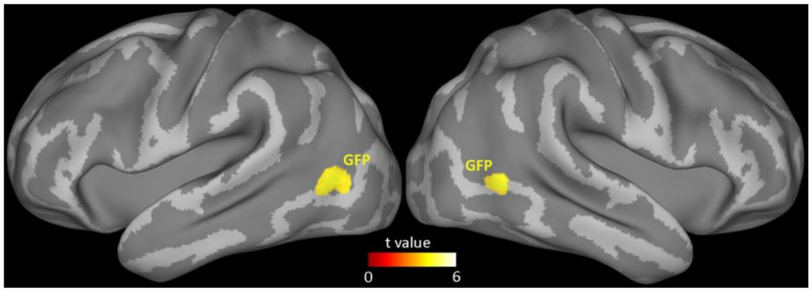

Fig. 6. MRI group data showing the BOLD response for the contrast gaze-following versus color-matching. Activation maximum in right hemisphere in Talaraich coordinates (50, −64, 2).

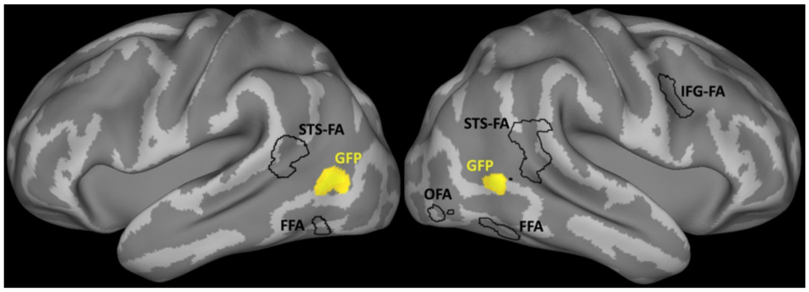

Fig. 7. Spatial organization of face-selective areas and the gaze-following patch.

functionally defined STS-FA, located at Talaraich coordinates right (51, −42, 12) and left (–57, −48, 8). After identifying the face-selective regions in the second-level analysis, the BOLD time series underwent spatial smoothing with an 8-mm FWHM Gaussian blur, and the clusters of face-selective regions were extracted as a mask mapped on the cortical surfaces to assess their spatial relationship to the GFP later on.

The pSTS gaze-following patch and the face patch are segregated

The fact that the GFP and the STS-FA, exhibiting the strongest BOLD contrast, were found in the same posterior part of the STS suggested that the two might overlap or, eventually, be even fully congruent. To investigate this possibility, we projected the two GFP and the face patches, including the one in the pSTS region, onto an inflated 3D representation of cortical surfaces using the PALS-B12 atlas of human cerebral cortex (Van Essen, 2005). This rendering did not exhibit any indication of overlap between the gaze-following patch and any of the face-selective regions. Actually, the boundaries of the GFP and the ones of the nearest STS-FA were separated by a gap of 4 mm (Fig. 7). We next defined the GFP and the STS-FA as our regions of interest (spheres with the diameter of 5 mm centered on the coordinates of the peak activities in these two areas in each individual subject to compare the response levels as captured by the contrast values for passive perception of static faces and gaze-following. As shown in Fig. 8, the average contrast values in the GFP for the passive face perception task did not differ significantly from zero (*t* test, $p = 0.20$), meaning that there was no selectivity for faces. Likewise, the mean contrast values in the STS-FA during gaze-following did not differ significantly from zero (*t* test, $p = 0.49$), correspondingly expressing a lack of selectivity to gaze-following. Hence, we may conclude that the GFP and the STS-FA are neighboring, yet nonoverlapping, areas with different functions. In six of 20 subjects, we could not delineate a significantly activated GFP and STS-FA at the level of the individual. Hence, these six subjects had to be excluded from a comparison of gaze-following–related activity with activity in individual delineated STS-FA and vice versa; i.e., face-selectivity test in the GFP. We also resorted to a conjunction analysis as an alternative to a random-effect analysis (Heller et al., 2007). This approach allows the assessment of how many subjects exhibit selective activations in each voxel and therefore shows the extent of overlap between gaze-following–related activity and activity evoked by static faces within and across subjects. This analysis did not show any overlap in individual subjects, passing the significance threshold of $p < 0.001$ (uncorrected).

Discussion

With two separate fMRI experiments, performed on the same subjects, we tried to map the cortical areas underlying gaze-following and the establishment of joint attention and/or the passive perception of static faces. The two experiments were run on the same subjects to find out whether the cortical structures involved overlap. In the first experiment, consisting of two tasks, subjects were asked to either follow the eye gaze direction of portrayed demonstrators toward distinct spatial targets or, alternatively, to shift attention to the target whose color corresponded to that of the demonstrator's iris. In accordance with previous work (Materna et al., 2008a), we found that a GFP lighted up bilaterally in the posterior part of the middle temporal gyrus when the BOLD signal evoked by eye gaze-following was contrasted with the BOLD signal evoked in the color-matching condition. Assuming that this contrast is able to eliminate activity due to visual

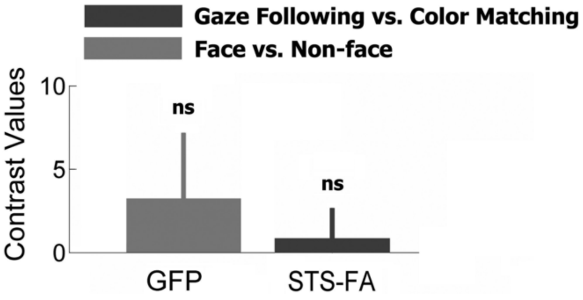

Fig. 8. Selectivity of the individually defined STS-FA to gaze-following in contrast to the selectivity of the GFP to static face perception. Error bars indicate 90% confidence intervals. In the right STS-FA (Talaraich coordinates of the peak: 51, −42, 12), the mean contrast value for gaze-following is not significantly different from zero, in accordance with the assumption of a lack of gaze-following selectivity (*t*-test, $p = 0.49$). On the other hand, the contrast value for face perception in the right GFP (Talaraich coordinates of the peak: 50, −64, 2) is not significantly different from zero, meaning no face-selectivity (*t*-test, $p = 0.20$).

stimulation or the indicative saccades required in both tasks, we may conclude that the neuronal machinery in the gaze-following patch in the pSTS might be responsible for the calculations needed to shift the observer's attention based on eye gaze. Unlike the shifts of attention evoked by our color-matching paradigm, gaze-following is reflexive (Friesen and Kingstone, 1998). However, this does not mean that it would not be subject to cognitive control. Indeed, careful psychophysical experiments on monkey head gaze-following (Marciniak et al., 2015), probably homologous to human gaze-following, clearly show that with the exception of a small early reflex component, a substantial part of the gaze-following response can be suppressed. Hence, we can be confident that the BOLD contrast used to identify the GFP reflects differences in gaze-following–related processing and its cognitive control. Our paradigm vetoed an immediate behavioral response to the gaze cue as subjects had to delay the response until the occurrence of the "GO" signal. Hence, one might be concerned that the GFP activity we observed in this experiment might differ from the normal pattern evoked by spontaneous gaze-following. However, the spatial coordinates of the GFP identified here are in accordance with our previous findings on activations evoked by spontaneous gaze-following (Materna et al., 2008a).

In the second experiment, we used a classical static face localizer to map the face-selective regions potentially involved in extracting information on face and eye gaze orientation to clarify the anatomic relationship between the GFP and the members of this face patch system. In fact, we did not observe any overlap between the GFP and any of the face patches, in particular not with a patch in the posterior STS (STS-FA), which in view of its localization as described by previous work (Kanwisher et al., 1997; Haxby et al., 2000) might have been expected to overlap with the GFP. One might argue that a lack of overlap between the two is not surprising, given that the GFP is orchestrating shifts of attention guided by the eyes, i.e., just one of many elements that make up faces and possibly not that influential in the STS-FA. However, the following consideration speaks against the validity of this criticism. As already shown by Wollaston in the 19th century (Wollaston, 1824), estimates of eye gaze depend on concurrent information on the orientation of the face. And this latter information is available in the GFP. This was shown by Laube et al. (2011) who could establish that the influence of head or face orientation on perceived eye direction, first described by Wollaston, finds its correlate in changes of the BOLD signal in the GFP. On the other hand, previous fMRI work on face perception has suggested that one of the hallmarks of the STS-FA is a stark interest in the changing aspects of faces which, like changes in eye and face orientation, are important for gaze-following (Hoffman and Haxby, 2000; Lee et al., 2010). Hence, the fact that the GFP and the STS-FA are distinct, although both handling information on oriented faces and most probably also oriented eyes, clearly indicates different functional roles. On the other hand, the anatomic vicinity may suggest an exchange of pertinent information between the two. However, if the GFP handles information on averted faces, why does it not light up in the passive viewing experiment? The answer is that its activation is most probably contingent on the presence of an object serving as goal for the gaze and observer's intention to follow gaze.

We found the maximum BOLD response to faces in the STS-FA rather than in the FFA or OFA as many other studies (Engell and Haxby, 2007). The reason is that, in our passive task to elicit maximal responses in the STS-FA, the set of face stimuli used was confined to pictures of emotionally neutral faces with averted eyes with the head straight, known to be less suitable for the FFA or OFA (Hoffman and Haxby, 2000; Narumoto et al., 2001). On the other hand, in most of the studies yielding stronger responses in the FFA or OFA, the emphasis was on faces exhibiting direct eye gaze, stimuli that seem to favor identity-processing.

In Pitcher et al. (2011), a face-selective area in the right pSTS was reported that responded three times more strongly to dynamic faces than to static faces. Hence, one may speculate that the current study using static stimuli underestimated the true size of the STS-FA and therefore failed to reveal an overlap between the GFP and the STS-FA. We cannot exclude the possibility that more powerful face stimuli might have expanded the activated areas with the consequence of some overlap to emerge. However, given the fact that the mean Talaraich coordinates of the pSTS patch center as given by Pitcher et al. (2011), (54, -38, 4), and the coordinates of the GFP in our study, (50, −64, 2), are separated by 26 mm Euclidean distance clearly supports the conclusion of largely noncongruent patches, at least when a static face localizer is used to map face-selective areas.

Nonhuman primates follow head gaze to establish joint attention. This behavior emerges very early during the development of the individual (Tomasello and Carpenter, 2005; Tomasello et al., 2007). According to Marciniak et al. (2015), it is characterized by key features that make human eye gaze-following reflexlike, namely swiftness and incomplete cognitive control. As described earlier, monkey head gaze-following activates a patch of cortex (the monkey GFP) whose location bilaterally in the posterior STS is reminiscent of the location of the human GFP. Also, the monkey GFP is anatomically distinct, not showing overlap with any of the face patches that can be activated by the passive vision of faces (Tsao et al., 2003, 2006). As a matter of fact, the spatial relationship of the monkey GFP with respect to the posterior face patch (PL) and the middle face patches (ML, MF) is reminiscent of the spatial relationship of the human GFP to the most posterior face-selective area (OFA) and the two more anterior ones (FFA, STS-FA). This lends further support to the notion of a close correspondence of the respective architectures. The major difference seems to be the ability of the human architecture to integrate social cues, providing directional information, other than head cues such as eye direction or the direction of fingers (Materna et al., 2008b; Laube et al., 2011). In other words, both species seem to exhibit a common core architecture for gaze-following, possibly reflecting homologous ancestry.

The notion of separate, yet possibly interdependent, cortical structures for the processing of faces and in particular faces showing gaze aversion and gaze-following is interesting with regard to observations on subjects with autism spectrum disorder (ASD). At least some people with ASD seem to be able to distinguish between different eye gaze positions when tested in discrimination tasks, suggesting an intact face-processing network. However, they fail to use information provided by the other's face to follow the gaze and establish joint attention (Baron-Cohen, 1995; Leekam et al., 1998, 2000). In accordance with these behavioral observations, Pelphrey et al. (2005a) reported a lack of differentiation in the STS BOLD responses of ASD subjects when confronted with averted target-directed and averted non–target-directed eye gaze stimuli, a deficit that may reflect an inability to integrate information on the other's gaze and the object of interest. The tentative conclusion suggested by these findings may be one of differential vulnerability of the face-processing network and the GFP, with the latter selectively compromised in ASD. However, what exactly is the added value of the GFP? At this stage, the lack of knowledge of the neuronal computations inside the GFP does not allow more than an admittedly rather vague speculation. We think that the GFP may be needed to convert directional information on eye and face/head orientation as well as directional information offered by other parts of the body into a "vector" describing the necessary shift of the observer's "spotlight of attention" to the place of interest. Moreover, to ultimately lead to the establishment of joint attention devoted to an object found in a particular place, the GFP may also help to integrate information on the object at stake. Finally, to be viable, these calculations require the integration of knowledge on the observer's viewpoint. A final remark pertains a possible role of the most anterior member of the face-processing network, the IFG-FA, located at the junction of inferior frontal sulcus and the precentral sulcus in gaze-following. There is evidence that the BOLD response of IFG-FA to faces is primarily driven by the eyes, i.e., the response to faces with eyes is lower than the presentation of the eyes alone and higher than to faces without eyes (Chan and Downing, 2011). In view of these findings and, moreover, the proximity of the IFG-FA to the frontal eye field, the authors speculated that it might contribute to analyze others' gaze to elicit gaze-following movements of the observer. Hence, future work will have to address the possibility that not the face patch immediately neighboring the GFP but a much more remote anterior face patch, the IFG-FA, may serve as the major source of directional information provided by the eyes and the face.

References

Allison T, Puce A, McCarthy G (2000) Social perception from visual cues: role of the STS region. Trends Cogn Sci 4:267–278. Medline

Baron-Cohen S (1994) How to build a baby that can read minds: cognitive mechanisms in mindreading. Curr Psychology Cogn 13: 513–552.

Baron-Cohen S (1995) Mindblindness: An Essay on Autism and Theory of Mind. MIT Press, Cambridge, MA.

Chan AW, Downing PE (2011) Faces and eyes in human lateral prefrontal cortex. Front Hum Neurosci 5:51. CrossRef Medline

Emery NJ (2000) The eyes have it: the neuroethology, function and evolution of social gaze. Neurosci Biobehav Rev 24:581–604. Medline

Engell AD, Haxby JV (2007) Facial expression and gaze-direction in human superior temporal sulcus. Neuropsychologia 45:3234–3241. CrossRef Medline

Fox CJ, Iaria G, Barton JJ (2009) Defining the face processing network: optimization of the functional localizer in fMRI. Hum Brain Mapp 30:1637–1651. CrossRef Medline

Fox E, Damjanovic L (2006) The eyes are sufficient to produce a threat superiority effect. Emotion (Washington, DC) 6:534–539. CrossRef Medline

Friesen CK, Kingstone A (1998) The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. Psychonom Bull Rev 5:490–495. CrossRef

Friston KJ, Holmes AP, Worsley KJ, Poline JB, Frith CD, Frackowiak RS (1995) Statistical parametric mapping in functional imaging: A general linear approach. Hum Brain Mapp 2:189–210.

Friston KJ, Holmes AP, Worsley KJ (1999) How many subjects constitute a study? NeuroImage 10:1–5. CrossRef Medline

Gobbini MI, Haxby JV (2006) Neural response to the visual familiarity of faces. Brain Res Bull 71:76–82. CrossRef Medline

Grill-Spector K, Knouf N, Kanwisher N (2004) The fusiform face area subserves face perception, not generic within-category identification. Nat Neurosci 7:555–562. CrossRef Medline

Haxby JV, Hoffman EA, Gobbini MI (2000) The distributed human neural system for face perception. Trends Cogn Sci 4:223–233. Medline

Heller R, Golland Y, Malach R, Benjamini Y (2007) Conjunction group analysis: an alternative to mixed/random effect analysis. NeuroImage 37:1178–1185. CrossRef Medline

Hoffman EA, Haxby JV (2000) Distinct representations of eye gaze and identity in the distributed human neural system for face perception. Nat Neurosci 3:80–84. CrossRef Medline

Hooker CI, Paller KA, Gitelman DR, Parrish TB, Mesulam MM, Reber PJ (2003) Brain networks for analyzing eye gaze. Brain Res Cogn Brain Res 17:406–418. Medline

Ishai A, Schmidt CF, Boesiger P (2005) Face perception is mediated by a distributed cortical network. Brain Res Bulletin 67:87–93. CrossRef Medline

Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci 17:4302–4311.

Langner O, Dotsch R, Bijlstra G, Wigboldus DHJ, Hawk ST, Van Knippenberg A (2010) Presentation and validation of the Radboud faces database. Cogn Emot 24:1377–1388. CrossRef

Langton SR, Bruce V (2000) You must see the point: automatic processing of cues to the direction of social attention. J Exp Psychol Hum Percept Perform 26:747–757. Medline

Laube I, Kamphuis S, Dicke PW, Thier P (2011) Cortical processing of head- and eye-gaze cues guiding joint social attention. NeuroImage 54:1643–1653. CrossRef Medline

Lee LC, Andrews TJ, Johnson SJ, Woods W, Gouws A, Green GG, Young AW (2010) Neural responses to rigidly moving faces displaying shifts in social attention investigated with fMRI and MEG. Neuropsychologia 48:477–490. CrossRef Medline

Leekam SR, Hunnisett E, Moore C (1998) Targets and cues: gaze-following in children with autism. J Child Psychol Psychiatry 39: 951–962. Medline

Leekam SR, López B, Moore C (2000) Attention and joint attention in preschool children with autism. Dev Psychol 36:261–273. Medline

Marciniak K, Dicke PW, Thier P (2015) Monkeys head-gaze following is fast, precise and not fully suppressible. Proc Biol Sci R Soc 282. CrossRef

Marciniak K, Atabaki A, Dicke PW, Thier P (2014) Disparate substrates for head gaze following and face perception in the monkey superior temporal sulcus. eLife 3. CrossRef

Materna S, Dicke PW, Thier P (2008a) Dissociable roles of the superior temporal sulcus and the intraparietal sulcus in joint attention: a functional magnetic resonance imaging study. J Cogn Neurosci 20:108–119.

Materna S, Dicke PW, Thier P (2008b) The posterior superior temporal sulcus is involved in social communication not specific for the eyes. Neuropsychologia 46:2759–2765. CrossRef Medline

Narumoto J, Okada T, Sadato N, Fukui K, Yonekura Y (2001) Attention to emotion modulates fMRI activity in human right superior temporal sulcus. Brain Res Cognitive Brain Res 12:225–231. Medline

Pelphrey KA, Viola RJ, McCarthy G (2004) When strangers pass: processing of mutual and averted social gaze in the superior temporal sulcus. Psychol Sci 15:598–603. CrossRef Medline

Pelphrey KA, Morris JP, McCarthy G (2005a) Neural basis of eye gaze processing deficits in autism. Brain J Neurol 128:1038–1048. CrossRef Medline

Pelphrey KA, Singerman JD, Allison T, McCarthy G (2003) Brain activation evoked by perception of gaze shifts: the influence of context. Neuropsychologia 41:156–170. Medline

Pelphrey KA, Morris JP, Michelich CR, Allison T, McCarthy G (2005b) Functional anatomy of biological motion perception in posterior temporal cortex: an FMRI study of eye, mouth and hand movements. Cereb Cortex 15:1866–1876.

Pitcher D, Dilks DD, Saxe RR, Triantafyllou C, Kanwisher N (2011) Differential selectivity for dynamic versus static information in face-selective cortical regions. NeuroImage 56:2356–2363. CrossRef Medline

Puce A, Allison T, Bentin S, Gore JC, McCarthy G (1998) Temporal cortex activation in humans viewing eye and mouth movements. J Neurosci 18:2188–2199.

Shimojo S, Simion C, Shimojo E, Scheier C (2003) Gaze bias both reflects and influences preference. Nat Neurosci 6:1317–1322. CrossRef Medline

Tomasello M, Carpenter M (2005) The emergence of social cognition in three young chimpanzees. Monogr Soc Res Child Dev 70:vii-132. CrossRef Medline

Tomasello M, Hare B, Lehmann H, Call J (2007) Reliance on head versus eyes in the gaze following of great apes and human infants: the cooperative eye hypothesis. J Hum Evol 52:314–320. CrossRef Medline

Tsao DY, Moeller S, Freiwald WA (2008) Comparing face patch systems in macaques and humans. Proc Natl Acad Sci U S A 105:19514–19519. CrossRef Medline

Tsao DY, Freiwald WA, Tootell RB, Livingstone MS (2006) A cortical region consisting entirely of face-selective cells. Science (New York) 311:670–674. CrossRef Medline

Tsao DY, Freiwald WA, Knutsen TA, Mandeville JB, Tootell RB (2003) Faces and objects in macaque cerebral cortex. Nat Neurosci 6:989–995. CrossRef Medline

Van Essen DC (2005) A population-average, landmark- and surface-based (PALS) atlas of human cerebral cortex. NeuroImage 28:635–662. CrossRef Medline

Wicker B, Michel F, Henaff MA, Decety J (1998) Brain regions involved in the perception of gaze: a PET study. NeuroImage 8:221–227. CrossRef Medline

Wollaston WH (1824) On the apparent direction of eye in a portrait. Philos Trans R Soc Lond Ser B 114:247–256. CrossRef

Chapter 2

A fronto-temporo-parietal network disambiguates potential objects of joint attention

Peter Kraemer*, Marius Görner*, Hamidreza Ramezanpour*, Peter Dicke, Peter Thier (*equal contribution)

Under Review at Cerebral Cortex and available on BioRxiv, doi: https://doi.org/10.1101/542555, *2019*

In the previous studies on the neural basis of gaze following, the target object could be identified unambiguously by gaze direction as for a given gaze direction the vector hit one object only. Hence, it remained unclear if the GFP helps to integrate the information needed to disambiguate the object choice in case the gaze vector hits more than one object. We hypothesized that singling out the relevant object was a consequence of recourse to prior information on the objects and their potential value for the other or for the observer. In order to test this hypothesis we carried out an fMRI study in which the selection of the object of joint attention required that the observer would have to recourse to complementary information aside from the gaze cue. In support of this hypothesis we could show that the disambiguation is based on a 3-component network. A first component, the well-known ‘gaze following patch’ in the posterior STS is activated by gaze following per se. BOLD activity here is determined exclusively by the usage of gaze direction and is independent of the need to disambiguate the relevant object. On the other hand, BOLD activity revealing *a-priori* information relevant for the disambiguation and starting early enough to this end is confined to a patch of cortex at the inferior frontal junction. Finally, BOLD activity reflecting the convergence of both, *a-priori* information and gaze direction, needed to shift attention to a particular object location is confined to the posterior parietal cortex.

A fronto-temporo-parietal network disambiguates potential objects of joint attention

P. M. Kraemer[1, 5†], M. Görner[1, 2, 3, †], H. Ramezanpour[1,2 ,3 ,†] , P. W. Dicke[1] and P. Thier[1,4,*]

[1] Department of Cognitive Neurology, Hertie Institute for Clinical Brain Research, University of Tübingen, 72076 Tübingen, Germany.

[2] Graduate School of Neural and Behavioural Sciences, University of Tübingen, 72074 Tübingen, Germany.

[3] International Max Planck Research School for Cognitive and Systems Neuroscience, University of Tübingen, 72074 Tübingen, Germany.

[4] Werner Reichardt Centre for Integrative Neuroscience, University of Tübingen, 72076 Tübingen, Germany.

[5] Center for Decision Neuroscience, Faculty of Psychology, University of Basel, 4055 Basel, Switzerland.

[†] These authors contributed equally to this work.

* Corresponding Author/Lead Contact: Peter Thier, Department of Cognitive Neurology, Hertie Institute for Clinical Brain Research, Hoppe-Seyler-Str. 3, 72076 Tübingen, Germany. E-mail: thier@uni-tuebingen.de.

Abstract

We use the other´s gaze direction to identify her/his object of interest and to shift our attention to the same object, i.e. to establish joint attention. However, gaze direction may not be sufficient to unambiguously identify the relevant object as the other´s gaze may hit more than one object. In this case, the observer must use other sources of information to choose the object. Using fMRI, we suggest that such a decision is based on a 3-component network. A first component, the ‘gaze following patch’ in the vicinity of the posterior STS, is activated exclusively by the extraction of the other´s gaze direction and is independent of the need to decide between possible objects in the line of sight. A second component, a patch of cortex at the inferior frontal junction is sensitive to the presence or absence of complementary information needed for the disambiguation of the objects on the gaze vector. Finally, BOLD activity in the posterior parietal cortex exhibits the convergence of both streams of information in accordance with the role of this part of cortex in coordinating attentional shifts to particular object locations.

Keywords: gaze following, superior temporal sulcus, inferior frontal junction, human lateral intraparietal area, fMRI

Introduction

We follow the other´s gaze to objects of her/his attention allowing us to shift our attention to the same object, thereby establishing joint attention. By associating our object-related intentions, expectations and desires with the other one, joint attention allows us to develop a *Theory of* (the other´s) *Mind* (TOM). TOM is a major basis of successful social interactions (Baron-Cohen 1994, 1995) and, arguably, its absence is at the core of devastating neuropsychiatric diseases such as autism. Human gaze following is geometric (Butterworth and Jarrett 1991; Atabaki et al. 2015). This means that we use the other´s gaze vector to identify the exact location of the object of interest. The features of the human eye such as the high contrast between the white sclera and dark iris allow us to determine the other´s eye direction at high resolution (Kobayashi and Kohshima 1997; Bock et al. 2008). However, knowledge of direction is not sufficient to pinpoint an object in 3D. In principle, differences between the directions of the two eyes, i.e. knowledge of the vergence angle, could be exploited to this end. Yet, this will work only for objects close to the beholder as the angle will become imperceptibly small if the objects are outside the confines of peripersonal space. On the other hand, gaze following remains precise also for objects quite far from the other one although the gaze vector will in many cases hit more than one object (Butterworth and Jarrett 1991). Hence, how can these objects be disambiguated? We hypothesized that singling out the relevant object is a consequence of recourse to prior information on the objects and their potential value for the other. For instance, let us assume that the day is hot and that the other´s appearance may suggest thirst and the desire to take a sip of something cool. If her/his gaze hit a cool beverage within a set of other objects of little relevance for a thirsty person, the observer might safely infer that the beverage is the object of desire. In this example, gaze following is dependent on prior assumptions about the value of objects for the other. Of course, also the value the object may have for the observer matters. For instance, Liuzza et al. showed that an observer´s appetence to follow the other´s gaze to portraits of political leaders is modulated by the degree of political closeness (Liuzza et al. 2011). If the politician attended by the other was a political opponent of the observer, the willingness to follow gaze was significantly reduced. Also knowing that gaze following may be inadequate in a given situation and that the other may become aware of an inadequate behavior will suppress it (Teufel et al. 2009, 2010). However, only assumptions about the object value for the other will help to disambiguate the scene.

Following the gaze of others to a particular object is accompanied by a selective BOLD signal in an island of cortex in the posterior superior temporal sulcus (pSTS), the “gaze-following patch (GFP)” (Materna et al. 2008; Laube et al. 2011; Marquardt et al. 2017). In these studies, the target object could be identified unambiguously by gaze direction as for a given gaze direction the vector hit one object only. Hence, it remains unclear if the GFP helps to integrate the information needed to disambiguate the object choice in case the gaze vector hits more than one object. In order to address this question, we carried out an fMRI study in which the selection of the object of joint attention required that the observer recoursed on another source of information aside from the gaze cue.

Materials and Methods

Participants

Nineteen healthy, right-handed volunteers (9 females and 10 males, mean age 27.4, *SD* = 3.6) participated in the study over three sessions. Participants gave written consent to the procedures of the experiment. The study was approved by the Ethics Review Board of the Tübingen Medical School and was carried out in accordance with the principles of human research ethics of the Declaration of Helsinki.

Task and procedure

The study was conducted in three sessions across separate days. On day 1, we instructed participants about the study goals and familiarized them with the experimental paradigms outside the MRI-scanner by carrying out all relevant parts of the fMRI experiments. The following fMRI-experiments included a functional localizer paradigm for the scanning session on day 2 as well as a contextual gaze following paradigm for the scanning session on day 3.

Behavioral session. After participants had been familiarized with the tasks, they were head-fixed using a chinrest and a strap to fix the forehead to the rest. Subjects were facing towards a frontoparallel screen (resolution = 1280×1024 pixels, 60 Hz) (distance to eyes ≈ 600 mm). Eye tracking data were recorded while participants had to complete 80 trials of the localizer paradigm and 72 trials of contextual gaze following.

Localizer task**.** We resorted to the same paradigm used in the study by Marquardt and colleagues, (Marquardt et al. 2017), to localize the gaze following network and in particular its core, the GFP. In this paradigm, subjects were asked to make saccades to distinct spatial targets based on information provided by a human portrait presented to the observer. Depending on the instruction, subjects either had to rely on the seen gaze direction to identify the correct target (*gaze following* condition) or, alternatively, they had to use the color of the irises, changing from trial to trial but always mapping to one of the targets, in order to make a saccade to the target having the same color (*color mapping* condition). In other words, the only difference between the two tasks was the information subjects had to exploit in order to solve the task, while the visual stimuli were the same.

This task is associated with higher BOLD activity in the GFP, a region, close to the posterior end of the superior temporal sulcus (pSTS), when subjects perform gaze following compared to color mapping. The task is further associated with the activation of regions in the posterior parietal cortex as well as the frontal cortex that take part in controlling spatial attention and saccade generation (Materna et al. 2008; Marquardt et al. 2017). Out of the 19 subjects of our study, 16 performed 6 runs (40 trials per run) and for reasons of time management during image acquisition, one subject performed five runs and two subjects performed four runs.

Contextual gaze following task**.** An example of a trial is shown in Figure 1. Each trial consisted of the following sequence of events. The trial started with the appearance of the portrait of an avatar image (6.7×10.5 degrees of visual angle) in the center of the screen together with four arrays of drawn objects (houses and hands, three objects per array). Subjects were asked to fixate on a red fixation dot (diameter) between the avatar´s eyes. After five seconds of baseline fixation, the portrait's gaze shifted towards one specific target object. Simultaneously, a spoken instruction either specified the object class of the target (spoken words "hand" or "house") or was not informative ("none"). While maintaining fixation, subjects needed to judge which object the target was (i.e. on which object the face was most likely looking at). After five seconds delay, the fixation dot vanished, an event that served as a *go*-signal. Participants had two seconds to make a saccade to the chosen target object and fixate it until a subsequent blank fixation screen was presented for eight seconds. The subjects were instructed to perform the task as accurately as possible. They were specifically instructed, when unsure about the actual target, to still rely on gaze and contextual information and choose the target they believed the avatar to be looking at.

The information provided by the spoken instruction distinguished three experimental conditions, an *unambiguous*, and two ambiguous conditions: *ambiguous-informative* and *ambiguous-uninformative*. The verbal instruction in the *unambiguous* condition reduced the number of potential targets from three to one by naming the object category with only one representative in the array. For instance, in Figure 1 the avatar gazes at the lower left array, specifying two hands and one house as potential gaze targets. An unambiguous instruction would be the auditory cue “house”. The *ambiguous-informative* instruction in this example, “hand”, reduced the number of potential gaze targets to two. In the *ambiguous-uninformative* condition the instruction would have been “none”, not suited to reduce the number of potential targets.

Participants performed six blocks of 30 trials each (10 per condition), summing up to 180 trials in total.

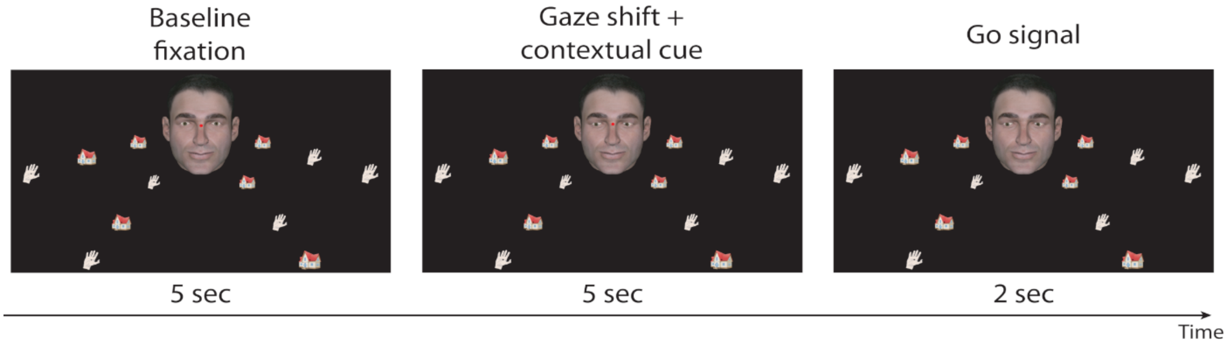

Figure 1. Contextual gaze following task. An avatar appeared in the center of the screen together with four linearly arranged sets of objects (houses and hands). After a baseline fixation period, the portrait’s gaze shifted towards one specific target object simultaneously with an auditory contextual instruction specifying the object class of the target (hand or house) or not, i.e. remaining uninformative (“none”). While maintaining fixation, subjects needed to decide on the target and make a saccade to the chosen target after a *go*-signal indicated by the disappearance of the fixation dot.

Stimuli

Control of visual and auditory stimuli as well as data collection were controlled by the Linux-based open source system *nrec (https://nrec.neurologie.uni-Tübingen.de/)*. The stimuli in the localizer task were identical to the stimuli used in a previous study (Marquardt et al. 2017). The stimuli of the contextual gaze following task consisted of an avatar and in total 12 target objects from two categories (houses and hands). The avatar was generated with the custom-made OpenGL library *Virtual Gaze Studio* (Benz 2008; Hübner 2008) which offers a controlled virtual 3D-environment in which an avatar can be set to precisely gaze at specific objects. More specifically, the program allows to place objects on a circle, parallel to the coronal axis, anterior to the avatar face. For each stimulus, we placed 12 objects in the surroundings of the avatar. The location of individual objects was fully determined by the distance to the coronal plane at the level of the avatar´s nasion, the radius of the circle and the angle of the object on that circle. By keeping the angle on the circle constant for sets of three objects, we created four arrays at angles 120°, 150°, 210° and 240°. The individual locations of these objects were specified by varying the distance and the circle radii based on trigonometric calculations. For these calculations, we assumed a right triangle from the avatar´s nasion with the hypotenuse pointing towards the object, an adjacent leg (length corresponded to the distance of the circle) proceeding orthogonal to the coronal plane, and an opposite leg which corresponded to the radius. By keeping tanα fixed to 0.268, we varied the distances and circle radii. For the 120° and 240° arrays, the circle radii were 335, 480, 580 and the distances were 90, 129 and 151 virtual mm. For the 150° and 210° arrays, the radii were 380, 510 and 590 and the distances were 102, 137 and 158 virtual mm. The reason for the difference in radii and distances between 120°/240° and 150°/210° arrays was that this allowed to exploit the total width of the screen. This procedure guaranteed that the angle of the gaze vector to all objects on an array was almost identical. This makes it relevant to take contextual information into account in order to choose the true target.

The objects were drawings of the two categories houses and hands, downloaded from freely available online sources (http://www.allvectors.com/house-vector/, https://www.freepik.com/free-vector/hand-drawn-hands_812824.htm#term=hands&page=1&%20position=37). The target objects were arranged in four radial directions (three objects in each direction) with the avatar eyes as the origin; in other words, the avatar’s gaze always hit one out of three objects along the gaze vector though participants were not able to tell which of the three it was. On each array, either two hands and one house or one hand and two houses were present. Further, we fixed the number of hands and houses per hemifield to three. The relative order of the objects was pseudo-randomized from trial to trial.

We created a pool of stimulus sets which satisfied three constraints: There was an equal number of trials in which a) the targets were hands or houses, b) targets were presented with an *unambiguous*, *ambiguous-informative* and *ambiguous-uninformative* instruction, and c) the spatial position (one out of twelve potential positions) of targets was matched.

This led to 72 stimulus sets. We exposed every subject to 180 trials in which each stimulus set was shown twice and for the residual 36 trials, stimuli were drawn from pseudo-randomly from the stimulus pool so that the three aforementioned criteria were met.

Auditory instructions were delivered via headphones (Sennheiser HD 201, Wedemark-Wennebostel, Germany, during the behavioral session and the standard air pressure headphones of the scanner system during the MRI sessions). The auditory instructions “hand”, “house” and “none” were computer generated with the web application imTranslator (http://imtranslator.net/translate-and-speak/speak/english/) and processed with the software Audacity 2.1.2. The sound files had a duration of 600 ms.

Eye tracking

During all three sessions, we recorded eye movements of the right eyes using commercial eye tracking systems (Behavioral sessions: Chronos Vision C-ETD, Berlin, Germany, sampling rate 400 Hz, resolution < 1° visual angle; Scanning sessions: SMI iView X MRI-LR, Berlin, Germany, sampling rate = 50 Hz, resolution ≈ 1° visual angle).

Eye tracking data was processed as follows. First, we normalized the raw eye tracking signal by dividing it by the average of the time series. Eye blinks were removed using a velocity threshold (> 1000 °/s visual angle). Next, we focused on a time window in which we expected the saccades to the target objects to occur ([*go*-signal – 500 ms, *go*-signal + 1800 ms]). Within this time window, we detected saccades by identifying the time point of maximal eye movement velocity. Pre- and post-saccadic fixation positions were determined by averaging periods of 200 ms before and after the saccade occurred. Due to partly extensive noise of the eye tracking signal recorded in the scanner, we did not automatize the categorization of the final gaze position. Instead, we plotted X- and Y-coordinates of the post-saccadic eye position for every run that was not contaminated by too much noise. An investigator (MG), who was blind to the true gaze target-directions of the stimulus face, manually validated which trials yielded positions that were clearly assignable to a distinct object location. For the behavioral analysis we only used valid trials (mean number of valid trials per participant = 80.2, *SD* = 45.4, range = [0, 153]) and weighted the individual performance values by its number in order to compute weighted means and *SD*s. Note, that we used these valid trials only for the behavioral analysis but used all trials of the participants for the fMRI analysis, assuming that eye tracking measurement noise was independent of the performance of the subjects.

fMRI acquisition and preprocessing

We acquired MR images using a 3T scanner (Siemens Magnetom Prisma, Erlangen, Germany) with a 20-channel phased array head coil at the Department of Biomedical Magnetic Resonance of the University of Tübingen. The head of the subjects was fixed inside the head coil by using plastic foam cushions to avoid head movements. An AutoAlign sequence was used to standardize the alignment of images across sessions and subjects. A high-resolution T1-weighted anatomical scan (MP-RAGE, 176×256×256 voxel, voxel size 1×1×1 mm) and local field maps were acquired. Functional scans were carried out using a $T2^*$-weighted echo-planar multi-banded 2D sequence (multi-band factor = 2, TE = 35 ms, TR = 1500 ms, flip angle = 70°) which covered the whole brain (44×64×64 voxel, voxel size 3×3×3 mm, interleaved slice acquisition, no gap).

For image preprocessing we used the MATLAB SPM12 toolbox (Statistical Parametric Mapping, https://www.fil.ion.ucl.ac.uk/spm/). The anatomical images were segmented and realigned to the SPM T1 template in MNI space. The functional images were realigned to the first image of each respective run, slice-time corrected and coregistered to the anatomical image. Structural and functional images were spatially normalized to MNI space. Finally, functional images were spatially smoothed with a Gaussian kernel (6 mm full-width at half maximum).

fMRI analysis

We estimated a generalized linear model (GLM) to identify regions of interest (ROIs) of single subjects. On these regions, we performed time course analyses to investigate event-related BOLD signal changes. In a first-level analysis, we estimated GLMs for the localizer task (GLM_{loc}) and the contextual gaze following task (GLM_{cgf}). The GLM_{loc} included predictors for the onset of directional cues and of the baseline fixation phase.

The GLM_{cgf} had predictors for the onset of the contextual instruction coinciding with the gaze cue. These event specific predictors of the two GLMs used the canonical hemodynamic response function of SPM to model the data. We corrected for head motion artifacts by the estimation of six movement parameters with the data of the realignment preprocessing step. Low-frequency drifts were filtered using a high-pass filter (cutoff at 1/128 Hz).

GFP and hLIP localizer. Before collecting the data, we specified the expected locations of two brain areas from the literature. We resorted to the parietal coordinates of the human homologue of monkey area LIP (hLIP) which had been identified using a delayed saccade task (Sereno et al. 2001). The GFP standard coordinates were taken from Marquardt, Ramezanpour and coworkers (2017). We transformed the standard coordinates for the hLIP and the GFP from Talairach space into MNI space, using an online transformation method of Lacadie and colleagues (Lacadie et al. 2008, http://sprout022.sprout.yale.edu/mni2tal/mni2tal.html).

To identify ROIs at the group level, we compared beta weights of the statistical parametric maps from the GLM_{loc} in a second-level analysis. The GFP weights were derived from the contrast *gaze following* vs. *color mapping*, the hLIP weights from the contrast *directional cue* vs. *baseline fixation*. To be characterized as GFP or hLIP, a cluster´s maximum weights had to be located in close proximity to their respective standard coordinates.

We aimed to identify ROIs on an individual subject level. To this end, we used the contrast maps from the first-level analysis of the GLM_{loc}. We selected the coordinates of the maximum contrast voxel which minimized the distance to the group level coordinates. This voxel had to be part of a statistically significant cluster (cluster size ≥ 6, $p < 0.05$). Due to relatively low signal-to-noise ratio in the *gaze following vs. color mapping* contrast and the corresponding increased risk of false-positive activations, we decided to introduce a second criterion to make GFP localization more rigorous in single subjects. This *proximity criterion* required that the cluster additionally had to be located at least partially within 10 mm range of the group level coordinates of the respective ROI.

Contrasts of context conditions. In addition to our *a-priori* ROIs, we were interested, whether the contextual gaze following task might activate regions which we did not consider beforehand. We performed a whole-brain analysis of the data from the contextual gaze following task. Using the GLM_{cgf}, we contrasted the weights of the two *ambiguous* conditions with the *unambiguous* condition at the group level (second-level analysis, significance threshold p < 0.001, cluster size ≥ 6 voxel) as well as at the single subject level (first-level analysis, significance threshold p < .05, cluster size ≥ 6 voxel).

Time course analysis. We determined the individual time courses of the BOLD signal within sphere-shaped ROIs. Whenever we identified a ROI on the single-subject level, spheres with a radius of 5 mm were centered at the individual ROI coordinates. In case the identification of a ROI on the single-subject level was not possible, we deployed spheres with a radius of 10 mm centered at the group level location, assuming these spheres would capture relevant single-subject activity.

For every subject and sphere, raw time series of the BOLD signal were extracted using the MATLAB toolbox *marsbar 0.44* (http://marsbar.sourceforge.net). Due to technical problems in the reconstruction of trial times, for five participants we included only five runs and for two only four runs into the analysis. The time course of every trial was normalized by the average signal intensity 5 seconds before the onset of the contextual instruction and transformed into % of signal change. For each participant, we averaged time courses across trials and runs and used the time courses of the three contextual conditions in the six ROIs for our analysis. To test differences across conditions for statistical significance, we performed permutation tests at each time point after contextual instruction delivery. To do so, we pooled the data of two experimental conditions, respectively, and produced 10,000 random splits for each pool. By computing the differences between the means of these splits, we obtained a distribution of differences under the null hypothesis. Calculating the fraction of values more extreme than the actual difference between means allowed us to obtain a p-value for each time bin. To account for the multiple comparison problem, we transformed p-values into FDR corrected q-values (Benjamini and Hochberg 1995) and considered each time bin with q < .05 as statistically significant.

We carried out an additional analysis in order to obtain credible intervals (CI) for the time courses. To do so we computed hierarchical models for each experimental condition and ROI allowing the intercept to vary for each participant. The model was a linear combination of seven sinusoidal basis functions. Model estimation was conducted using the nideconv package (Hollander and Knapen 2017) which interfaces with the Stan language for Bayesian model estimation (Stan Development Team 2018). Analogous to frequentist confidence intervals, non-overlapping 95%-CIs imply a statistically significant difference. Unlike the interpretation of confidence intervals, CIs can be interpreted such that the estimate lies within the given range with a probability of 0.95.

Results

Our subjects participated in two fMRI experiments. The first one was a *localize*r task that allowed us to identify two regions of interest of which we know that they are relevant for attentional shifts based on social cues, the GFP and hLIP (Materna et al. 2008; Marquardt et al. 2017). Our main intention was to investigate the BOLD activity of these regions in a *contextual gaze following* task (experiment 2). In this task, the subjects used the gaze direction of a human avatar, complemented by a spoken instruction. In one out of three conditions the observer was able to unambiguously identify the relevant object out of several hit by the other´s gaze vector. This was the case in the *unambiguous condition (ua)* in which the spoken instruction identified an object class represented by only one exemplar on the avatar´s gaze vector. In the two other conditions (to which we refer collectively as ambiguous conditions) the spoken information was insufficient. Either because two exemplars of the relevant object category were available (*ambiguous-informative condition* (*inf*)) or because the verbal instruction was uninformative (*ambiguous-uninformative condition* (*uninf*)). In the latter case, observers were left with the choice between three objects.

Behavioral performance. In the localizer task, subjects were able to hit targets reliably and without significant difference between the two conditions (median hit rates: *gf*: 0.94 ± 0.13 *SD*; *cm*: 0.92 ± 0.09 *SD*; *p* = 0.6, two-tailed t-test, *N* = 19, Fig. 2). Using the gaze-following performance in the localizer task as reference we assumed the following expected hit rates for the contextual gaze following task: 0.94 for the *unambiguous condition*, 0.94*1/2 for the *ambiguous-informative* and 0.94*1/3 for the *ambiguous-uninformative* condition (Fig. 2). As summarized in Figure 2, the measured performances matched the assumptions in the contextual gaze following task very well (comparison by two-tailed t-tests, n.s.). This result indicates that the probability to identify an object as a target was exclusively determined by the information provided by gaze direction and the verbal instruction and not influenced by biases or uncontrolled strategies.

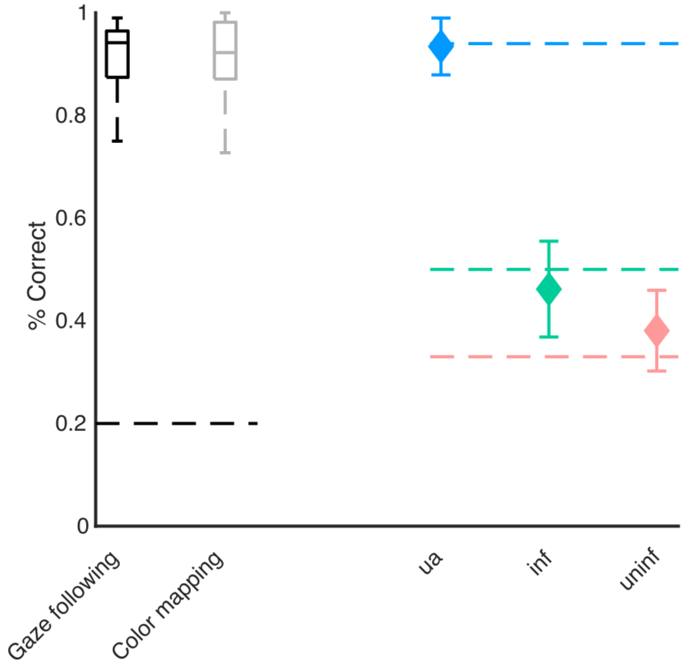

Figure 2. Behavioral performance. Left: Boxplots (black and gray) showing the percentage of correct response in the localizer paradigm (dashed line depicts chance level performance). Right: Plots of correct responses in the contextual gaze following paradigm (weighted mean performance and weighted *SD*, dashed lines depict expected performance; blue: unambiguous, green: ambiguous-informative, red: ambiguous-uninformative).

ROI localization. To localize the GFP we contrasted *gaze following* with *color mapping* trials in the first experiment. At the group level (*N* = 19) this contrast yielded a patch of significantly larger activity for *gaze following* close to the pSTS in both hemispheres. The contrast maxima (blue spheres in Figure 3, left column) were located at *x*, *y*, *z* = -57, -61, -1 in the left and at *x*, *y*, *z* = 48, -67, -1 in the right hemisphere. These locations closely match those known from previous studies, visualized as green and cyan spheres for comparison (Materna et al. 2008; Marquardt et al. 2017). In addition to the GFP, the *gf* > *cm* contrast was also significant in a few more regions, not consistently seen as activated in previous work using the same paradigm (see supplementary material Tab. S1 for a list of all activated regions).

We localized the right hemispheric GFP in nine individual subjects (mean distance to group coordinates = 6.6 mm, *SD* = 3.1 mm) and the left GFP in six subjects (mean distance = 7.7 mm; *SD* = 1.4 mm) (white spheres in Figure 3, left column).

An analogous procedure was applied to localize the hLIP, using the contrast *directional cue* vs. *baseline fixation*. The location of maximum activation at the group level was found to be at *x*, *y*, *z* = 21, -67, 50 (right) and *x*, *y*, *z* = -21, -67, 53 (left) (blue spheres in Figure 3, middle column) in good accordance with previous work on saccade-related activity in the parietal cortex (Sereno et al. 2001; Figure 3, middle). We identified the hLIP regions bilaterally in all 19 subjects individually with a mean distance of 13.4 mm (*SD* = 3.9 mm) to the standard coordinates in the right hemisphere and 11.93 mm (*SD* = 3.7 mm) in the left hemisphere (white spheres in Figure 3, middle column).

In order to determine if BOLD activity in regions not delineated by the localizer experiment was modulated by the conditions of the contextual gaze following task, we contrasted activity in each of the ambiguous conditions with the *unambiguous* condition. This contrast was significant for a region in the inferior prefrontal cortex (Figure 3, bottom) whose group level maxima were found in slightly different locations in the two hemispheres, namely at *x*, *y*, *z* = -39, 11, 29 in the left and *x*, *y*, *z* = 48, 20, 23 in the right hemisphere (blue spheres), corresponding to the most lateral part of left BA 8 and the upper right BA 44. In 15 subjects we could delineate individual contrast locations (white spheres ibid., *SD* of individual locations (in mm): right x, y, z = 5, 6, 6; left x, y, z = 5, 8, 6). The individual locations scattered around BA 44, BA 8 and BA 9 and henceforth we will refer to this region as the inferior frontal junction (IFJ).

Weaker, albeit still significant *inf/uninf* > *ua* contrasts were also found in the medial part of left BA 8 at *x*, *y*, *z* = -3, 11, 50, bilaterally in BA 6 at *x*, *y*, *z* = -21, -4, 50 and *x*, *y*, *z* = 24, -1, 50 and at *x*, *y*, *z* = 36, 8, 47 (right hemisphere) not far from the IFJ (cf. Supplementary material Tab. S1). Reversing the contrast, i.e. *ua* > *inf/uninf*, we observed bihemispheric significance within BA 13 (insula), BA 40, within the cingulate cortex (BA 24 and 31) and within BA 7 (all *p* = 0.001, and a minimum of 6 adjacent voxel, cf. Supplementary material Tab. S1).

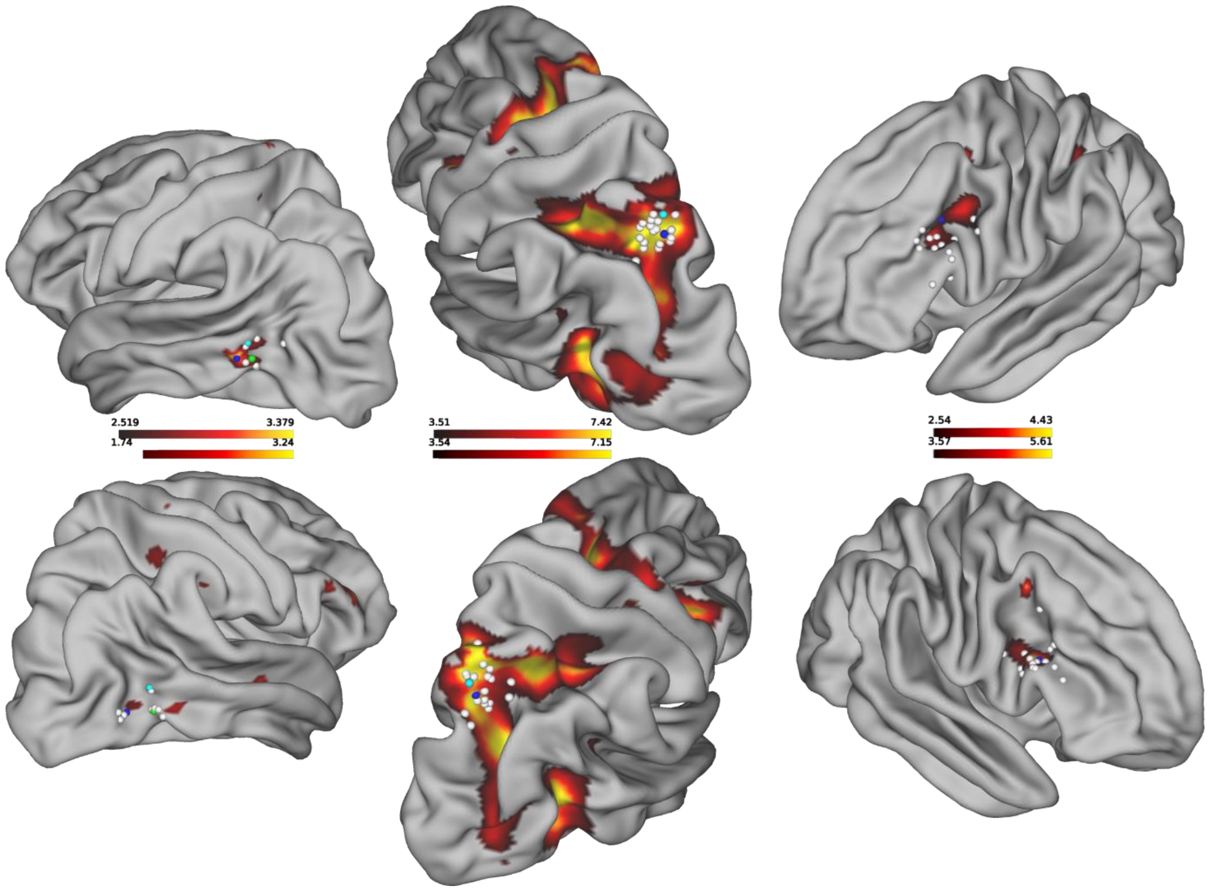

Figure 3. Activation maps. Left column: contrast gf > cm (localizer paradigm) used to identify the GFP. Blue dots mark maximum activation on the group level closest to locations taken from literature (green dots: Marquardt et al. 2017; cyan dots: Materna et al. 2008), white dots mark the maximum activation of those locations which were identifiable on the individual level. Middle column: contrast cm > bl (localizer paradigm) used to identify saccade-related activity in the hLIP closest to location taken from Sereno et al. (2001) (cyan dot). Blue and white dots mark again group level and individual coordinates; Right column: uninf > ua (contextual gaze following paradigm). Blue and white dots mark the group level and individual locations of the maximum IFJ-activity.

Time course of BOLD signals. We wanted to know how the BOLD signal depended on time relative to trial onset in both of the predefined ROIs and the IFJ. In order to characterize the time course of the signal in the ROIs, we quantified the BOLD activity in spheres with a radius of 5 mm in case individual ROIs had been determined. Otherwise spheres with a radius of 10 mm centered at the group level coordinates were used to extract the averaged activity of the included voxels. Figure 4 shows the averaged time courses of the BOLD signal for each condition of the *contextual gaze followin*g experiment separately for the GFP and hLIP.

We performed two types of analysis to investigate effects of context condition (*ua*, *inf*, *uninf*) on the BOLD activity; (1) permutation tests on each time point of the extracted BOLD signals (FDR corrected) and (2) estimation of hierarchical models to infer CIs of the time courses (cf. Fig. S1) (cf. Materials and Methods section for details). Since both methods yielded qualitatively identical results (with one exception described below) we will focus on the model-free analysis, here.

In the GFP, we observed two peaks throughout the trial, one at 10 sec and the other one after 16.5 sec. Considering the latency of the BOLD signal of about 5 sec we assume that the first peak is related to the onset of the cue (at 5 sec) and the second to the *go*-signal at 10 sec. For the GFP we did not observe significant difference between any conditions at any time point.

The hLIP region depicted a similar two-peak pattern in response to the cue and the *go*-signal. Permutation tests indicated that the BOLD response in both hemispheres was significantly different between *ua* and *uninf* trials after 15 sec, in temporal correspondence to the *go*-signal ($q < 0.5$, gray-shaded areas in Figure 4, bottom row). Qualitatively, the differentiation between the corresponding BOLD signals started earlier, after around 12 sec. There was no significant difference between the *inf* and *uninf* conditions ($q > 0.5$). Here, the second analysis method did not allow us to infer a statistical difference between conditions for the right hemisphere but only for the left one. Note, however, that the overlap of the 95%-CIs for the right hemisphere does not imply acceptance of the Null-Hypothesis of no difference. Indeed, since the pattern closely resembles the one for the left hemisphere and CIs are only barely overlapping, we tend to attribute this outcome to the low signal-to-noise ratio.

To rule out that the difference between the *ua* and the *uninf* condition was a reflection of a larger numbers of saccades caused by the higher uncertainty, we performed a t-test on the number of saccades across subjects, yielding no significant difference ($p > 0.5$).

To summarize, hLIP exhibited a significant stronger activity in *uninf* trials compared to *ua* trials (at least in the left hemisphere,) while this was not the case for the GFP.

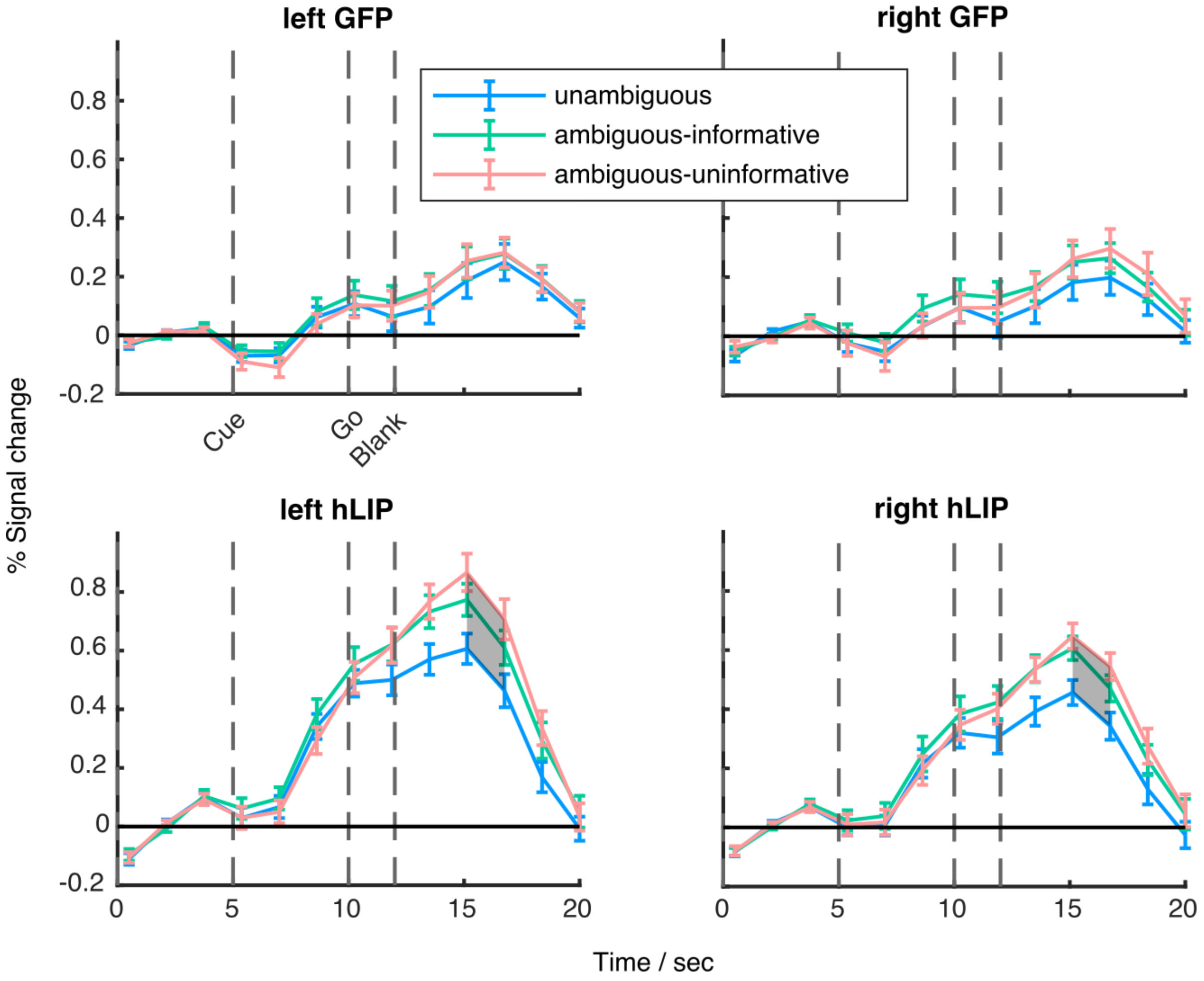

Figure 4. Time courses of activation. Time course of mean percent signal change in the contextual gaze following experiment in areas identified in the localizer experiment (error bars are SEM). Areas in which conditions showed significant differences are shaded (permutations test, $q < 0.05$).

We also performed time-course analyses on the regions identified by contrasting the context conditions of the *contextual gaze following* experiment to further assess their response characteristics in relation to events during trials. Of the identified regions, only the IFJ (Figure 5) survived this arguably more conservative analysis, even though the contrast itself suggests their sensitivity to the experimental conditions.

Compared to the *a-priori* ROIs, the BOLD signal in the IFJ exhibited a qualitatively different activity pattern during trials: In ambiguous conditions the signal appeared to rise in response to the onset of the gaze cue and the verbal instruction but there was no pronounced second peak in relation to the *go*-signal. The signal evoked in *unambiguous* trials was weak at best and dropped back to baseline after an initial bump which probably corresponds to the presentation of the cue. This was not the case in ambiguous trials where the signal sustained at a higher level until the end of the trial. Permutation tests yielded significant differences between the *unambiguous* and the *ambiguous-uninformative* conditions between 12 sec and 17 sec (left) and 12 sec and 15 sec (right) (q < 0.05). The second analysis suggested earlier discrimination between conditions starting around 9–10 sec which is clearly too early to be related to the *go*-signal and allows its alignment to the onset of the gaze and verbal cue. It appeared that the IFJ differentiated earlier between ambiguity condition than the hLIP. However, given the low temporal resolution of BOLD responses, one should treat this observation with caution. The profiles for *ambiguous-informative* and the *ambiguous-uninformative* were statistically not different from each other.

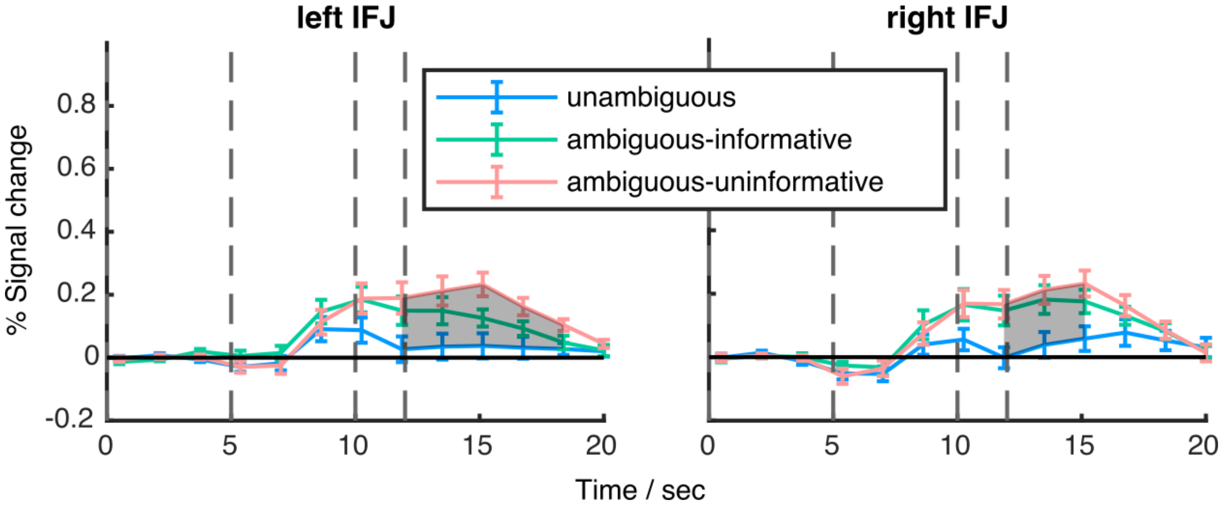

Figure 5. Time courses of activation in the IFJ. Time course of mean percent signal change during the contextual gaze following experiment of the IFJ (error bars are SEM). Areas in which conditions showed significant differences are shaded (permutations test, q < 0.05).

Discussion

In this study, we carried out two experiments in an attempt to clarify the processes allowing us to select the object of interest to the other in case the other´s gaze vector hits more than one object. In experiment 1 we delineated two brain regions (GFP and hLIP) known from previous work to be involved in processing the other´s gaze direction and in shifting spatial attention respectively (Sereno et al. 2001; Materna et al. 2008; Marquardt et al. 2017). In experiment 2, subjects were tested on a *contextual gaze following* task in which they needed to integrate gaze direction and auditory information in an attempt to disambiguate sets consisting of several objects hit by the other´s gaze in order to identify the target object. While BOLD activity of the GFP was not modulated by the ability of the auditory information to disambiguate the object set, hLIP showed increased activity when the information provided was insufficient to specify the target. The BOLD contrast between the condition unambiguously specifying the targets and the two ambiguous conditions identified yet another area only involved in contextual gaze following, missed by the GFP-localizer paradigm lacking the need to disambiguate the object choices. This area exhibited a continuously elevated response if and only if the evidence about the target was low. Unlike the other two areas, the IFJ did not show a general response to events of the trials in all experimental conditions; apart from an initial bump resembling the early part of the activity profiles during the two ambiguous conditions, its activity during *unambiguous* trials was close to or undistinguishable from baseline activity.

This study confirms our previous finding that the GFP located close to the pSTS plays a major role in processing information on the others' gaze. The present work shows that no matter if one or more potential target objects are hit by the other´s gaze vector, the BOLD activity in the GFP is the same. The need to differentiate between objects in case more than one lies on the gaze vector requires contributions from additional areas that exhibit differential activity. One of these areas, the hLIP in the posterior parietal lobe is also activated in the traditional, restricted gaze following paradigms in which the gaze hits one object only. hLIP is necessary for the control of spatial attention (Corbetta and Shulman 2002).

Work on monkey area LIP, arguably homologous to hLIP, has suggested that this area constitutes a priority or saliency map providing a representation of the environment that highlights locations that serve as attractors of attention. The saliency map may be modulated by bottom-up sensory cues, symbolic cues or gaze cues (Walther and Koch 2006; Bisley and Goldberg 2010). The latter is suggested by single unit recordings from area LIP. Many LIP neurons respond to the appearance of a gaze cue provided the gazed-at location lies within the neuron´s receptive field (Shepherd et al. 2009). Spatial selectivity for gazed-at locations and objects at these locations is also exhibited by many neurons in monkey GFP (Ramezanpour and Thier 2019). However, unlike neurons in area LIP, those in the GFP are selective for gaze-direction cueing and do not respond to bottom-up sensory cues highlighting a specific spatial location. This selectivity suggests that the priority map in LIP might draw on input from the GFP. The yoked activation of the hLIP/LIP and the GFP in BOLD imaging studies of gaze following is in principle in accordance with this scenario (Materna et al. 2008; Shepherd et al. 2009; Marquardt et al. 2017). However, the poor temporal resolution of the BOLD signals does not allow us to critically test if the assumed direction of information flow holds true. In any case, bidirectional projections are known to connect monkey area LIP and parts of the STS (Seltzer and Pandya 1994). One well-established pathway links area LIP and PITd, an area in the lower STS, probably close to the GFP, known to contribute to the maintenance of sustained attention (Stemmann and Freiwald 2016; Sani et al. 2019). Yet, the anatomical data available does not allow us to decide if the GFP does indeed contribute to this fiber bundle.

In the present study the BOLD signal evoked by gaze following in the hLIP was overall much stronger than in the GFP. Moreover, unlike the GFP signal, it exhibited a clear dependence on the conditions of the *contextual gaze following* experiment. Higher activity was associated with the *ambiguous-informative* and the *ambiguous-uninformative* conditions, both associated with unresolved uncertainty about the object requiring a decision of the participant that could only partially be based on information provided by the cue. Why should a region thought to coordinate spatial shifts of attention show an influence of target ambiguity, i.e. the need to choose between several potential targets? One possible answer may be that the higher hLIP activity reflects an increased attentional load. More specifically, increased uncertainty in ambiguous trials may have prompted more shifts of attention from one object to the other in an attempt to resolve the ambiguity. Although we found no difference in the number of exploratory saccades after the *go*-signal across conditions, we cannot rule out that participants covertly shifted attention between targets in ambiguous trials more than in the other trials. However, a more parsimonious explanation could be that hLIP constitutes a neural substrate for making decisions under uncertainty independent of the attentional load as suggested by several studies such as by Vickery and Jiang (2009).

The BOLD signal in an area we identified as the IFJ (between premotor cortex (BA 6), BA 44 and BA 8) exhibited a dependency on condition as well. However, the time course analysis revealed a fundamental difference compared to response profiles of BOLD activity in hLIP or the GFP. Sustained activity could only be observed in trials of the two *ambiguous* conditions, i.e. when the participants needed to make decisions under sensory uncertainty. This suggests that the condition dependency of the IFJ signal may be a consequence of shifts of attention between the two object categories, houses and hands. This interpretation draws on an MEG-fMRI study carried out by Baldauf and Desimone that demanded the allocation of attention to distinct classes of visual objects such as faces and spatial scenes (Baldauf and Desimone 2014). Depending on the object of attention, gamma band activity in the IFJ was synchronized either with the fusiform face area (FFA) or the parahippocampal place area (PPA). Additional support for this view comes from spatial cueing paradigms, which suggest that the IFJ primarily supports transient attentional processes, such as covert attentional shifts (Asplund et al. 2010; Tamber-Rosenau et al. 2018). We speculate that the time course of activity in the IFJ reflects the coordination of covert shifts of attention until the choice for the saccade target is made. In unambiguous trials, the lack of ambiguity allows fast decisions and since no attentional shifts are necessary the IFJ is not required.

The functional characteristics of the GFP, hLIP and the IFJ attribute complementary functions to each area which, in sum, allows gaze following under sensory ambiguity. We propose that information on the direction of the other´s gaze is provided by the GFP and modulates the saliency map generated by area hLIP such that spatial positions in the direction of the gaze vector are highlighted. In this situation the choice which of the possible objects is the most relevant one requires the resolution of uncertainty which is accomplished by the IFJ. In this scenario the intersection between the spatial information provided by the GFP-hLIP complex and the object-based information provided by the IFJ singles out one object that then will become the target of the observer´s gaze following response, elicited by the hLIP.

Several points need to be addressed by future work in order to test and to further refine this concept. As a first step, it will be necessary to investigate the temporal interplay between these regions in an attempt to establish causal interactions in order to critically test the model. Our hypothesis assumes that the IFJ has a leading role in processing information on competing objects on the gaze vector, resolving the uncertainty as to which one the target is. The conclusion that IFJ has a leading role in the disambiguation of the object set is primarily based on the fact that ambiguity-related information arises first in IFJ and only later in hLIP. Yet, we cannot rule out that this sequence might be an artifact of region-specific differences in the statistical power of the BOLD time course analysis, eventually in conjunction with region-specific differences in the variability of BOLD signal latencies.

***Funding*:** This work was supported by the Deutsche Forschungsgemeinschaft (TH 425/12-2).

***Acknowledgments*:** We are grateful to Friedemann Bunjes and Michael Erb for technical support.

***Author Contributions*:** PT developed the conceptual framework of the research. PT, PK and HR designed the experiments. PK and PWD performed the experiments. PK and MG analyzed the data. All authors contributed to the interpretation of results and the writing.

Competing Interests Statement: All authors declare to have no competing interests of any sort.

References

Asplund CL, Todd JJ, Snyder AP, Marois R (2010) A central role for the lateral prefrontal cortex in goal-directed and stimulus-driven attention. Nature Neuroscience. 13:507–512.

Atabaki A, Marciniak K, Dicke PW, Thier P (2015) Assessing the precision of gaze following using a stereoscopic 3D virtual reality setting. Vision Res. 112:68–82.

Baldauf D, Desimone R (2014) Neural Mechanisms of Object-Based Attention. Science. 344:424–427.

Baron-Cohen S (1994) How to build a baby that can read minds: Cognitive mechanisms in mind reading. Curr Psychol Cogn. 13:513–552.

Baron-Cohen S (1995) Mindblindness: An essay on autism and theory of mind, Mindblindness: An essay on autism and theory of mind. Cambridge, MA, US: The MIT Press.

Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological). 57:289–300.

Benz PF (2008) Erstellung einer Software zur partiellen Virtualisierung von Experimentierumgebungen für die perzeptionelle Blickrichtungsbestimmung (Master’s Thesis).

Bisley JW, Goldberg ME (2010) Attention, Intention, and Priority in the Parietal Lobe. Annual Review of Neuroscience. 33:1–21.

Bock SW, Dicke PW, Thier P (2008) How precise is gaze following in humans? Vision Res. 48:946–957.

Butterworth G, Jarrett N (1991) What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy. British journal of developmental psychology. 9:55–72.

Corbetta M, Shulman GL (2002) Control of goal-directed and stimulus-driven attention in the brain. Nature Reviews Neuroscience. 3:201–215.

Hollander G de, Knapen T (2017) Nideconv. Vrije Universiteit and the Spinoza Centre for Neuroimaging.

Hübner CVB (2008) Erstellung einer Software zum Design virtueller Experimentumgebungen: Bestimmung des Referenzsystems der visuellen Verarbeitung artifizieller Blicke (Master’s Thesis).

Kobayashi H, Kohshima S (1997) Unique morphology of the human eye. Nature. 387:767–768.

Lacadie CM, Fulbright RK, Rajeevan N, Constable RT, Papademetris X (2008) More accurate Talairach coordinates for neuroimaging using non-linear registration. NeuroImage. 42:717–725.

Laube I, Kamphuis S, Dicke PW, Thier P (2011) Cortical processing of head- and eye-gaze cues guiding joint social attention. Neuroimage. 54:1643–1653.

Liuzza MT, Cazzato V, Vecchione M, Crostella F, Caprara GV, Aglioti SM (2011) Follow My Eyes: The Gaze of Politicians Reflexively Captures the Gaze of Ingroup Voters. PLOS ONE. 6:e25117.

Marquardt K, Ramezanpour H, Dicke PW, Thier P (2017) Following Eye Gaze Activates a Patch in the Posterior Temporal Cortex That Is Not Part of the Human “Face Patch” System. eNeuro. 4:1–10.

Materna S, Dicke PW, Thier P (2008) Dissociable Roles of the Superior Temporal Sulcus and the Intraparietal Sulcus in Joint Attention: A Functional Magnetic Resonance Imaging Study. J Cogn Neurosci. 20:108–119.

Ramezanpour H, Marciniak K, Dicke PW, Thier P (2014) Neurons in the posterior STS extract facial information for the guidance of gaze following and the establishment of joint attention. Society for Neuroscience Abstract.

Sani I, McPherson BC, Stemmann H, Pestilli F, Freiwald WA (2019) Functionally defined white matter of the macaque monkey brain reveals a dorso-ventral attention network. eLife. 8:e40520.

Seltzer B, Pandya DN (1994) Parietal, temporal, and occipita projections to cortex of the superior temporal sulcus in the rhesus monkey: A retrograde tracer study. J Comp Neurol. 343:445–463.

Sereno MI, Pitzalis S, Martinez A (2001) Mapping of Contralateral Space in Retinotopic Coordinates by a Parietal Cortical Area in Humans. Science. 294:1350–1354.

Shepherd SV, Klein JT, Deaner RO, Platt ML, Shepard RN (2009) Mirroring of attention by neurons in macaque parietal cortex. Proc Natl Acad Sci. 106:9489–9494.

Stan Development Team 2018 PyStan: the Python interface to Stan.

Stemmann H, Freiwald WA (2016) Attentive Motion Discrimination Recruits an Area in Inferotemporal Cortex. J Neurosci. 36:11918–11928.

Tamber-Rosenau BJ, Asplund CL, Marois R (2018) Functional dissociation of the inferior frontal junction from the dorsal attention network in top-down attentional control. Journal of Neurophysiology. 120:2498–2512.

Teufel C, Alexis DM, Clayton NS, Davis G (2010) Mental-state attribution drives rapid, reflexive gaze following. Atten Percept Psychophys. 72:695–705.

Teufel C, Alexis DM, Todd H, Lawrance-Owen AJ, Clayton NS, Davis G (2009) Social Cognition Modulates the Sensory Coding of Observed Gaze Direction. Curr Biol. 19:1274–1277.

Vickery TJ, Jiang YV (2009) Inferior Parietal Lobule Supports Decision Making under Uncertainty in Humans. Cereb Cortex. 19:916–925.

Walther D, Koch C (2006) Modeling attention to salient proto-objects. Neural Networks, Brain and Attention. 19:1395–1407.

Chapter 3

A neural substrate for volitional control of gaze following

Maria-Sophie Breu*, Hamidreza Ramezanpour*, Peter Dicke, Peter Thier (*equal contribution)

Under Review at Neuropsychologia

We also tried to address the hypothesis that the volitional control of gaze following demanded by specific behavioral needs may be a consequence of prefrontal control of the GFP. In order to identify the cortical substrate of cognitive control of gaze following behavior, we carried out an event-related fMRI experiment, in which human subjects were exposed to social gaze cues in two distinct contexts: a normal gaze following condition in which subjects had to use social gaze cues to shift their attention to the gazed-at spatial targets, or, alternatively, a control condition in which the subjects had to ignore the direction of gaze cues and shift their attention to the same spatial targets according to a color-matching rule. We could identify BOLD activity in two frontal brain areas when gaze following had to be suppressed: dorsolateral prefrontal cortex (DLPFC), part of Brodmann area 46, and the anterior cingulate cortex (ACC). Our results suggest that DLPFC and ACC play a central role in the context-dependent control of human gaze following behavior.

A neural substrate for volitional control of gaze following

M.-S. Breu[a,*], H. Ramezanpour[a,b,c,*] , P. W. Dicke[a] and P. Thier[a,d]

[1]Department of Cognitive Neurology, Hertie Institute for Clinical Brain Research, University of Tübingen, 72076 Tübingen, Germany.

[2]Graduate School of Neural and Behavioral Sciences, University of Tübingen, 72074 Tübingen, Germany.

[3]International Max Planck Research School for Cognitive and Systems Neuroscience, University of Tübingen, 72074 Tübingen, Germany.

[4]Werner Reichardt Centre for Integrative Neuroscience, University of Tübingen, 72076 Tübingen, Germany.

* These authors contributed equally to this work.

Correspondence to: Peter Thier or Hamidreza Ramezanpour, Department of Cognitive Neurology, Hertie Institute for Clinical Brain Research, Hoppe-Seyler-Str. 3, 72076 Tübingen, Germany. E-mail: thier@uni-tuebingen.de; hamidreza.ramezanpour@uni-tuebingen.de.

Abstract

Gaze following is an essential part of non-verbal communication and indispensable for successful social interactions. Human gaze following is a fast and almost reflex-like behavior, yet, it can be volitionally controlled and suppressed to some extent if inappropriate or unnecessary given the social context. In order to identify the neural basis of the cognitive control of gaze following, we carried out an event-related fMRI experiment, in which human subjects were exposed to social gaze cues in two distinct contexts: a baseline gaze following condition in which subjects were encouraged to use gaze cues to shift their attention to a gazed-at spatial target and a control condition in which the subjects were required to ignore the gaze cue and instead to shift their attention to a distinct spatial target to be selected based on a color-mapping rule, requiring the suppression of gaze following. We could identify suppression-related BOLD activity in the dorsolateral prefrontal cortex (dlPFC) and the anterior cingulate cortex (ACC). These results are in line with the established role of the dlPFC and ACC in executive control and conflict monitoring.

Keywords: gaze following, cognitive control, dorsolateral prefrontal cortex, anterior cingulate cortex, gaze following patch

Introduction

Humans have developed a complex communication system based on information provided by the face and the eyes (Andrew 1963; Kobayashi and Kohshima 1997; Emery 2000). Prompted by the other´s gaze direction, determined by the direction of the eyes and the head, human observers shift their focus of attention to the object of interest to the other, thereby establishing joint attention to the object with the other. This ability is so important because it allows the observer to map her/his object-related aspirations and intentions on to the other, thereby establishing a *Theory of* (the other´s) *Mind* (Baron-Cohen 1994; Perrett and Emery 1994; Baron-Cohen 1995; Emery 2000; Langton and Bruce 2000)*.*

Gaze following is a fast and quasi reflex-like behavior that emerges very early during ontogeny (Friesen and Kingstone 1998; Hood, Willen et al. 1998; Driver, Davis et al. 1999; Baki, Baron-Cohen et al. 2000; Langton and Bruce 2000), hence meeting Fodor’s criteria of a domain-specific, probably largely innate capacity (Fodor 1983). Although we probably always feel a certain urge to follow the other´s gaze, we are able to control gaze following if alternative behaviors might be more pertinent in a given moment. For instance, following the other´s gaze to her or his object of desire would be highly inappropriate if all of a sudden something dangerous appeared on the scene requiring the observer´s full attention. However, not only the significance of competing stimuli modifies the willingness to follow the other´s gaze but also the other´s identity and the affective links between the two agents. For instance, as shown by Liuzza and coworkers, observers are more poised to follow the gaze of their favorite political leader than the gaze of the representative of an opposing party (Liuzza, Cazzato et al. 2011). Hence, gaze following is embedded into a broader behavioral context and can only be understood if we learn how pertinent contextual information is integrated (Ristic and Kingstone 2005). Previous fMRI studies have implicated a patch of cerebral cortex (the “gaze following patch”, GFP) in the posterior superior temporal sulcus (STS) in the geometric calculations for shifting one´s own gaze in accordance to the gaze orientation of the counterpart (Hoffman and Haxby 2000; Materna, Dicke et al. 2008; Marquardt, Ramezanpour et al. 2017; Kraemer, Görner et al. 2019). In this study, we tried to address the hypothesis that the volitional control of gaze following demanded by specific behavioral requirements may be a consequence of prefrontal control of the GFP. To this end, we performed an event-related fMRI experiment that allowed us to compare activation patterns evoked by gaze following and its rule-based suppression.

Methods

Subjects

Twenty subjects (10 female, 10 male) participated in our study. Subjects were between 20 and 32 years old, right-handed and had normal or corrected-to-normal (lenses) vision. The study was approved by the Ethics Review Board of the Tübingen Medical School and complied with the guidelines of the Declaration of Helsinki. All subjects received oral and written information and provided written consent to participate in our study.

Paradigm

The images offering gaze stimuli were the same as used by Marquardt and coworkers (2017). They were portrait photographs of a white, caucasian female ("sender") and manipulated using Adobe Photoshop 7.0. The portrait shown in the fixation period of a trial was the female face ("fixation portrait") in front of a random pattern background (gray and black dots) with her eyes straight ahead and a green iris (Figure 1A). In front of her were five targets, all the same in size and shape, but each with a different color (from left to right: dark blue, light blue, green, light brown, dark brown). The visual angle between the targets was 12.5° for the sender portrait. For the subsequent spatial task epoch, the portrait was manipulated in two ways (Figure 1B): first, the eye gaze direction could change to hit one of the four outer targets or, alternatively, stay on the central target. Second, the color of the eyes could change to dark blue, light blue, light brown, dark brown or stay green, corresponding to the color of one of the five targets. In our experiment subjects were instructed to perform two different tasks. In "gaze following trials" subjects were asked to execute a saccade to the target the portrait was looking at, ignoring the color of the iris. In "color mapping trials" subjects were conversely asked to perform a saccade to the target corresponding to the color of the iris of the sender, this time ignoring the direction of the eyes.

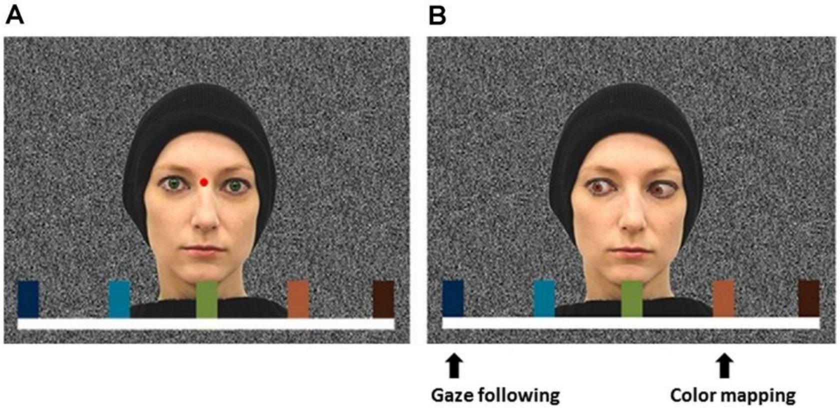

Figure 1. A) Picture presented in the fixation period: Portrait of a female with green eyes looking straight ahead. In front of her, 5 target objects with distinct colors are arranged. B) Picture presented in the subsequent spatial attention period: The actor is shown gazing at one of the targets. Moreover, the color of

the iris changed such as to match the color of one of the objects. The arrows are pointing to the target depending on the task rule named below.

Gaze following and color mapping trials were presented in a pseudo-randomized manner, allowing no more than three consecutive trials chosen from the same condition. At the beginning of each trial, a written rule was provided to inform the subjects about the upcoming condition (Figure 2). Between subsequent trials, there was a randomly varying interval of 14–15 seconds in which only the red fixation point (dimension: 0.3°) was presented on the otherwise black screen. The long intertrial intervals were chosen to minimize the spillover of BOLD responses from a preceding trial on a given trial (Dale, 1999; Bandettini and Cox, 2000). Subjects were asked to keep their eyes fixating on the red fixation point whenever visible. A trial started with the presentation of the written rule, followed by the onset of the fixation point. After a delay of 1–5 seconds, the fixation picture appeared for 5 seconds, followed by the spatial task portrait available for 4 seconds. The red fixation point was constantly on until 1 second after the appearance of the task picture. The offset of the fixation point was the go-signal for subjects to make a saccade to the spatial target identified by the conjunction of the spatial information provided by the sender portrait and the prevailing rule (i.e. gaze following vs. color mapping). Each subject performed 90 trials.

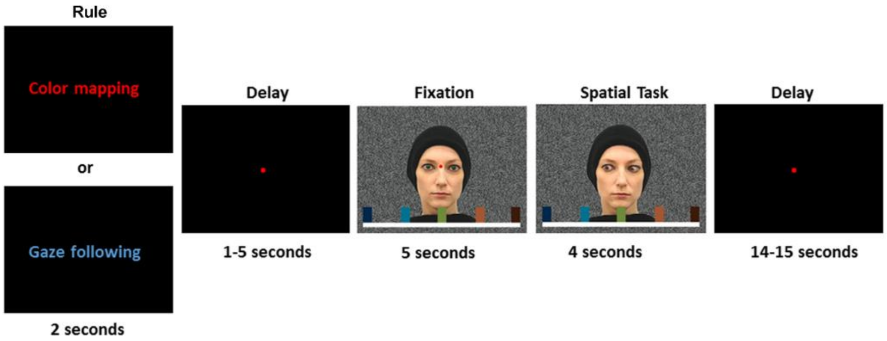

Figure 2. The sequence of events in a trial. A trial started with the presentation of the instruction specifying which rule to apply in order to identify the target in the upcoming trial. Following a further delay of 1–5 seconds the fixation picture appeared, stayed on for 5 seconds and then was replaced by the spatial task picture. One second later, the fixation point disappeared, serving as go-signal to perform a targeting saccade. The consecutive trial started after an intertrial period of 14–15 seconds.

fMRI Recording

Prior to the fMRI experiment, subjects completed a training session involving the behavioral paradigms discussed before. The session took place in a darkened room and lasted approximately 45 minutes. Participants were seated on a comfortable chair in front of a screen (distance: 90cm, dimension: 120cm x 80cm, size of images presented: 40cm x 30cm projected from the back by a beamer and were asked to rest their head in a chin rest to prevent head movement.

Scanning took place 1–5 days later. Subjects lay supine in the MRI scanner and their heads were fixed by foam rubber to minimize head movements. Visual stimuli (dimension: 45cm x 34cm) were backprojected on a translucent screen positioned behind the subject and seen via a mirror attached to the head coil. The resulting viewing distance between observer and image was 102cm. Images were acquired by a 3-Tesla MRI scanner (Prisma, Siemens, Erlangen, Germany) using a 12-channel head coil (acquisition matrix: 64x64). A volume of approximately 1200 T2-weighted echo-planar (epifid) images (TR: 3000ms, TE: 35ms, TA: 2.93s, flip angle: 90°) was taken. The images covered the whole brain (44 transverse slices, slice order: [44:-1:1], slice thickness: 2.5mm, gap: 0.5mm, pixel spacing: 3mm x 3mm). Additionally, anatomical T1-weighted images were taken for each subject, using a magnetization prepared, rapid acquisition gradient-echo sequence (mprage) (TE: 2.96ms, TR: 2300ms, TI: 1100ms, flip angle: 8°, voxel size: 1.0mm x 1.0mm x 1.0mm).

Vertical and horizontal eye movements were recorded during both training and scanning sessions. Eye position recordings during training were acquired using a the Cronos Vision C-ET video eye tracker. During scanning, we deployed a certificated, MRI-compatible eye-tracker (SMI iView X™ MRI-LR; sampling rate of 60 Hz). Calibration of the eye-tracker output was performed three times during the experiment. To this end subjects had to alter fixation between nine positions on the screen, allowing the comparison of known spatial position and tracker output.

Data Analysis

The whole stack of images of each subject was preprocessed and analyzed deploying the SPM8 statistic parametric mapping software (Welcome Department of Cognitive Neurology, London, UK, http://www.fil.ion.ucl.ac.uk/spm/).

For preprocessing functional images were first realigned and slice time corrected. Anatomical images, mean image and functional images were coregistered to enlarge mutual information. Anatomical images were segmented using templates provided by SPM (T1.nii 1) and used to normalize functional images. Finally, functional images were spatially smoothed using a full-width half-maximum Gaussian filter (FWHM: 6mm).

Data analysis was performed by modeling the events of the two tasks (gaze following and color mapping) with a canonical hemodynamic response function and applying the general linear model (GLM). As onset times we used the appearance of the spatial task picture or the appearance of the rule. Regressors representing estimated head movements (translation and rotation with six degrees of freedom) were added to the model as covariates of no interest to reduce the influence of head movements during scanning. In order to eliminate slow, not task-related fluctuations/changes, the BOLD signal was high-passed filtered (cut off frequency 1/128Hz). For each subject two contrasts were calculated: *gaze following* versus *color mapping* during the spatial task as well as *color mapping* versus *gaze following* aligned to the onset of the rule in each trial. Significant changes were assessed using t-statistics.

In order to establish the response pattern for the group of subjects, single-subject contrasts were analyzed on a second level using a random effects model that compared the average activation for a given voxel with the variability of that activation in the examined population (Friston 1995, Friston, Holmes et al. 1999). BOLD responses were considered significant and reported if the statistical significance exceeded p<0.01 on the level of single voxels and, moreover, involved clusters of more than 10 neighboring voxels. We visualized these responses on the SPM template of single_subj_T1. For the time course analysis, the percent signal change of the BOLD signal was calculated using SPM toolbox NERT4SPM (NERT4SPM; by Axel Lindner and Christoph Budziszewski; https://svn.discofish.de/MATLAB/spmtoolbox/NERT4SPM) in regions of interests delineated in individual subjects (ROIs). To this end, we determined the individual peak activity in a region encompassing a region of 3–10 voxels radius, centered on the peak activity of the group, satisfying a statistical significance threshold of $p < 0.05$, uncorrected and a minimum size of significant voxels of >= 6 voxel). In case more than one peak was found, we chose the one closest to the peak of the group. Differences between BOLD responses in this ROI for individual points in time were considered significant based on a running paired t-test with a threshold of $p<0.05$.

Results

Behavioral performance

Eye data were available for 19 out of 20 subjects, allowing us to assess the percentage of correct target-directed saccades and the measurement of their latencies relative to the disappearance of the red fixation point. We used the time of peak saccade velocity as a proxy of saccade onset. Although this measure certainly overestimated saccade onset times, it had the advantage of substantially reduced variance. In order to exclude predictive saccades, not necessarily driven by the spatial information provided by the paradigm, we excluded saccades with reaction times less than 200ms. There was no significant difference (2-way ANOVA, $p>0.05$) between the two conditions, neither for the percentage of correct saccadic choices (gaze following: mean: 83.4%, SD: 13.3%; color mapping: mean: 82.2%, SD: 13.3%) nor for saccadic reaction times (gaze following: mean: 573.7ms, SD: 154.3ms; color mapping: mean: 560.2ms, SD: 120.3ms), indicating that both tasks were experienced equally demanding (Figure 3).

In one out of the 20 subjects, the eye position records were too noisy to allow a reliable judgment of target choices and reaction times. In view of the fact that the aforementioned behavioral analysis of the other 19 subjects had demonstrated good performance without exception, we nevertheless decided to consider the fMRI data of this subject as well.

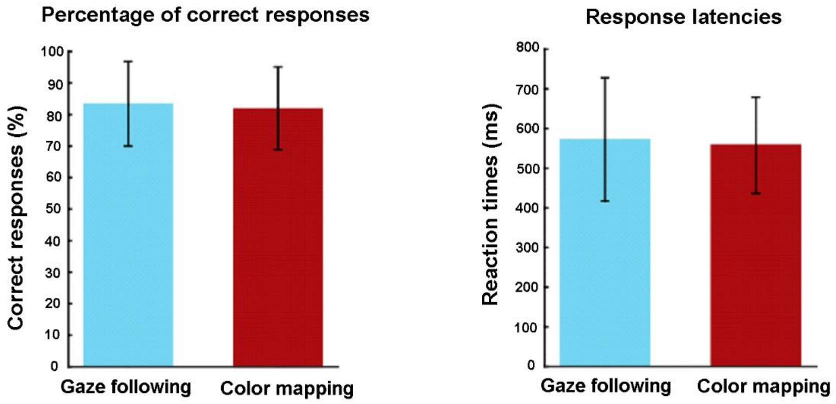

Figure 3. Behavioral performance: There was no significant difference (ANOVA, $p<0.05$) in behavioral performance between gaze following and color mapping trials, neither with respect to the number of correct trials (left panel) nor with respect to the response latencies (right panel).

BOLD results

As a first step in testing our hypothesis that the volitional control of gaze following is a consequence of prefrontal control of the GFP, we identified the gaze following patch (GFP) in the posterior superior temporal sulcus in both hemispheres, characterized by significantly larger BOLD responses to gaze following as compared to color mapping. The peak activity of the group based contrast was found at coordinates [x,y,z]=[48, -37, 15] and [x,y,z]=[-45, -34, 20] in the right and left respectively ($p<0.02$ uncorrected, cluster size: >=10 adjacent voxels). These coordinates were similar to those previously found in (Materna, Dicke et al. 2008, Marquardt, Ramezanpour et al. 2017). Considering the statistical criteria laid out in the methods section allowed us to delineate individual GFPs in 14 out of the 20 subjects and to define individual ROIs. The radius of the individual ROIs ranged from 3 to 10 voxels and the coordinates of the individual peak contrasts scattered between 42 to 55 for x, -30 to -44 for y, and 8 to 22 for z.

The next step was to determine BOLD signals related to the rule to follow gaze or, alternatively, to suppress it by mapping eye color. To this end, we performed a whole-brain search for significant changes in the BOLD signal in the color mapping task in comparison to the gaze following task. We reasoned that the preparatory rule signals must be established before the actual spatial task in order to have enough time to act on the reflexive gaze following responses. Hence, we looked at the contrast activity at the time of each trial rule. Significant rule-related activity was found in two cortical regions in the right hemisphere (Figure 4), the right dorsolateral prefrontal cortex (dlPFC; MNI coordinates of the maximum activity at the group level: [x,y,z]=[42, 32, 30], part of the Brodmann area 46) and the right anterior cingulate cortex (ACC; MNI coordinates of the maximum activity at the group level: [x,y,z]=[15, 29, 18], part of the Brodmann area 32) (p<0.02 uncorrected, cluster size: >=10 adjacent voxels). To optimally visualize and measure the cortical representations, statistical *t*-maps were projected onto inflated and flattened reconstructions of cortical surface gray matter using Caret (http://brainvis.wustl.edu/wiki/index.php/caret).

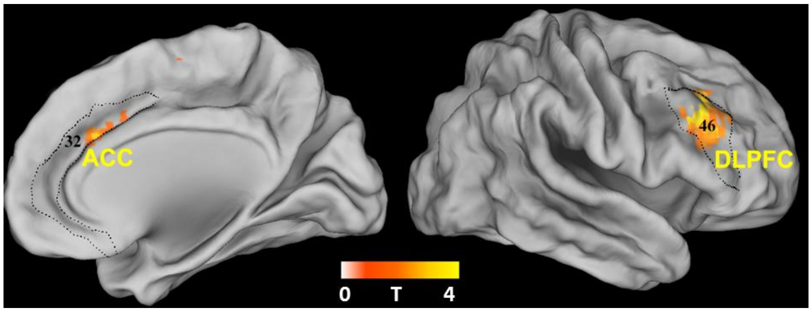

Figure 4. BOLD contrast in the rule period between responses to the rule to select a target based on eye color versus the rule to follow gaze. There was a significant contrast in the right dlPFC (part of Brodmann area 46) and the right ACC (part of Brodmann area 32), p<0.01 (uncorrected), cluster size: >10.

The time course analysis of the BOLD signals revealed that just after the rule to apply the color mapping rule, the BOLD signal started to rise in the dlPFC and the ACC relative to the BOLD signal in gaze following trials within 5 seconds after the rule peaking at 10 seconds (Figure 5, 6A). On the other hand, being instructed to follow gaze in gaze following trials did not evoke a significant BOLD response relative to the baseline in the two regions exhibiting activity related to the rule to deploy the color mapping rule. A preparatory BOLD signal following the presentation of the color mapping rule might be related to a specific contribution to the processing of color and its association with particular targets. On the other hand, it might reflect the need to suppress gaze following if color mapping is called for. In contrast to the BOLD signals in these two cortical areas, BOLD activity evoked by the presentation of the rule did not differ for the two rules in the GFP (Figure 6C). For both, it showed similar fluctuations until around the presentation of the face with averted eyes (onset of the spatial task). Only then did the BOLD signals start to diverge. As to be expected based on the definition of the GFP captured by the ROI, the gaze following-related BOLD signal surpassed the one for color mapping (Figure 6D). If the assumption is correct that the preparatory rule-related signal in the prefrontal cortex and the ACC is associated with the volitional suppression of gaze following if impertinent one might expect to see an influence on the signal in the GFP. The assumption here is that cognitive control is based on a control of signals orchestrating gaze following in the GFP. In order to identify the impact of prefrontal cortex and the ACC on the GFP, we calculated pairwise correlations. However, the correlation between the average signal changes in the GFP and the two frontal regions during the availability of the rule failed to reach significance (Spearman correlation, p>0.05).

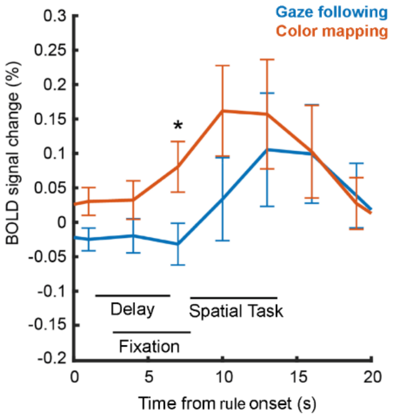

Figure 5. Comparison of the BOLD signal as function of time after the appearance of the rule (t=0) in the ACC. Only the BOLD response to the rule to rely on eye color (red) rises significantly but not the one associated with the gaze following rule (blue). The asterisk indicates the time bin in which the difference between two conditions reached a significant level (paired t-test, p<0.05)

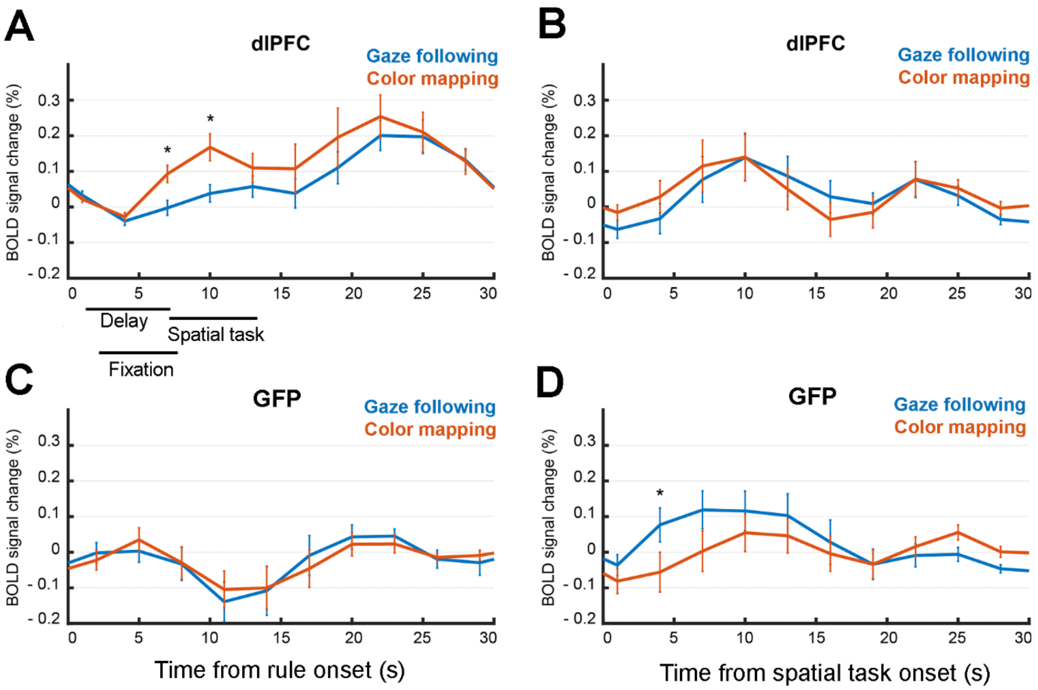

Figure 6. A) BOLD signal as function of time in the right dlPFC after the appearance of the rule (t=0). The signal rises in the right dlPFC only after the rule to map eye color (red), but not after the rule to follow gaze (blue). B) BOLD signal as function of time in the right dlPFC after the presentation of the spatial task picture (t=0). The curves do not exhibit differences between the color mapping and gaze following trials C) BOLD signal change in the right GFP after the appearance of the rule. There are no significant differences in the activity level for color mapping and gaze following trials. D) BOLD signal change in the right GFP after the presentation of the spatial task picture. The activity rises only in the gaze following trials, but not in color mapping trials (significant differences (paired t-test, p<0.05) are marked by asterisks). The horizontal lines between figure parts A and C indicate the potential onset of the various task events considered for the analysis. Note that because of the variable time of 1-5 seconds before the appearance of the initial fixation picture, the activity in the period following the offset of the fixation period is no longer properly aligned in the plots of rule related activity. Conversely, the activity in the period before the onset of the spatial task is based on trials that are not properly aligned before the onset of the spatial task image. Asterisks indicate time bins with a significant level of difference between conditions (paired t-test, p<0.05.

Discussion

Complementary roles of dlPFC and ACC in the cognitive control of gaze following

We deployed an event-related fMRI design in an attempt to identify cortical areas exhibiting BOLD signals related to the need to suppress gaze following if not pertinent. In our experiment, the need to suppress a gaze following response was a consequence of the rule to ignore the other´s gaze and instead to use the other´s eye color to shift attention to locations associated with particular eye colors based on prior learning. The presence of the rule to suppress gaze following was associated with the build-up of a BOLD signal in two prefrontal areas, dorsolateral prefrontal cortex (dlPFC) and the anterior cingulate cortex (ACC). BOLD activity in the gaze following patch (GFP) in the superior temporal sulcus (STS), well-known to be involved in the translation of gaze cues into an appropriate gaze following response (Materna, Dicke et al. 2008; Marquardt, Ramezanpour et al. 2017) exhibited the expected positive gaze following vs. color mapping contrast confined to the period around the shift of attention prompted by the other´s gaze, yet unaffected by the preceding rule.

The dlPFC is connected with a wide range of neocortical areas, garnering input from any sensory modality and in turn projecting to cortical and subcortical areas orchestrating purposeful behavior (Miller and Cohen 2001). However, the dlPFC is anything but a structure underlying elementary sensorimotor transformations. This is clearly indicated by the numerous non-sensorimotor influences on neuronal activity in the dlPFC such as information on past events and experiences (Shimamura 1995), expected reward (Leon and Shadlen 1999), *a-priori* information on object features, places of particular interest (Pochon, Levy et al. 2001; Lebedev, Messinger et al. 2004) or knowledge of the value of behavior checked against the subject´s needs (Duncan, Emslie et al. 1996). As the dlPFC has access to information on past events, bodily needs, and future ambitions, it is in a position, well suited to modify the behavioral impact of the flood of sensory signals raining down on the subject in each moment, taking the longer-term interests of the subject into account (Nauta 1971; Duncan, Emslie et al. 1996). It is this ability to cognitively control behavior that frees us from the inevitability of automatic or reflex-like behaviors facilitated by powerful preformed sensorimotor pathways (MacLeod 1991; MacDonald, Cohen et al. 2000; Miller and Cohen 2001; Miller, Freedman et al. 2002, Aron, Robbins et al. 2004). The need to choose a hard-learned behavior, the mapping of eye color onto distinct spatial positions rather than to release gaze following, an ontogenetically preformed reflex-like behavior as demanded in our experiment, is a paradigmatic manifestation of our ability to deploy cognitive control. Hence, the finding of significant BOLD activity in the dlPFC, evoked by the rule to select the target based on eye color and to suppress following the other´s gaze is in line with the well-accepted role of the dlPFC in cognitive control. One might have expected to see a correlation between the dlPFC BOLD signal associated with the requirement to suppress gaze following and the later shift of attention-related gaze BOLD signals in the GFP. The reasoning would be that a stronger color mapping-related BOLD signal in the rule period found in the dlPFC would be associated with a stronger color mapping-related BOLD signal in the spatial task period in the GFP. However, such a correlation could not be seen in our data, probably because the assumption is too simplistic. Assuming that the GFP serves as a hub for both gaze following and color mapping, a stronger dlPFC control signal, thought to strengthen the decision for color mapping may not necessarily change the color mapping-related activity in the spatial task period in the GFP. This may be so because the dlPFC control signal may have two largely compensatory consequences, namely an increase of color mapping-related neuronal activity but also an accompanying decrease in residual gaze following-related neural activity. However, assuming that more color rule-related BOLD activity in the dlPFC may reflect better cognitive control, one might expect to see fewer false decisions. Unfortunately, the number of error trials was too small to allow us to test if this prediction applied.

There is one serious caveat to the conclusion that the dlPFC BOLD response to the color mapping rule reflects cognitive control, namely the possibility that it may reflect a restricted contribution to identifying the target. In color mapping trials, eye color has to be extracted and compared with four possible color-target pairs, kept in memory. This comparison entails access to a long-term memory store but, arguably, also a component of working memory, required for the provision of eye color. Considering the well-established role of the dlPFC in working memory alluded to earlier (Goldman-Rakic 1988), BOLD activity elicited by the processing of the eye color rule may not be too surprising. Unfortunately, our paradigm does not allow us to decide between the two possible interpretations of the color mapping-related BOLD signal in the dlPFC. We may mention, though, that a role in accommodating working memory in the context of the color mapping task would of course not preclude a more general role in cognitive control.

Also, the ACC exhibited a significant BOLD response to the color response. Such a response is not unexpected in view of standard models of the role of the ACC that center around performance monitoring and response overriding (Kerns, Cohen et al. 2004). The latter term emphasizes a central aspect of cognitive control, namely the need to deploy processing resources to ignore a prepotent stimulus that otherwise prompts a quasi-automatic response like the other´s gaze in our task. Unlike suppression-related activity in the dlPFC, the one in the ACC may be more closely related to the subjective experience of the conflict of giving preference to a stimulus that is non-dominant. Numerous studies have demonstrated the occurrence of ACC activity in conjunction with the need to override such dominating stimuli (MacDonald, Cohen et al. 2000). Response monitoring is a complementary aspect of cognitive control, needed to optimize its efficiency (Botvinick, Braver et al. 2001). Activity in the ACC is known to be elicited in reaction to the occurrence of errors due to insufficient response suppression (Carter, Braver et al. 1998). Not surprisingly in light of the role of the dlPFC in mediating cognitive control, error-related activity in the ACC is known to be associated with an increase in dlPFC activity. Such a correlation supports the notion of a feedback loop helping to boost control signals in the dlPFC if needed and a conceptual model, in which the dlPFC is thought to implement the rule and the ACC to monitoring the quality of the rule-based performance. According to Botvinick et al., differences in the timing of activity in ACC may allow differentiation of contributions to response overriding and to performance monitoring with the former appearing earlier than the latter (Botvinick, Braver et al. 2001). In this vein, the early color mapping-related ACC activity observed in our experiment might be associated with response overriding, or more concretely, with the experience of the conflict when preferring the response to the weaker stimulus over the prepotent one. Later activity, in the period in which attention is actually shifted, may reflect the effort to detect errors, i.e. to monitor the performance.

In sum, our study suggests that the dlPFC helps to control gaze following by integrating contextual information for the suppression of gaze following in situations in which it may be inappropriate. The suppression of gaze following most probably involves the generation of bias signals in the dlPFC, broadcasted to the GFP and other dependent cortical structures representing the behavioral options. Furthermore, the need to control gaze following seems to entail an important contribution of the ACC in monitoring the rule dependent performance, arguably fine-tuning the control function of the dlPFC resorting to feedback from the ACC.

***Contribution statement*:** HR and PT designed the study; MSB and PWD collected the data. MSB and HR analyzed the data. MSB, HR, and PT wrote the manuscript.

***Acknowledgment*:** This work was supported by the Werner Reichardt Centre for Integrative Neuroscience (CIN) at the Eberhard Karls University of Tübingen. The CIN is an Excellence Cluster funded by the Deutsche Forschungsgemeinschaft (DFG) within the framework of the Excellence Initiative (EXC 307). Further support was granted by the German Research Foundation (DFG) project TH 425/12-2.

References

Andrew RJ (1963) Evolution of facial expressoin. Science 142(3595): 1034-1041.

Aron AR, Robbins TW, Poldrack RA (2004) Inhibition and the right inferior frontal cortex. Trends Cogn Sci 8(4): 170-177.

Baki A, Baron-Cohen S, Wheelwright S, Connellan J, Ahluwalia J, Kashima H (2000) Is there an innate gaze module? Evidence from human neonates. Infant Behavior and Development 23(3): 223-229.

Baron-Cohen S (1994) How to build a baby that can read minds: Cognitive mechanisms in mindreading. Cahiers de Psychologie Cognitive/ Current Psychology of Cognition 13: 513-552.

Baron-Cohen S (1995) Mindblindness: an essay on autism and theory of mind." MIT Press.

Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD (2001) Conflict monitoring and cognitive control. Psychol Rev 108(3): 624-652.

Carter CS, Braver TS, Barch DM, Botvinick MM, Noll D, Cohen JD (1998) Anterior cingulate cortex, error detection, and the online monitoring of performance. Science 280(5364): 747-749.

Driver J, Davis G, Ricciardelli P, Polly K, Maxwell E, Baron-Cohen S (1999) Gaze Perception Triggers Reflexive Visuospatial Orienting. Visual Cognition 6(5): 509-540.

Duncan J, Emslie H, Williams P, Johnson R, Freer C (1996) Intelligence and the frontal lobe: the organization of goal-directed behavior. Cogn Psychol 30(3): 257-303.

Emery NJ (2000) The eyes have it: the neuroethology, function and evolution of social gaze. Neurosci Biobehav Rev 24(6): 581-604.

Fodor J (1983) Modularity of Mind: An Essay on Faculty Psychology. MIT Press.

Friesen C, Kingstone A(1998) The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. Psychonomic Bulletin & Review 5(4): 490-495.

Friston KJ, Holmes AP, Worsley KJ (1999) How many subjects constitute a study? NeuroImage 10: 1-5.

Friston KJ, Worsley KJ, Poline JB, Frith CD, Frackowiak RS (1995) Statistical parametric mapping in functional imaging: A general linear approach. Human Brain Mapping 2: 189-210.

Goldman-Rakic PS (1988) Topography of cognition: parallel distributed networks in primate association cortex. Annu Rev Neurosci 11: 137-156.

Hoffman EA, Haxby JV (2000) Distinct representations of eye gaze and identity in the distributed human neural system for face perception. Nat Neurosci 3(1): 80-84.

Hood BM, Willen JD, Driver J (1998) Adult's eyes trigger shifts of visual attention in human infants. Psychological Science 9: 131-134.

Kerns JG, Cohen JD, MacDonald AW, Cho RY, Stenger VA, Carter C (2004) Anterior cingulate conflict monitoring and adjustments in control. Science 303(5660): 1023-1026.

Kobayashi H, Kohshima S (1997) Unique morphology of the human eye. Nature 387(6635): 767-768.

Kraemer P, Görner M, Ramezanpour H, Dicke PW, Their P (2019) A fronto-temporo-parietal network disambiguates potential objects of joint attention. bioRxiv 542555; doi: https://doi.org/10.1101/542555.

Langton SR, Bruce V (2000) You must see the point: automatic processing of cues to the direction of social attention. J Exp Psychol Hum Percept Perform 26(2): 747-757.

Lebedev MA, Messinger A, Kralik JD, Wise SP (2004) Representation of attended versus remembered locations in prefrontal cortex. PLoS Biol 2(11): e365.

Leon MI, Shadlen MN (1999) Effect of expected reward magnitude on the response of neurons in the dorsolateral prefrontal cortex of the macaque. Neuron 24(2): 415-425.

Liuzza MT, Cazzato V, Vecchione M, Crostella F, Caprara GV, Aglioti SM (2011) Follow my eyes: the gaze of politicians reflexively captures the gaze of ingroup voters. PLoS One 6(9): e25117.

MacDonald AW, Cohen JD, Stenger VA, Carter CS (2000) Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. Science 288(5472): 1835-1838.

MacLeod CM (1991) Half a century of research on the Stroop effect: an integrative review. Psychol Bull 109(2): 163-203.

Marquardt K, Ramezanpour H, Dicke PW, Their P (2017) Following eye gaze activates a patch in the posterior temporal cortex that is not part of the human "face patch" s ystem. eNeuro 4(2).

Materna S, Dicke PW, Their P (2008) Dissociable roles of the superior temporal sulcus and the intraparietal sulcus in joint attention: a functional magnetic resonance imaging study. J Cogn Neurosci 20(1): 108-119.

Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. Annu Rev Neurosci 24: 167-202.

Miller EK, Freedman DJ, Wallis JD (2002) The prefrontal cortex: categories, concepts and cognition. Philos Trans R Soc Lond B Biol Sci 357(1424): 1123-1136.

Nauta WJH (1971) The problem of the frontal lobe: a reinterpretation. J. Psychiatr. Res. 8: 167-170.

Perrett DI, Emery NJ (1994) Understanding the intentions of others from visual signals: Neurophysiological evidence. Current Psychology of Cognition 13(5): 683-694.

Pochon JB, Levy R, Poline JB, Crozier S, Lehericy S, Pillon B, Deweer B, Le Bihan, Dubois B (2001) The role of dorsolateral prefrontal cortex in the preparation of forthcoming actions: an fMRI study. Cereb Cortex 11(3): 260-266.

Ristic J, Kingstone A (2005) Taking control of reflexive social attention. Cognition 94(3): B55-65.

Shimamura AP, (1995) Memory and the prefrontal cortex. Ann N Y Acad Sci 769: 151-159.

Chapter 4

Decoding of the others’ focus of attention by a temporal cortex module

Hamidreza Ramezanpour, Peter Thier

Under Review at PNAS (revision prompted), and available on BioRxiv,
doi: https://doi.org/10.1101/681957, *2019*

We electrophysiologically explored the GFP in monkeys and studied the properties of neurons in this region using a battery of highly controlled behavioral paradigms and showed that we are able to delineate a distinct part of cortex in the fundus of the middle-posterior superior temporal sulcus, congruent with the *gaze following patch (GFP),* identified by BOLD imaging. This patch is characterized by the presence of gaze following neurons and other types of neurons, exhibiting a complexity of features beyond any expectations prompted by previous fMRI work. We demonstrate that the information provided by gaze following neurons in this GFP fully predicts the monkey´s performance when asked to shift attention to the object singled out by the other’s gaze, strongly suggesting causality. We could also reliably show that many GFP neurons are able to integrate high-level rule signals with the spatial information provided by faces in order to make the gaze following behavior adaptive and controllable. The properties of these neurons establish the GFP as a key switch in controlling social interactions based on the other´s gaze.

Decoding of the other´s focus of attention by a temporal cortex module

H. Ramezanpour[1,2,3,*] and P. Thier[1,4,*].

[1] Department of Cognitive Neurology, Hertie Institute for Clinical Brain Research, University of Tübingen, 72076 Tübingen, Germany.

[2] Graduate School of Neural and Behavioural Sciences, University of Tübingen, 72074 Tübingen, Germany.

[3] International Max Planck Research School for Cognitive and Systems Neuroscience, University of Tübingen, 72074 Tübingen, Germany.

[4] Werner Reichardt Centre for Integrative Neuroscience, University of Tübingen, 72076 Tübingen, Germany.

*Correspondence to: Peter Thier and Hamidreza Ramezanpour, Department of Cognitive Neurology, Hertie Institute for Clinical Brain Research, Hoppe-Seyler-Str. 3, 72076 Tübingen, Germany. E-mail: thier@uni-tuebingen.de; hamidreza.ramezanpour@uni-tuebingen.de.

Abstract

Faces attract the observer´s attention towards objects and locations of interest for the other, thereby allowing the two agents to establish joint attention. Previous work has delineated a network of cortical “patches” in the macaque cortex, processing faces, eventually also extracting information on the other´s gaze direction. Yet, the neural mechanism that links information on gaze direction, guiding the observer´s attention to the relevant object, has remained elusive. Here we present electrophysiological evidence for the existence of a distinct “gaze-following patch (GFP)” with neurons that establish this linkage in a highly flexible manner. The other´s gaze and the object, singled out by the gaze, are linked only if this linkage is pertinent within the prevailing social context. The properties of these neurons establish the GFP as a key switch in controlling social interactions based on the other´s gaze.

Significance Statement

We follow the other´s gaze to objects of interest to the other and share attention to the object, a key step towards a theory of (the other´s) mind. Also monkeys follow gaze and establish joint attention. Although monkeys depend more on head gaze, i.e. the orientation of the other´s face, than humans, monkey gaze following exhibits many parallels to human gaze following, rendering monkeys the perfect model for studies of its neural underpinnings. Here we report the identification of a gaze following hub in the monkey STS characterized by neurons that link information on the other´s gaze with distinct targets. Importantly, this link is modifiable by contextual information, allowing the executive control of gaze following.

Introduction

We use the other´s gaze direction to shift attention to the object the other one is attending to, thereby establishing joint attention. Joint attention allows us to develop a theory of (the other´s) mind (ToM) (Baron-Cohen 1995) by mapping one´s own thoughts, beliefs and desires associated with the attended object onto the other one. Although it is questionable whether monkeys also possess a full ToM, they follow the other´s gaze to establish joint attention (Emery, Lorincz et al. 1997; Tomasello, Call et al. 1998;

Tomasello, Hare et al. 2007). An important distinction between human and non-human primate gaze following is the different weight of eye and head gaze cues. Not surprisingly in view of the fact that the eyes of non-human primates lack the conspicuous features of the human eye (Kobayashi and Kohshima 1997), monkeys´ gaze following relies primarily on head gaze rather than on eye gaze cues (Emery, Lorincz et al. 1997). This important difference notwithstanding, the evidence available emphasizes close similarities of human and non-human gaze following behavior, suggesting the possibility of a homologous system shared within the primate order. For instance, comparative fMRI work has delineated a distinct cortical node in the posterior temporal cortex of both rhesus monkeys and man specifically activated by gaze-following. In both species this gaze-following patch (GFP) is located in the immediate vicinity of the more posterior elements of the so-called face patch system (Marciniak, Atabaki et al. 2014; Marquardt, Ramezanpour et al. 2017), a system that has been implicated in the extraction of different aspects of information on faces such as identity, face or head orientation or facial expression (Tsao, Freiwald et al. 2003; Tsao, Freiwald et al. 2006; Freiwald, Tsao et al. 2009; Freiwald and Tsao 2010). Actually, not only face orientation is important for the guidance of the observer´s gaze but also information on identity or facial expressions as both are known to modulate human gaze-following (Shepherd 2010). Hence, it is likely that the GFP may draw on information from the face patch system. Yet, the neural mechanisms that may allow the GFP to use the information on the other´s facial features into a gaze-following response establishing joint attention are not known. It is also unclear if neurons in the GFP may possibly contribute to the cognitive control of gaze-following, integrating contextual information relevant for the modulation of the behavior. After all, although human and monkey gaze-following has features of a quasi reflex-like behavior that kicks in at short latency, it can be suppressed to a considerable degree if not appropriate within a given context (Ricciardelli, Carcagno et al. 2013; Marciniak, Dicke et al. 2015).

With these questions in mind, we explored the GFP of rhesus monkeys and adjoining regions of the superior temporal cortex (STS), deploying tasks that asked the observer to follow the other´s gaze or to suppress gaze-following if not expedient. Our results suggest that neurons in the GFP link information on the other´s gaze and the object singled out by the gaze, provided that this linkage is pertinent within the prevailing social context.

Results

We recorded the activity of well-isolated single neurons in the GFP and adjacent regions of the right STS of two rhesus monkeys. In one of the two, the location of the GFP had been delineated in preceding fMRI experiments in which we had searched for BOLD activity associated with gaze-following as described by (Marciniak, Atabaki et al. 2014). In the second monkey, we relied on the same coordinates as a reference, when exploring the STS. Both monkeys had learned to follow the direction of a monkey head ("demonstrator") presented on a monitor. The demonstrator turned to one out of four spatial targets ((head) gaze-following task). Alternatively, the monkeys had to use the facial identity of the portrayed monkeys to determine the relevant target. To this end, they had to rely on a learned association between the four targets and the four possible identities (identity-mapping task). An instructive color cue presented on a baseline portrait before the appearance of the four spatial cues and targets told the monkey whether to deploy the gaze or the identity rule when dealing with the monkey portraits. The two trial types were presented randomly interleaved (Figure 1A, B).

The design of the paradigm allowed us to dissociate neural activity evoked by features of the portraits from activity associated with the shift of attention to a particular target object, prompted by two different social cues, gaze direction and facial identity respectively. As the portraits and the overt behavior they caused were the same, independent of the rule, any difference in neural responses had to be a reflection of differences between the cognitive processes responsible for the rule-based selection of target objects. Monkey L reached a mean performance of 84 ± 5% correct trials on the gaze-following task and 82 ± 7% on the identity-mapping task, whereas monkey T attained 71 ± 8% and 73± 11% respectively (mean ± std) (Figure 1C). Both monkeys´ performance was significantly above the 25% chance level ($P < 0.001$, binomial test) and independent of the specific task (Wilcoxon signed-rank test, $p > 0.05$). We also tested the responses of the same neurons to the passive viewing of faces and a variety of biological and non-biological objects (Figure 1D).

Single pSTS neurons encode gazed-at targets

Altogether, we tested 923 neurons recorded from the posterior STS (pSTS) of the two monkeys on all three tasks. Out of these, 426 neurons (172 neurons in monkey T and 254 neurons in monkey L) exhibited significant changes of their discharge rate relative to baseline (Kruskall-Wallis ANOVA, $p<0.05$) in at least one of two key phases of a trial, namely during the presentation of the rule and/or during the subsequent availability of the spatial information provided by gaze direction or facial identity. In total, 109 (out of 426 task-related) neurons exhibited selectivity for head gaze ("gaze-following (GF) neurons"), 37 neurons for facial identity specifying spatial locations ("identity-mapping (IM) neurons") and 12 neurons were responsive to both gaze and identity ("mixed selectivity neurons")(see pie chart in Figure 2A).

Figure 2B depicts the distribution of spatial preferences of GF and IM neurons based on the target yielding the maximal response. It shows that all four possible targets are well represented in the data set without any bias for the left or the right side. The discharge profiles of two exemplary spatially-selective neurons and one exemplary classical face-selective neuron lacking interest in spatial information are assembled in Figures 2C-E.

Figure 2C shows a typical GF neuron. Its discharge profile was characterized by very similar discharge rates in the gaze following and the identity mapping tasks until the time the monkey portrait provided information on the target location to be chosen. In case the rule demanded gaze following, the discharge rate was significantly higher than for identity-mapping if the cued target was the one at 10° on the right (target G4, corresponding to 40° left from the view of the demonstrator monkey). The difference became significant shortly after the onset of the spatial cue, reached its maximum 145ms later (first peak) and stayed until the time of the indicative saccade. Figure 2D depicts another neuron exhibiting a qualitatively similar discharge pattern, yet with some preference for identity-mapping defining the target at 10° on the left (target ID1). Both neurons lacked specificity for faces when tested for visual responses to the presentation of faces and a variety of biological and non-biological objects during stationary fixation ("object vision task"). On the other hand, the neuron shown in Figure 2E was a classical face-selective neuron when tested in the object vision task, characterized by a strong preference for face stimuli. Clear bursts of activity evoked by the appearance of the portraits also characterized the active tasks without any difference between the two conditions or between the spatial targets within each task.

A clear preference for distinct targets was also exhibited by the population tuning curve based on all 109 spatially-selective GF neurons. To assess the selectivity of the population for distinct spatial targets we ranked the strength of the responses of all individual GF neurons to the four gaze targets and calculated population responses for each rank. Rank 1 stood for the most preferred gaze target (highest mean discharge in the period of 50 ms after the onset of the spatial cues until the appearance of the *go*- cue) and rank 4 for the least preferred gaze target. As can be seen in Figure 3A, the rank-specific population responses were very distinct with the largest burst of activity for rank 1, a smaller one for rank 2 and clear activity suppression for the two lowest ranked targets.

Figure 3B compares the population responses of the GF neurons for the highest and lowest ranked targets in the GF task with the responses evoked by the same targets cued by facial identity. In case of identity-mapping, the difference between the population responses for the two targets associated with the most and the least preferred target in the gaze-following task was dramatically reduced to a non-significant level (Mann-Whitney U -test, p=0.32). Both discharge profiles, in each case averaging over all four identities, lay in between the rank 1 and the rank 4 responses evoked by gaze cueing. The residual response modulation in the spatial cueing period, uninfluenced by target position, may reflect the need to process facial identity in this task. An analogous analysis for the IM neurons did not reveal any significant difference between the population responses to the rank 1 and the rank 4 targets (Mann-Whitney U test, p=0.1)(Figure S1B). In other words, at the population level IM neurons do not convey information on spatial targets. This may suggest that the significant preferences for particular identity-targets association, exhibited by 37 out of the 76 IM neurons may actually reflect identity tuning rather than spatial tuning.

Under the assumption that the GF neurons underlie a monkey´s ability to follow gaze to the relevant target, error trials, in which the monkey fails to hit the target identified by the other´s gaze should be associated with reduced selectivity of the GF population discharge. To test this prediction, we calculated a spatial selectivity index (SSI), capturing the difference between the population responses to the most (rank 1) and the least preferred target (rank 4) divided by the sum. For the population of GF neurons, the distribution of SSI varied between 0 and 1 with a median of 0.48 for correct trials. For incorrect trials the whole distribution shifted to the left with a median of 0.33, significantly smaller than the one for correct trials ($p < 0.001$, Mann-Whitney U test; $n = 109$). Unlike the distribution for correct trials, the one for error trials spread into the negative range, indicating that quite a few neurons changed their spatial preferences (Figure 3C). The notion that errors in gaze-following trials are a consequence of compromised selectivity of the GF neuron population signal is also supported by a time-resolved decoding analysis based on a linear support vector machine (SVM) classifier with 5-fold cross-validation, determining the amount of information available on the correct target. We obtained a decoding time course by performing this analysis in a 100ms window and advancing in steps of 20ms during the whole spatial cueing period. We performed this analysis separately for the two pools of GF neurons, tested with spatial cue windows of 200ms (n=28) and 500ms (n=81) respectively. As shown in Figure 3D, information about the position of the spatial targets is present almost throughout the whole time. These results demonstrate that the population of GF neurons offers reliable information on the gazed-at target throughout a period from the onset of the spatial cue until the time of the *go*-signal. For error trials, the maximum of the decoder classification performance dropped by about 10%, in line with the notion that precise shifts of attention to the gazed at target require a specific population signal.

To test if task-related neurons in the pSTS are indeed tuned only to social cues such as information on gaze direction or facial identity, we ran a control task with abstract symbols replacing the faces. Four specific symbols —a square, a circle, a triangle, and a star— had been learned to be associated with one out of the four possible targets each. Out of the pool of 426 GF or IM task-related neurons 131 neurons were tested on this task. Only very few (n=8) showed weak, albeit significant responses to targets cued by symbols. Moreover, the population response failed to distinguish correct and error trials. The same holds for a topographically distinct separate population of face-selective neurons (n=23). These neurons (with the exception of n=6) lacked spatial tuning in the gaze-following and identity-mapping tasks and as a group failed to discriminate between the correct and error trials (see Supplementary materials, Figure S1A).

pSTS neurons encode abstract rules and bias monkeys´ social choices

Spatial selectivity was not the only feature characterizing neurons in the pSTS. We could also identify 104 rule-selective neurons, either encoding the rule to follow gaze or to map identity. The population of rule-selective neurons overlapped with the one exhibiting spatial selectivity with 22% of the latter (43 out of 195) showing both spatial and rule selectivity. Figure 4A depicts two exemplary rule-selective neurons, one preferring the gaze-following rule, the other one the identity-mapping rule. Both exhibited a clear increase of their discharge rates for the respective preferred rule, 131ms and 159ms (latency of first peak) respectively after the onset of the information on the prevailing rule. As exemplified by these two neurons, the rule-associated activation depended on whether the monkey was able to convert the rule into successful shifts of attention to the correct target or not. In case of error trials, the differential response of the GF neuron dropped dramatically, while, conversely, the IM neuron showed a higher amount of differentiation between the two rules. Note that both neurons lacked a significant response to the portraits comprising the non-preferred rule, i.e. similar to many spatially-selective neurons they were not sensitive to the vision of faces.

The pie chart in Figure 4B gives a breakdown of the numbers of rule-selective neurons in each category and each monkey. Figure 4C depicts the population responses of rule-selective neurons preferring the GF and the IM rule respectively. In both cases, the population plots exhibit excitatory responses to the preferred rule. However, whereas the population responses for neurons preferring the gaze-following rule show a weak excitatory response to the non-preferred rule, the one for those preferring the identity-mapping rule is characterized by a discharge suppression in the presence of the non-preferred rule. Unlike identity-mapping, gaze-following has all the features of a domain-specific cognitive function, among others characterized by a significant degree of automaticity. In other words, reliably suppressing the gaze-following reflex may need extra efforts involving the active suppression of the distracting gaze-following rule-related activity.

As alluded to above quite a few rule-selective neurons integrated rule selectivity and sensitivity to spatial targets, either identified by gaze or by identity, in any case in a congruent manner. That is to say that neurons that preferred the gaze rule also preferred gaze-following and vice versa neurons selective for the identity-mapping rule attentional shifts guided by identity. An example of a neuron selective for the gaze rule and for a target selected by gaze is depicted in Figure 4D. We wondered if the degree of rule selectivity might predict the later spatial selectivity. This was indeed the case as shown by a quantitative analysis of the rule selectivity based on a rule selectivity index (RSI) calculated by the normalized difference of the mean discharges for the GF and IM conditions in the 400ms of the rule window (details in the supplementary materials). In other words, the ability to decode the rule is relevant for the ability to shift attention to the right target. This is also indicated by a consideration of error trials. In the case of GF rule-selective neurons, median RSI values dropped from a median of 0.16 for correct trials to 0.12 for error trials ($p < 0.05$, Mann Whitney U test; $n = 104$)(Fig. 4E). Similarly, for the population of IM rule-selective neurons the median RSI values decreased from -0.21 for correct trials to -0.15 for error trials ($p < 0.001$, Mann Whitney U test; $n = 104$). As a consequence, although still significantly different ($p < 0.001$, Mann Whitney U test, two-tailed; $n = 104$) the distribution of RSI values for the two groups of neurons exhibited considerable overlap for error trials, an impairment that is in principle in accordance with the drop in behavioral selectivity (Figure 4E). The same conclusion can be drawn from a decoding analysis deploying a support vector machine classifier. Here we asked how well the responses evoked by the two rules predicted the behavioral decisions. As shown in Figure 4F, the classifier performance dropped significantly for erroneous decisions.

Topography of the neural response types in the STS

We reconstructed the locations of recorded neurons in stereotactic coordinates based on a 3D rendering of the pSTS using anatomical MRI data sets available for two monkeys. The positions of neurons were then used to construct 2D-density maps of response features following unfolding of the pSTS. To this end, we counted the number of neurons in each 0.5 mm^2 of unfolded cortex and finally passed the resulting distribution through a 2D-spatial filter (Gaussian, $\sigma=2$). Figure 5 depicts the resulting density maps for the two monkeys. As can be seen, GF and IM neurons had similar locations in the pSTS with the highest density around stereotactic coordinates A0-A2 in monkey T and P1-A1 in monkey L (A0 represents the interaural plane). This hot spot is located on the ventral bank of the pSTS and encroaches on the fundus and the dorsal part of the posterior inferotemporal cortex (pITd). To compare the location of the neurons found in this study with the topography of the GF patch as delineated by a significant contrast between BOLD signals evoked by GF and IM respectively (Marciniak, Atabaki et al. 2014), we calculated an analogous contrast map based on the electrophysiological heat maps for GF and IM. Despite the general overlap of GF and IM neurons, the contrast map exhibited a clear dominance of GF-related activity, due to the larger number of GF neurons. The location of this GF-IM hot spot is comparable to the location of the GFP obtained by BOLD imaging.

Discussion

The posterior STS exhibits a clear functional topography with neurons presenting gaze-following related activity confined to a relatively small area in the lower bank and fundus of the STS around A0-A2 in one monkey and A1-P1 in the other monkey, the gaze-following patch (GFP), clearly separated from neighboring face-selective cortex. The location of the GFP as determined by the properties of single neurons is in good accordance with the location of the GFP as identified by fMRI (Marciniak, Atabaki et al. 2014). Although boundaries of areas delineated by fMRI are based on somewhat arbitrary statistical thresholds they are often perceived as being sharp and, moreover, associated with qualitatively different functions on both sides. However, our electrophysiological exploration clearly showed that the boundaries of the electrophysiologically defined GFP are gradual with quite a few GF neurons located many millimeters away from the gaze-following hotspot.

Shifts of endogenous attention have recently been shown to elicit BOLD activity in the dorsal posterior inferotemporal cortex (pITd) of monkeys (Stemmann and Freiwald 2016). Considering the published coordinates, this area might be close to the GFP or even overlap with it. Hence, could it be that the GFP is a generic node in an attention network rather than playing a distinct role in gaze-following and joint attention? We think that this possibility can be rejected given the fact that just a few (n=8) of the GF neurons tested responded in a control task in which spatial targets had to be identified by a learned association with distinct abstract objects. On the other hand, some of the spatially-selective GF neurons also exhibited an at least weak interest in shifts of spatial attention guided by learned associations between facial identities and target locations and in general, the regions in which GF- and IM-related activity were found, overlapped, although the former clearly dominated. In any case, it is the usage of facial information for the purpose of focusing spatial attention that characterizes the GFP. Faces without inherent or learned spatial value do not drive neurons in the GFP. And conversely, the face-selective neurons outside the GFP do not respond to shifting spatial attention, no matter whether the faces provide directional information or not. Of course, this does not preclude the possibility, actually suggested by anatomical proximity, that the GFP depends on input from face-selective neurons. The availability of information on spatial targets derived from gaze and in general faces does not necessarily imply a key role in controlling the behavior. Indeed a causal role in guiding behavior is suggested by the fact that the discriminatory power of the population signal on the correct target predicted the behavioral choice.

Further support for a causal role comes from a previous study which demonstrated that reversible inactivation of the pSTS compromised the ability of monkeys to use gaze cues to guide target choices (Roy, Shepherd et al. 2014). While the selection of injection sites was based on the response of neurons to a passive face-viewing task and ignorant of physiological landmarks reflecting the preferences of neurons for active GF responses, the reported coordinates suggest that the relatively large injections might have involved the GFP.

By stimulating parietal area LIP, Crapse and Tsao have recently been able to evoke BOLD activity in a region of the monkey pSTS whose coordinates correspond to the GFP (Crapse and Tsao 2013). Hence, it is tempting to speculate that LIP may draw on information on spatial choices prompted by the other´s gaze, originating from the GFP. This input might allow LIP to update its spatial saliency map and to reallocate spatial attention. Such a transfer of information would explain the fact that neurons in LIP present activity related to spatial shifts of attention evoked by gaze cues (Shepherd, Klein et al. 2009).

The GFP does not seem to be confined to ignite gaze-following but also to help suppressing it if not pertinent. This is suggested by the fact that many GFP neurons are sensitive to the rule specifying if the gaze should be followed or not. The observer´s ability to implement a prevailing rule such as to inhibit gaze-following is predicted by the discriminatory power of the population-based rule-related activity in a given moment. This suggests a key role of the GFP in controlling gaze-following. The prefrontal cortex is thought to be important for the encoding of rules (Wallis, Anderson et al. 2001). This is for instance indicated by the difficulties of patients with prefrontal lobe damage in following rules (Szczepanski and Knight 2014). Hence, it may well be assumed that the rule sensitivity of neurons in the GFP might be a consequence of the integration of top-down information from prefrontal cortex. Although we know that output from Brodmann areas 8 and 46 of prefrontal cortex reaches the posterior parts of the STS (Kawamura and Naito 1984; Yeterian, Pandya et al. 2012), it remains open if the GFP is among the target structures.

Like human gaze-following, also monkey gaze-following seems to be a domain-specific faculty that does not have to be learned from scratch, resorting to domain-general machinery. Arguably, the latter is needed to learn to associate particular spatial targets with facial identities or abstract objects as required in our study. And we would interpret the existence of identity-mapping-related signals in the GFP as reflections of the learned association. Yet, how sure can we be that the gaze-following-related activity in the GFP is not also a signature of a learned association? We think that the following arguments render this possibility unlikely. 1. The notion that monkey head gaze-following is domain specific has received substantial support from a previous behavioral study which delineated the position of a monkey´s focus of attention guided by gaze or by identity cues (Marciniak, Dicke et al. 2015). Shifts of spatial attention could be fully suppressed if prompted by identity cues. However, shifts of attention guided by gaze cues were blocked only largely, yet not entirely, even after extensive periods of training (Marciniak, Dicke et al. 2015). The inability to unlearn gaze-following completely suggests an inborn behavioral capacity not modifiable by learning. 2. A patient suffering from right hemisphere temporal lesion no longer benefitted from gaze direction cues when detecting peripheral targets while the ability to use arrow cues remained intact (Akiyama, Kato et al. 2006a, 2006b). 3. The mark of gaze-following related activity in the GFP is considerably stronger than the one of identity-mapping with the number of GF neurons around 4-fold more (Fig.2a). Hence, the evidence available is in line with the interpretation that the GFP is a central, and possibly domain-specific node in a network for the ignition and the control of gaze-following. The emergence of identity-mapping related activity in the GFP is most probably a consequence of the need to control gaze-following based on identity information.

Materials and Methods

Animals, surgery, and recording methods

All experimental preparation and procedures were approved by the local animal care committee (Regierungspräsidium Tübingen, Abteilung Tierschutz) and fully complied with German law and the National Institutes of Health's Guide for the Care and Use of Laboratory Animals. Two male rhesus (Macaca mulatta) monkeys (T and L) of weights 8 kg and 11 kg respectively were used in this study. Before the recording chamber was implanted, we acquired structural Magnetic Resonance Imaging (MRI) scans to identify implant locations. Scans were carried out in a Siemens 3T scanner. Then monkeys were implanted with a titanium head-post to restrain the head during the experiment, scleral search coils for eye position recording and a cylindrical titanium chamber for the introduction of microelectrodes.

Monkey L had participated in a previous fMRI study that had led to the identification of the GFP (1). This allowed us to use the stereotactic data available to determine the position and orientation of the chamber on the skull in order to approach the GFP. For the placement of the chamber in monkey T, we relied on the average location of the GFP in the two monkeys that had participated in the fMRI study. All surgeries were carried out under combination anesthesia with isoflurane and remifentanil (1–2 μg/kg/min) with monitoring of all physiological parameters (heart rate, blood oxygen saturation, blood pressure, body temperature). After surgery, opioid analgesics (buprenorphine) were administered until no sign of pain was evident anymore. The experiments commenced only after full recovery about 12 days after surgery.

Single-unit recording

We recorded single unit activity with vertically movable glass insulated microelectrodes (Alpha Omega, 0.5–1 MΩ at 1 KHz) using conventional techniques. In brief, microelectrodes were driven by a homemade multi-channel micromanipulator attached to the recording chamber in every recording session. Up to four microelectrodes were inserted at the same time with at least 1mm distance to each other. The micromanipulator allowed the selection of microelectrodes positions relative to the chamber walls plane with a spatial resolution of 0.5mm. Single units were isolated online using the spike waveform matching option of the Alpha Omega SnR system. Quality of isolation was again checked offline and only units whose spikes had been stable throughout the whole session were considered for further analysis.

Behavioral tasks

Two monkeys were trained on two “active” tasks requiring either following the head gaze of a demonstrator monkey portrayed on a monitor towards distinct spatial targets or, alternatively, the identification of the same targets based on learned associations with the identity of the portrayed demonstrators. Moreover, they were tested on a “passive” task, requiring fixation of a central target, while a series of behaviorally irrelevant face and non-face images, centred on the target, were presented. In the active tasks, trials started with a white fixation point on a dark background. After 500ms, a neutral monkey face, centered on the fixation point, always looking straight ahead, appeared. 400ms later, the central fixation changed its color to either red or green, informing the monkey on the rule for target selection to be applied to the upcoming view of an oriented monkey face (“demonstrator”). In case of red, the observer was required to follow the demonstrator’s gaze to one of four spatial targets. The green color required the monkey to make a saccade to the target chosen based on a learned association between the four target positions and four possible facial identities, while ignoring gaze orientation. Note that we replaced the green color used to indicate the identity mapping rule by blue in all figures for the sake of better visibility. The demonstrator appeared immediately after the disappearance of the straight face and remained on until the end of the trial. The targets became available together with the onset of the demonstrator. The elimination of the central fixation point 200ms or 500ms after the appearance of the demonstrator served as *go*-signal for the observer. The monkeys received a drop of water as a reward if they kept fixation of the central fixation point and later made a successful saccade to the target as demanded by the prevailing rule. Trials were aborted if monkeys were not keeping their eyes within a window of 2º around the fixation point and the target respectively and were unable to reach the target within 300ms after the *go*-signal.

The images of straight and oriented faces had a size of 5.6º x 5.6º and were presented in the center of a monitor placed at a distance of 60 cm from the observer. Spatial targets were small red dots (diameter of 0.8º) and were aligned on a virtual horizontal line 1o below the center of the portraits at horizontal eccentricities of -10º, -5º, 5º and 10º with respect to the observer monkey (-40º, -20º, 20º and 40º with respect to the demonstrator monkeys). As the portrait of each individual monkey could be shown in four different head gaze orientations, corresponding to the four spatial targets, the stimulus set involved 16 stimuli (Figure 1B). We used an open source recording and stimulation system for recording of eye movement data and presenting the stimulus images (nrec.neurologie.uni-tuebingen.de/nrec). Gaze-following and identity mapping trials were identical in visual terms except for the color of the instruction cue, available in a short period only, and identical with respect to the motor responses required. Hence, any differences in the associated neuronal responses outside the short presence of the instruction cue had to be a consequence of differences in cognitive strategies and operations.

Finally, the monkeys had to perform the aforementioned passive viewing task in which images of faces and non-face stimuli, centered on the fixation dot, were presented and the monkeys had to keep fixation of the central fixation point (Figure 1C). In this task, we used the same set of images used in a previous study (1) in addition to the 16 monkey portraits used in the active tasks and additional 16 human faces (four identities with four gaze directions similar to those monkeys' head directions in the active tasks taken from the Radboud Face Database (2)). Monkeys saw in total 144 images of 6^ox6^o, each lasting for 400ms and followed by a 400ms black and white random dot background (pixel size 0.05^o). Monkeys were rewarded in this experiment if they were keeping their eyes within a window of 2^ox2^o around the central fixation point for each image.

Statistical Analysis

In order to characterize the discharge patterns evoked in the two active tasks, we determined the mean discharge rate in three periods: (1) the baseline period: the last 100ms of the portrait fixation period right before the onset of the rule presentation period, (2) the rule period: the 400ms after onset of the instructive cue, and (3) the spatial information period: the period during which the demonstrator was available and the observer waited for the *go*-signal cue. The duration of this latter period was either 200ms or 500ms. We refrained from considering later periods because we expected them to be influenced by a complicated mixture of variables like saccade execution, saccade-induced visual stimulation, reward expectancy and preparation or outcome evaluation. We determined the task-related preferences of neurons by comparing the mean firing rates in the three periods by a non-parametric 1-way ANOVA (Kruskal-Wallis test) (considering $p<0.05$). When an effect in the 1-way ANOVA was found, the specific phase (rule or spatial cue periods) significantly different from baseline was identified by means of a post-hoc analysis ($p < 0.05$ Bonferroni corrected for multiple comparisons). Neurons which exhibited a significant change of their discharge rate in the rule and/or the spatial cueing period were selected for further analysis. A neuron was considered to be spatially selective in the spatial cueing period if its firing rates (only correct trials considered) to the four different targets were significantly different (Kruskal–Wallis one-factor ANOVA, $P < 0.05$, carried out separately for gaze-following and identity-mapping). The population responses associated with the most and the least preferred target were compared by a Mann-Whitney U test ($p<0.05$).

A rule selectivity index (RSI), with a theoretical maximum of 1 for gaze following, -1 for identity mapping and a theoretical minimum of 0 for unselective neurons, was calculated for the 400ms rule according to

$$RSI = \frac{< R_{gaze} > - < R_{Identity} >}{< R_{gaze} > + < R_{Identity} >}$$

In a similar way, a spatial selectivity index (SSI), with a theoretical maximum of 1 and minimum of 0, was calculated for the 200/500ms spatial cueing periods according to

$$SSI = \frac{< R_{most\ preferred(rank\ 1)} > - < R_{least\ preferred(rank\ 4)} >}{< R_{most\ preferred(rank\ 1)} > + < R_{least\ preferred(rank\ 4)} >}$$

with the operator < > denoting the mean firing rate in correct trials. Changes in the distribution of the RSI and SSI values were evaluated by means of Mann-Whitney U tests.

In the decoding analysis, we deployed a support vector machine (SVM) with 5-fold cross-validation to determine how well the population discharge predicted the spatial target selected and the choice of the monkeys, considering the prevailing rule. We obtained a decoding time course by performing this analysis in a100ms window, advanced in steps of 20ms throughout the whole spatial cueing period. We performed this analysis separately for the two different spatial cueing periods of 200ms and 500ms when testing for target location sensitivity. Error trials were classified according to the class of correct trials that they resembled most. Standard errors of the decoding performance were obtained by bootstrapping (n=1000).

***Acknowledgments*:** We are grateful to Friedemann Bunjes and Peter W. Dicke for valuable technical support and Ian Chong for critical reading of the manuscript.

***Funding*:** This work was supported by a grant from the DFG (TH 425/12-2).

Author contributions**:** P. T. developed the conceptual framework of the research. P. T. and **H. R.** designed the experiments, interpreted the results and wrote the paper. **H. R.** performed the experiments and analyzed the data.

Competing interests**:** The authors declare no competing financial interests.

Data and materials availability**:** Data is available on reasonable request from authors.

References

Akiyama T, Kato M, Muramatsu T, Saito F, Nakachi R, Kashima H (2006) A deficit in discriminating gaze direction in a case with right superior temporal gyrus lesion. Neuropsychologia 44(2): 161-170.

Akiyama T, Kato M, Muramatsu T, Saito F, Umeda S, Kashima H (2006) Gaze but not arrows: a dissociative impairment after right superior temporal gyrus damage. Neuropsychologia 44(10): 1804-1810.

Baron-Cohen S (1995) Mindblindness: an essay on autism and theory of mind. MIT Press.

Crapse TBC, Tsao D (2013) A strong input to area LIP from two distinct regions in IT cortex revealed by combined fMRI and microstimulation." Program No. 824.07. Neuroscience 2013 Abstracts. San Diego, CA: Society for Neuroscience.

Emery NJ, Lorincz EN, Perrett DI, Oram MW, Baker CI (1997) Gaze following and joint attention in rhesus monkeys (Macaca mulatta). J Comp Psychol 111(3): 286-293.

Freiwald WA, Tsao DY (2010). Functional compartmentalization and viewpoint generalization within the macaque face-processing system. Science 330(6005): 845-851.

Freiwald WA, Tsao DY, Livingstone MS (2009) A face feature space in the macaque temporal lobe. Nat Neurosci 12(9): 1187-1196.

Kawamura K, Naito J (1984) Corticocortical projections to the prefrontal cortex in the rhesus monkey investigated with horseradish peroxidase techniques. Neurosci Res 1(2): 89-103.

Kobayashi H, Kohshima S (1997) Unique morphology of the human eye. Nature 387(6635): 767-768.

Langner O, Dotsch R, Bijlstra G, Wigboldus DHJ, Hawk ST, Van Knippenberg A (2010) Presentation and validation of the Radboud faces database. Cognition & Emotion 24(8): 1377—1388.

Marciniak K, Atabaki A, Dicke PW, Their P (2014) Disparate substrates for head gaze following and face perception in the monkey superior temporal sulcus. ELife 3.

Marciniak K, Dicke PW, Their P (2015) Monkeys head-gaze following is fast, precise and not fully suppressible. Proc Biol Sci 282(1816).

Marquardt K, Ramezanpour H, Dicke PW, Their P (2017) following eye gaze activates a patch in the posterior temporal cortex that is not part of the human "face patch" system." eNeuro 4(2).

Ricciardelli P, Carcagno S, Vallar G, Bricolo E (2013) Is gaze following purely reflexive or goal-directed instead? Revisiting the automaticity of orienting attention by gaze cues. Exp Brain Res 224(1): 93-106.

Roy A, Shepherd SV, Platt ML (2014) Reversible inactivation of pSTS suppresses social gaze following in the macaque (Macaca mulatta). Social cognitive and affective neuroscience 9(2): 209-217.

Shepherd SV (2010) Following Gaze: Gaze-Following Behavior as a Window into Social Cognition. Front Integr Neurosci 4.

Shepherd SV, Klein JT, Deaner RO, Platt ML (2009) Mirroring of attention by neurons in macaque parietal cortex. Proc Natl Acad Sci U S A 106(23): 9489-9494.

Stemmann H, Freiwald WA (2016) Attentive motion discrimination recruits an area in inferotemporal cortex. J Neurosci 36(47): 11918-11928.

Szczepanski SM, Knight RT (2014) Insights into human behavior from lesions to the prefrontal cortex. Neuron 83(5): 1002-1018.

Tomasello M, Call J, Hare B (1998) Five primate species follow the visual gaze of conspecifics. Anim Behav 55(4): 1063-1069.

Tomasello M, Hare B, Lehmann H, Call J (2007) Reliance on head versus eyes in the gaze following of great apes and human infants: the cooperative eye hypothesis. J Hum Evol 52(3): 314-320.

Tsao DY, Freiwald WA, Knutsen TA, Mandeville JB, Tootell RB (2003) Faces and objects in macaque cerebral cortex. Nat Neurosci 6(9): 989-995.

Tsao DY, Freiwald WA, Tootell RB, Livingstone MS (2006) A cortical region consisting entirely of face-selective cells. Science 311(5761): 670-674.

Wallis JD, Anderson KC, Miller EK (2001). Single neurons in prefrontal cortex encode abstract rules. Nature 411(6840): 953-956.

Yeterian EH, Pandya DN, Tomaiuolo F, Petrides M (2012) The cortical connectivity of the prefrontal cortex in the monkey brain. Cortex 48(1): 58-81.

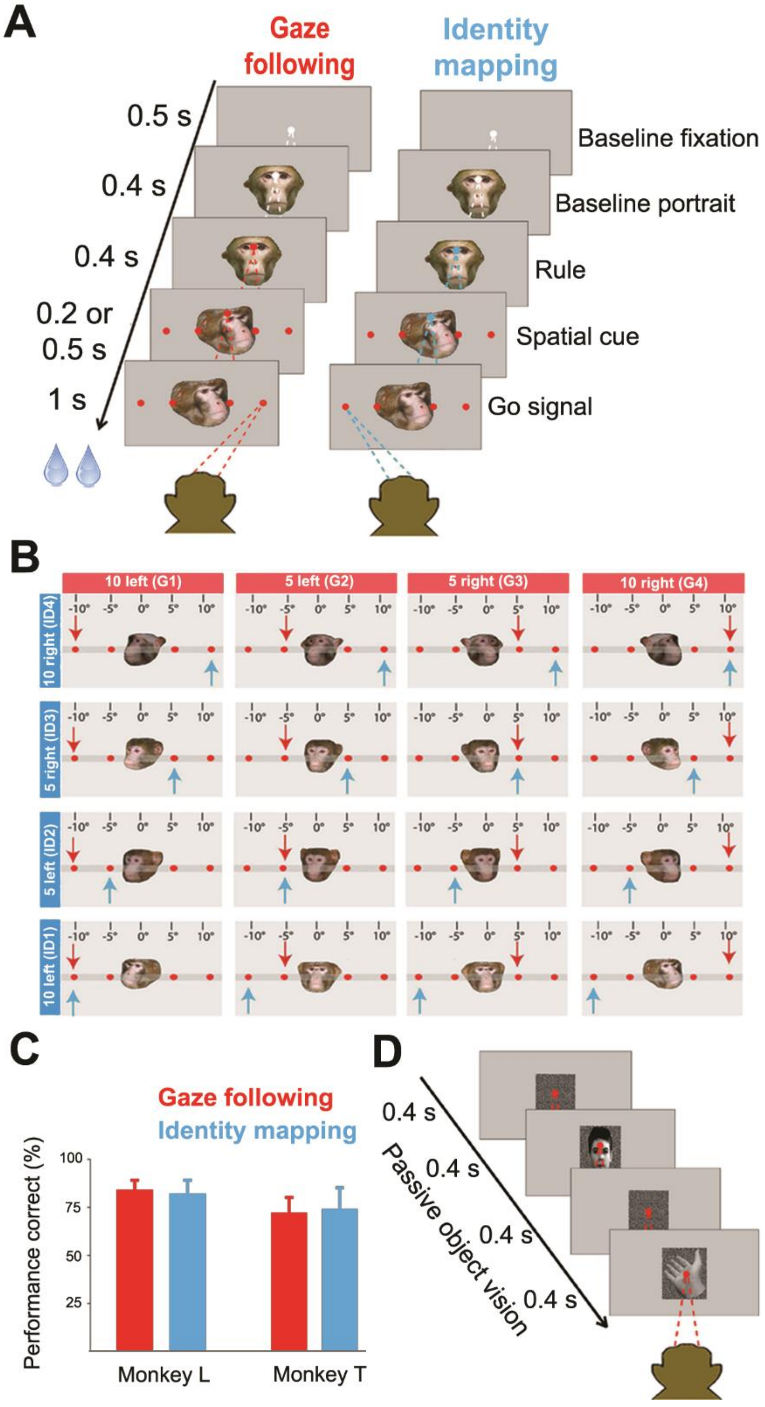

Figure 1. Behavioral Paradigms

(A) Each trial started with the presentation of a white central fixation dot, 500ms later supplemented by a straight ahead looking portrait. 400ms after the onset of the portrait, the fixation dot changed color, specifying the rule to be applied on the trial. Red indicated gaze-following, green identity-mapping (note:

rendered blue in all figures for better visibility). 400ms later, the straight ahead portrait was replaced by the portrait of another monkey (“demonstrator”) looking at one out of the four targets, looming up at the same time. The disappearance of the central fixation point, 200ms or 500ms later —depending on the day— served as the *go*-signal to make a saccade to the target specified by the demonstrator. In the gaze-following task the relevant cue was the demonstrator´s head gaze whereas in the identity-mapping task, the observer was asked to ignore the gaze direction and to make a saccade and respond to the target specified by the portrait´s identity, resorting to learned associations between the four targets and the four possible identities of the demonstrators. These two tasks were randomly interleaved. (B) The 4x4 cue matrix defined by the four possible demonstrator identities and the four possible orientations of the demonstrator´s head (40° left, 20° left, 10° right, 20° right from the demonstrator´s viewpoint corresponding to targets at 10° left, 5° left, 5° right, 10° right from the perspective of the observer). The blue arrow in each cell specifies the target to be chosen according to the prevailing identity, the red arrow the one singled out by head gaze. (C) The behavioral performance of the two monkeys in each of the two behavioral paradigms was very good, well above the 25% chance level, not significantly different for the two tasks (p>0.05) and without significant difference between the two monkeys. Error bars represent standard error. (D) The passive viewing task required the observer to fixate a 0.2° dot while exposed to a sequence of images of faces and non-face stimuli, presented randomly interleaved. Each image was on for 400ms and followed by a 400ms duration random dot background.

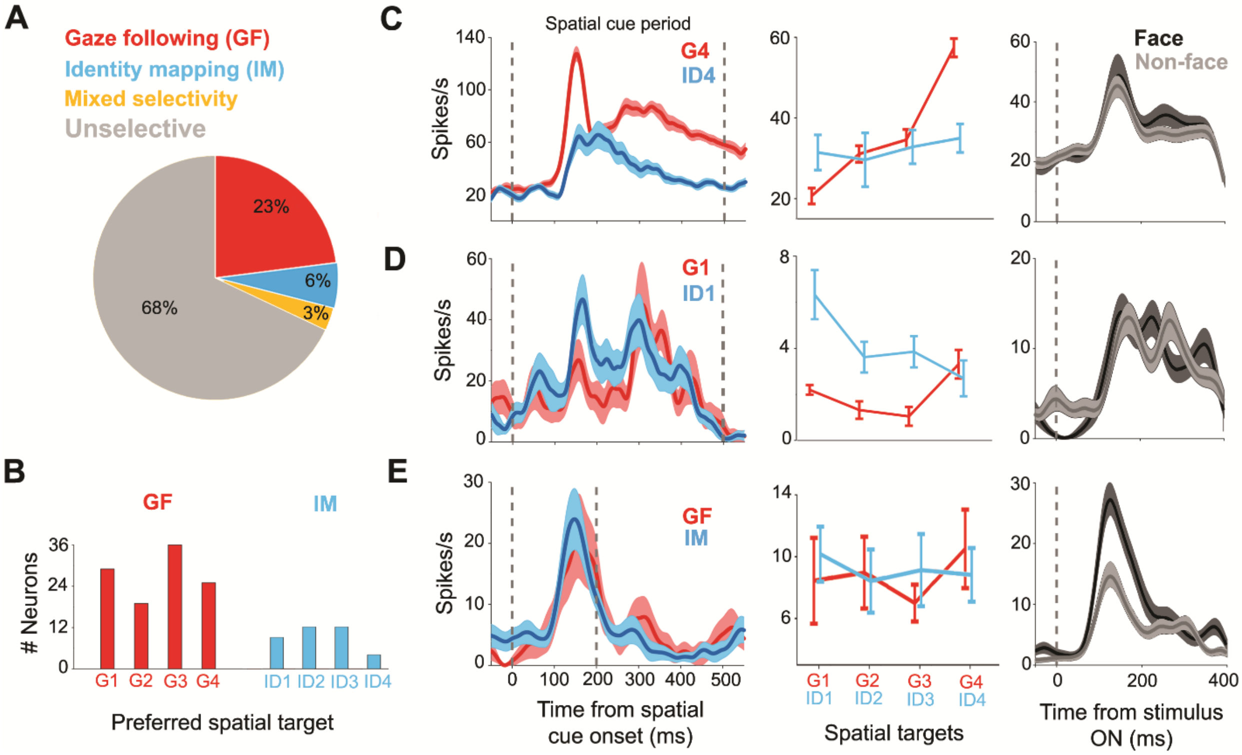

Figure 2. The variety of neuronal response features in the pSTS

(A) Breakdown of response preferences of neurons in the GFP during the spatial cueing period. (B) Distribution of targets preferred by spatially-selective GF and IM cells respectively, (G1=10° left , G2=5° left, G3= 5° right, G4=10° right, ID1=10° left , ID2=5° left, ID3= 5° right, ID4=10° right) according to their most preferred targets pooled over both monkeys. (C) Response profiles of an exemplary spatially-selective GF neuron tested in the GF and the IM task (left and middle panel) and the passive viewing task (right panel). This neuron was activated in both the GF and the IM task, yet clearly more in the former with preference for target G4. It did not show face selectivity in the passive viewing task. (D) Exemplary IM neuron with preference for target ID1 in the identity-mapping task. Also this neuron failed to exhibit face selectivity in the passive viewing task. (E) Example of a classical face-selective neuron that preferred faces over non-face stimuli in the passive viewing task and a clear face response in both the gaze-following and the identity-mapping task without exhibiting any sensitivity to the other aspects of the two tasks. In all panels the vertical line at t=0 identifies the onset of the four targets while the second vertical line at 500ms (C, D) or 200ms (E) identifies the time of the *go*-signal. Error bars and shaded areas represent standard error in all figure parts.

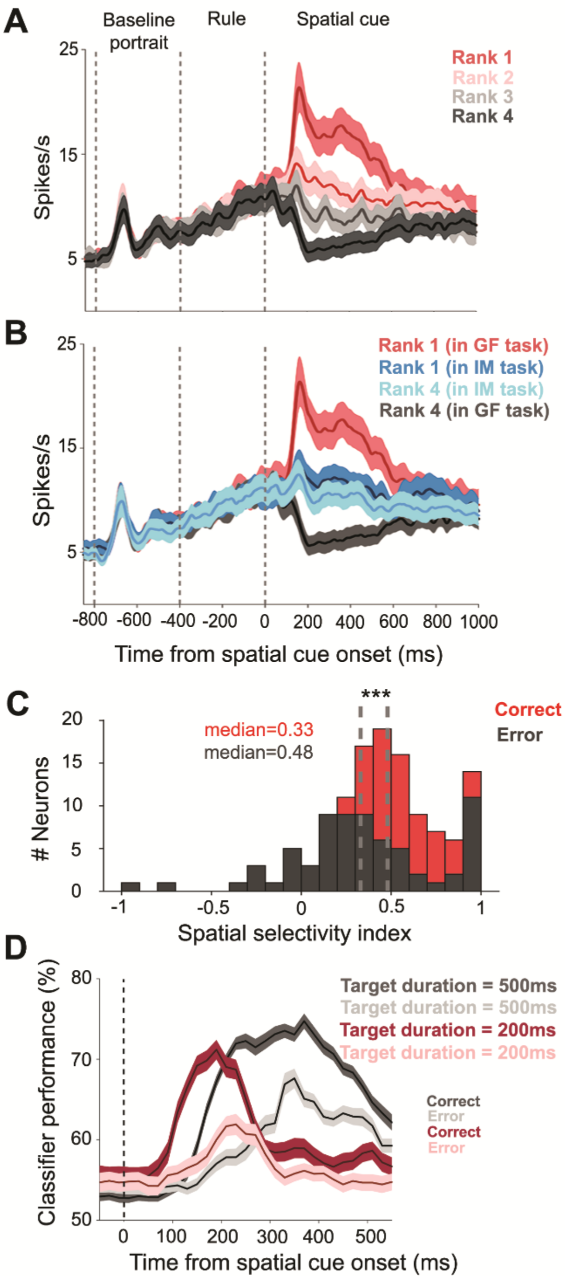

Figure 3. Population responses of spatially-selective GF neurons

(A) Population responses of 109 GF neurons from both monkeys for the target eliciting the strongest response (rank 1) in the spatial cue period, the second strongest (rank 2), the third (rank 3) and the fourth strongest response (rank 4). The population discharge associated with the two most preferred targets exhibited an increase in discharge rate, the ones associated with the least preferred targets a suppression. (B) Population responses of GF neurons for the most preferred and the least preferred targets in the GF task compared with the population responses to the same targets cued by identity in the IM task. (C) For the population of spatially-selective GF neurons, the median SSI values dropped significantly (Mann Whitney U test, p<0.001) from 0.48 for correct trials to 0.33 for error trials. SSI values for error trials could even become negative, indicating a reversal of preference, i.e. a target that was the most preferred one in the GF task became less effective than the least preferred one when studied in the IM task. (D) SVM decoding of the discrimination between the most preferred and least preferred target based on the activity of all spatially-selective GF, shown separately for correct and error trials and separately for the two durations of the spatial cueing periods (200ms vs. 500ms). Standard errors were obtained by deploying a bootstrapping procedure (n=1000). Shaded areas represent standard errors in all figures parts.

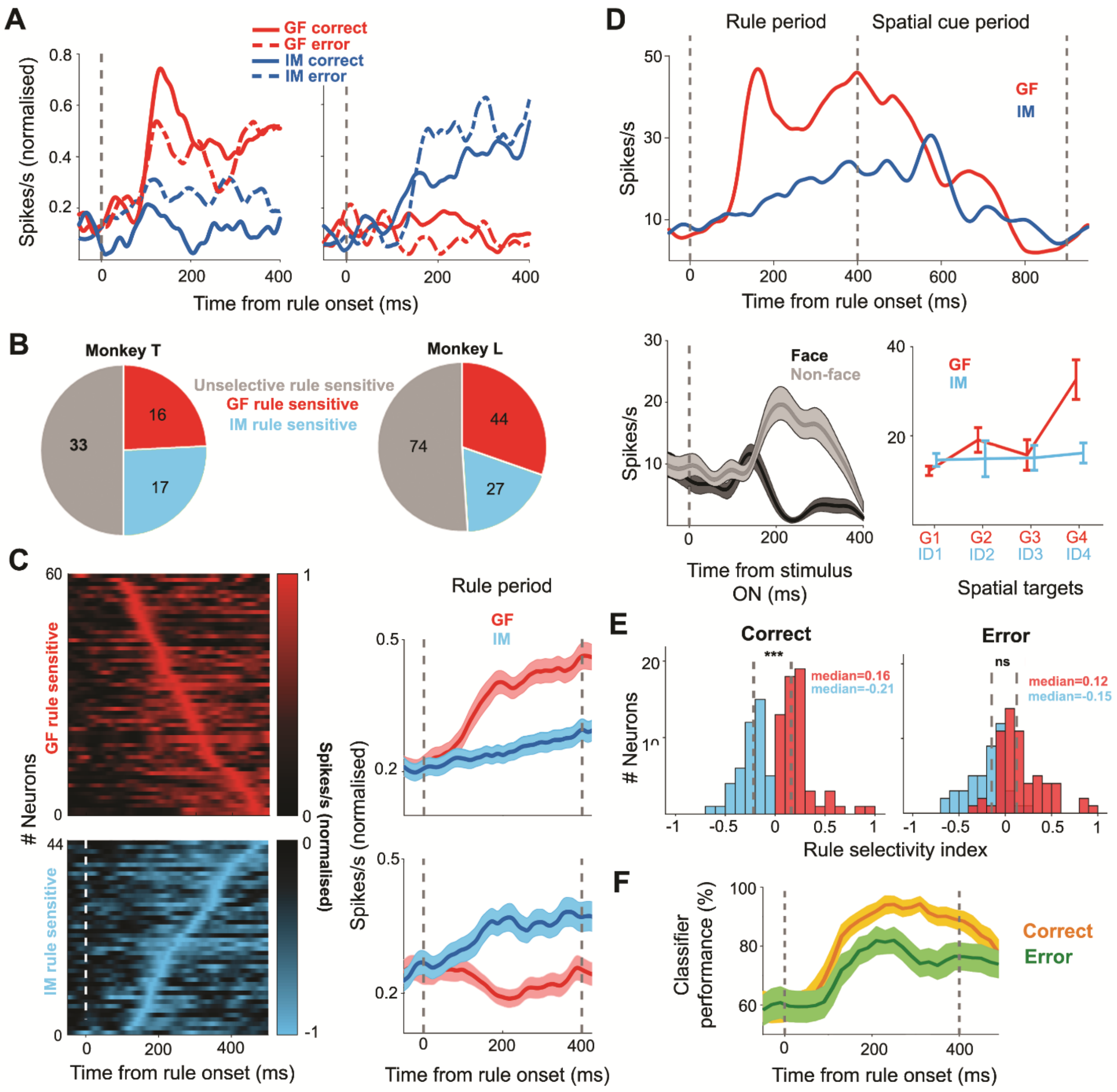

Figure 4. Rule selectivity of GFP neurons

(A) Responses of two exemplary GFP neurons to the presentation of the two rules, shown separately for correct and for error trials. The one on the left exhibited selectivity for the GF rule, while the one on the right preferred the IM rule. (B) Breakdown of rule preferences of rule-sensitive GFP neurons for both monkeys. (C) The right panel depicts the time course of the population responses of the two pools of neurons preferring the GF rule (n=60) and the IM rule (n=44) respectively, tested in both the GF and the IM task. The panel on the left plots the contributions of individual neurons ordered according to the latency of their peak discharge rates. (D) Exemplary GFP neuron demonstrating that preferences for rules and spatial cues are yoked. This neuron preferred the gaze rule and exhibited a much stronger response to the gaze cue in the subsequent 500ms of the spatial cueing period. The vertical line at 400ms denotes the onset of spatial cue. This neuron preferred non-face objects over faces in the passive viewing task. (E) Deterioration of the rule selectivity of the GFP as captured by the RSI results in erroneous decisions. (F) SVM decoding accuracy obtained from GFP rule-selective neurons for correct trails as compared to error trails. The shaded area represents standard deviations obtained by bootstrapping (n=1000). Note the clear drop in performance for error trials. Error bars and shaded areas represent standard error in figure parts except for (F) which shows standard deviations.

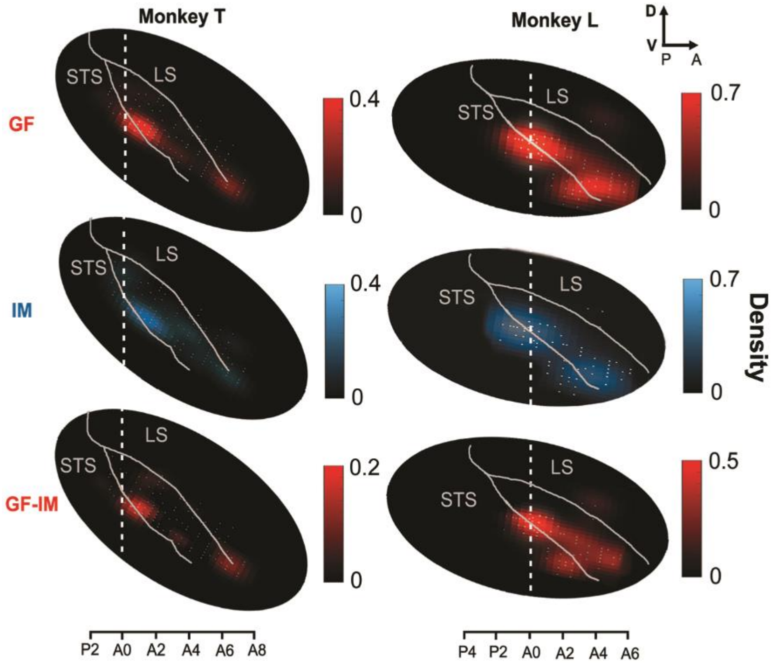

Figure 5. Topography of the GF and IM neurons in the pSTS

Heat maps of the density of GF neurons (red) and IM neurons (blue) in the pSTS of the two monkeys. The two lowest panels depict the contrast between the heat maps of GF neurons and IM neurons density. A0 is the interaural plane. The density scale represents the number of neurons in elements of 0.5 mm^2 of the pSTS surface after passing through a 2D-spatial filter (Gaussian, σ=2mm).

Supplementary Information for

Decoding of the other´s focus of attention by a temporal cortex module

H. Ramezanpour[1,2,3,*] and P. Thier[1,4,*].

[1] Department of Cognitive Neurology, Hertie Institute for Clinical Brain Research, University of Tübingen, 72076 Tübingen, Germany.

[2] Graduate School of Neural and Behavioural Sciences, University of Tübingen, 72074 Tübingen, Germany.

[3] International Max Planck Research School for Cognitive and Systems Neuroscience, University of Tübingen, 72074 Tübingen, Germany.

[4] Werner Reichardt Centre for Integrative Neuroscience, University of Tübingen, 72076 Tübingen, Germany.

*Correspondence to: Peter Thier and Hamidreza Ramezanpour, Department of Cognitive Neurology, Hertie Institute for Clinical Brain Research, Hoppe-Seyler-Str. 3, 72076 Tübingen, Germany. E-mail: thier@uni-tuebingen.de ; hamidreza.ramezanpour@uni-tuebingen.de .

This PDF file includes:

Supplementary text

Figure S1

Legend for Movie S1

Other supplementary materials for this manuscript include the following:

Movies S1

Supplementary information text

Low spatial selectivity of face-selective neurons

We also recorded the activity of the same pSTS neurons tested in the two active tasks during passive vision of face and non-face stimuli in order to investigate to what extent the spatially selective neurons are the same classically face-selective neurons. Twentythree out of 334 neurons tested exhibited significantly stronger responses to the passive vision of faces than to non-face stimuli ($p<0.05$; Wilcoxon signed-rank test comparison of mean response to faces and non-face stimuli). Out of this total of 23 face-selective neurons sampled from both monkeys, only six were from the set of 109 gaze following neurons with spatial selectivity. Figure S1A compares the population response of these face-selective neurons in the three tasks, as well as the amount of their spatial modulation for correct and incorrect trials. As can be seen, the average SSI values were very similar and statistically not different for correct and incorrect trials. Also, the response profiles for the most preferred and least preferred targets did not show any significant difference in the spatial cueing period (Mann Whitney U test, $p>0.05$). Hence, these neurons cannot have immediate relevance for the behavioral choices.

Abstract symbol-matching responses

We also investigated if neurons activated in the identity mapping task might also be responsive to learned associations between non face images and spatial locations. To this end, we trained the monkeys to associate four abstract symbols (square, circle, triangle, star) almost the same size (5^o x 5^o) with the four targets. The trial structure and timing were similar for the GF and IM tasks with the exception that there was no change in the color of the central fixation which could serve as a rule and the four abstract symbols were presented right after the same portrait fixation face used in the previous tasks. As this control task was usually run at the end of a session, once sufficient data in the three main tasks had been collected, only a minority of neurons could be tested on this abstract symbol mapping control. In many other cases, the monkeys were no longer motivated to work or the quality of the spike isolation had deteriorated too much. We only considered neurons for which we had collected at least eight correct trials for each of the four spatial targets. Both monkeys learned the task well and their performance was very similar to their performance in the gaze-following and identity-mapping tasks (monkey T: mean ± std= 73 ± 1%, monkey L: mean ± std= 74 ± 2% correct). In total, we could test 131 out of the 426 neurons tested on the gaze following and the facial identity mapping task (44 neurons from monkey L and 87 from monkey T) in the symbol mapping task. Only eight out of the 131 neurons (two from monkey L, six neurons from monkey T) exhibited spatial selectivity, characterized by significant target specific responses in the spatial cueing period (1-way ANOVA (Kruskal-Wallis test, $p<0.05$). In three of the eight neurons, the number of error trials was sufficiently large to allow a comparison of the responses between correct and incorrect trials. These comparisons did not reveal any significant difference (Mann Whitney U test, $p>0.05$). Hence, rather than reflecting spatial selectivity, these may be more elementary visual responses evoked by particular abstract objects. On the other hand, given the low probability of neurons exhibiting significant responses to the presence of abstract symbols (eight out of 131 neurons tested; i.e. 6% of the population), these neurons may simply be statistical artifacts.

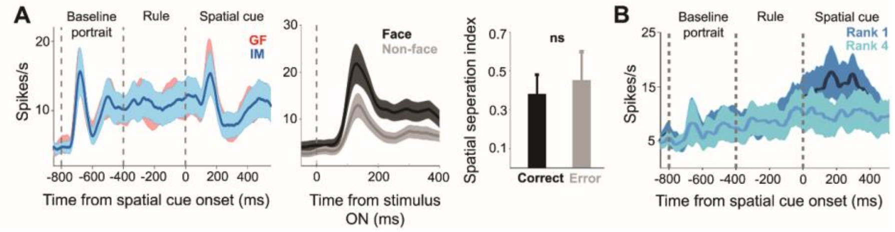

Figure S1. Population response profile of the passive face-selective and IM neurons

(A) The population discharge profiles of classical face-selective neurons, exhibiting a significantly larger response to faces as compared to non-face stimuli when tested in the passive viewing task. Both profiles show significant responses to the presence of the portraits in both the baseline portrait and the spatial cueing periods, yet without any difference between the gaze following and the identity matching tasks. Moreover, the population discharge did not differentiate between the most and the least preferred targets, no matter if spatial choices were correct or not. Hence, they hardly contribute to shaping monkeys´ spatial choices. (B) The population discharge profile of the spatially-selective IM neurons did not differentiate significantly between the most preferred and the least preferred IM targets. Error bars and shaded areas represent standard error in all subfigures.

Movie S1.

Exemplary neuron exhibiting a burst of activity for shifts of the experimenter´s gaze to the left.

Overall conclusions and outlook

The scope of this dissertation was to shed more light on the neural substrates and mechanisms affording gaze following, a key step in establishing joint attention and eventually forming a TOM. Disturbances of joint attention and gaze following are arguably the key cognitive impairments in autism spectrum disorders. Hence, the conceptual advance offered by this dissertation may help to better understand the pathophysiology of an important neuropsychiatric disease.

Another interest associated with this work was the question if gaze following of humans has a relationship to gaze following exhibited by old world monkeys, in particular macaques, eventually suggesting a shared evolutionary history. This is why we tried to add information to the comparison of the brain regions responsible the for computation of gaze following-related information in monkeys and man. In an fMRI experiment on healthy human subjects, we could identify a gaze following patch (GFP) in an anatomical location relative to the face-patch system very similar to the one previously established for monkeys by Marciniak and colleagues (Marciniak, Atabaki et al. 2014). The lack of face selectivity of the GFP together with its anatomical dissociation from the neighboring face patches suggested that the GFP computes distinct features of the face confined to the specific purpose of supporting gaze following rather than providing information on identity or expression, functions thought to rely on the face patch system. The hypothesis of a highly specific role in gaze following could be confirmed by recording from single neurons within and outside of the GFP in monkeys, based on the well-founded assumption that the monkey gaze following system is very similar to the human one if not actually homologous. We found that many gaze following neurons show precise spatial tuning to the gazed-at targets in the GFP in the middle-posterior STS, a region as yet not thought to be involved in the processing of spatial information. Many of the gaze following-selective neurons were also sensitive to contextual information such as the rule to follow or alternatively to suppress gaze following suggesting that the high order cognitive signals for the fine-tuning and control of gaze following originating from prefrontal cortex (see Appendix 3) take control at the level of individual neurons in the GFP. We then demonstrate that the information provided by gaze following neurons in this GFP indeed fully predicts the monkey´s performance when asked to shift attention to the object singled out by the other’s gaze, strongly suggesting causality. Finally and particularly importantly, we could show that the GFP is not able to select the potential target when there are more than one lying on the gaze vector. This ambiguous situation requires contributions from a cortical region located at the junction between prefrontal and premotor cortex, the inferior frontal junction region (IFJ) providing the information needed to disambiguate the potential target.

The properties of the GFP as revealed by BOLD imaging and the study of single neurons clearly support the notion that the GFP is a domain-specific module underlying the ability to use the other´s gaze to shift attention to objects of interest to the other and eventually to establish joint attention.

There are several questions left unanswered for future studies. One is how downstream areas, putatively downstream of the GFP such as parietal area LIP, needed to control covert and overt shifts of attention can use the GFP output. The idea would be that signals from the GFP enhance the priority of the location in the LIP priority map corresponding to the location of the gazed-at target. Another question is if the monkey GFP may offer not only gaze vector information but also information on a variety of other directions offered by the other´s body, integrating the various directional sources into a more general social attention vector. Finally, it will be very interesting to check if object-related information such as saliency or value can modulate the gaze following behavior at the level of the GFP.

References

Akiyama T, Kato M, Muramatsu T, Saito F, Nakachi R, Kashima H (2006) A deficit in discriminating gaze direction in a case with right superior temporal gyrus lesion. Neuropsychologia 44(2): 161-170.

Akiyama T, Kato M, Muramatsu T, Saito F, Umeda S, Kashima H (2006). Gaze but not arrows: a dissociative impairment after right superior temporal gyrus damage. Neuropsychologia 44(10): 1804-1810.

Allison T, Puce A, McCarthy G (2000) Social perception from visual cues: role of the STS region. Trends Cogn Sci 4(7): 267-278.

Aron AR, Robbins TW, Poldrack RA (2004) Inhibition and the right inferior frontal cortex. Trends Cogn Sci 8(4): 170-177.

Baron-Cohen S, (1994) How to build a baby that can read minds: Cognitive mechanisms in mindreading. Cahiers de Psychologie Cognitive/ Current Psychology of Cognition 13: 513-552.

Baron-Cohen S, (1995) Mindblindness: an essay on autism and theory of mind. MIT Press.

Baron-Cohen S, Baldwin DA, Crowson M (1997) Do children with autism use the speaker's direction of gaze strategy to crack the code of language? Child Dev 68(1): 48-57.

Butterworth G, Jarrett N (1991) What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy. British J Dev Psych 9(1): 55-72

Callejas A, Shulman GL, Corbetta M (2014) Dorsal and ventral attention systems underlie social and symbolic cueing. J Cogn Neurosci 26(1): 63-80.

Emery NJ, (2000) The eyes have it: the neuroethology, function and evolution of social gaze. Neurosci Biobehav Rev 24(6): 581-604.

Fodor JA, (1983) Modularity of Mind: An Essay on Faculty Psychology. MIT Press.

Friesen C, Kingstone A (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. Psychonomic Bulletin & Review 5(4): 490-495.

Haxby JV, Hoffman EA, Gobbini MI (2000) The distributed human neural system for face perception. Trends Cogn Sci 4(6): 223-233.

Hoffman EA, Haxby JV (2000) Distinct representations of eye gaze and identity in the distributed human neural system for face perception. Nat Neurosci 3(1): 80-84.

Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci 17(11): 4302-4311.

Kobayashi H, Kohshima S (1997) Unique morphology of the human eye. Nature 387(6635): 767-768.

Kobayashi H, Kohshima S (2001) Unique morphology of the human eye and its adaptive meaning: comparative studies on external morphology of the primate eye. J Hum Evol 40(5): 419-435.

Langton SR, Bruce V (2000) You must see the point: automatic processing of cues to the direction of social attention. J Exp Psychol Hum Percept Perform 26(2): 747-757.

Laube I, Kamphuis S, Dicke PW, Their P (2011) Cortical processing of head- and eye-gaze cues guiding joint social attention. Neuroimage 54(2): 1643-1653.

Marciniak K, Atabaki A, Dicke PW, Their P (2014) Disparate substrates for head gaze following and face perception in the monkey superior temporal sulcus. ELife 3.

Marciniak K, Dicke PW, Their P (2015) Monkeys head-gaze following is fast, precise and not fully suppressible. Proc Biol Sci 282(1816).

Materna S, Dicke PW, Their P (2008) Dissociable roles of the superior temporal sulcus and the intraparietal sulcus in joint attention: a functional magnetic resonance imaging study. J Cogn Neurosci 20(1): 108-119.

Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. Annu Rev Neurosci 24: 167-202.

Pelphrey KA, Morris JP, McCarthy G (2004) Grasping the intentions of others: the perceived intentionality of an action influences activity in the superior temporal sulcus during social perception. J Cogn Neurosci 16(10): 1706-1716.

Puce A, Allison T, Bentin S, Gore JC, McCarthy G (1998). Temporal cortex activation in humans viewing eye and mouth movements. J Neurosci 18(6): 2188-2199.

Ricciardelli P, Carcagno S, Vallar G, Bricolo E (2013) Is gaze following purely reflexive or goal-directed instead? Revisiting the automaticity of orienting attention by gaze cues. Exp Brain Res 224(1): 93-106.

Ridderinkhof KR, van den Wildenberg WP, Segalowitz SJ, Carter CS (2004) Neurocognitive mechanisms of cognitive control: the role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. Brain Cogn 56(2): 129-140.

Shepherd SV (2010) Following Gaze: Gaze-Following Behavior as a Window into Social Cognition. Front Integr Neurosci 4.

Shimojo S, Simion C, Shimojo E, Scheier C (2003) Gaze bias both reflects and influences preference. Nat Neurosci 6(12): 1317-1322.

Tomasello M, Carpenter M (2005) The emergence of social cognition in three young chimpanzees. Monogr Soc Res Child Dev 70(1): vii-132.

Tomasello M, Hare B, Agnetta B (1999) Chimpanzees, Pan troglodytes, follow gaze direction geometrically. Anim Behav 58(4): 769-777.

Tomasello M, Hare B, Lehmann H, Call J (2007) Reliance on head versus eyes in the gaze following of great apes and human infants: the cooperative eye hypothesis. J Hum Evol 52(3): 314-320.

Tsao DY, Moeller S, Freiwald WA (2008). Comparing face patch systems in macaques and humans. Proc Natl Acad Sci USA 105(49): 19514-19519.

Tsao DY, Schweers N, Moeller S, Freiwald WA (2008) Patches of face-selective cortex in the macaque frontal lobe. Nat Neurosci 11(8): 877-879.

Acknowledgments

I am very thankful to my thesis supervisor Prof. Dr. Peter Thier for trusting in me and giving me the unique opportunity to work in his lab and for his tremendous support during my PhD. Under Prof. Thier’s supervision, I have learned to be a critical thinker, to be honest, optimistic and goal-oriented. I learned how to tackle a scientific question from different angles and how not to be disappointed when facing failures. The most important lesson I learned was that science is always secondary and humanity and helping others must be first. He holds a very special place in my heart.

I am very grateful to all members of the lab for their great friendship and very constructive scientific discussions we had together, especially Ian, Akshay, Joern, Mohammad K and Mohamad S. I am also very grateful to Dr. Peter Dicke and Dr. Friedemann Bunjes for their invaluable technical support. Big thanks to Dagmar Heller-Schmerold and Ute Grosshennig for the kindness and help with all administrative stuff. I thank my advisory board members Prof. Dr. Uwe Ilg and Prof. Dr. Martin Giese for very useful advice and comments during my advisory board meetings.

Lastly, I would like to thank my family. All of my academic achievements, if we can really call them achievements, would have never been real without having great parents who supported me in every step. And most of all I want to thank a very special person, my wife, Ghazal for the tremendous support, patience and care she provided during this long journey. She always stayed next to me with love and helped me overcome problems, difficulties, and disappointments. Thank you, Ghazal.