Graduate School of
Systemic Neurosciences
LMU Munich

Visual Attention in Social Contexts

Jairo Irenarco Pérez Osorio

Dissertation der

Graduate School of Systemic Neurosciences der

Ludwig-Maximilians-Universität München

Munich, August 2016

Date of submission: August 10th, 2016
Date of examination: December 21st, 2016

Supervisors: Prof. Dr. Agnieszka Wykowska
Prof. Dr. Hermann J. Müller
Second Reviewer: PD Dr. Markus Conci
Third Reviewer: Dr. Bert Timmermans

"Remember to look up at the stars and not down at your feet.
Try to make sense of what you see
and wonder about what makes the universe exist.
Be curious".
Stephen Hawking

Acknowledgments

My words fail short to thank all the persons that, willingly or inadvertently, contributed to make this dissertation –and my PhD– a reality.

Foremost, I would like to extend my infinite gratitude to my supervisor Prof. Dr. Agnieszka Wykowska who has supported me throughout my scientific carrier. A truly admirable person, absolutely passionate for science, and moved by a tirelessness curiosity. Her remarkable readiness to not only open and share her knowledge and experience, but also to listen others ideas, makes her a very unique person among the selfish and greedy academic landscape. She guided me with her worthwhile advice, being present and at the same giving me the chance to work independently. It is hard to wish for a better supervisor. I feel grateful with Prof. Dr. Hermann Müller who offered his broad knowledge and valuable experience to guide my project. His broad knowledge and experience served as a valuable support to face critical situations. I fully appreciate the time he devoted to make the manuscripts excellent. I would like to thank my supervisor Prof. Dr. Gordon Cheng, member of my thesis advisory committee (TAC). His extensive experience and up-to-date knowledge of both academia and private sectors offered a rich and meaningful advice in order to expand the findings of my research into further applications.

I am also very grateful with Marius t'Hart for his invaluable help in programming the eye tracking experiments. I would like to thank all the NCP students that participated in the studies, and with Lee, Ella, and Mallory who helped me to develop the images used in my research, Nona

and Ebru for the data collection. As well, I sincerely appreciate the financial support given by the Excellence Cluster of CoTeSys and the Graduate School of Systemic Neuroscience.

I would like to thank as well to my fellow colleagues from the LMU psychology department Eva, Iris, Ondrej, Kelsey, Julia, Melanie, Donatas, Anna, Dragan, David, Adriana, Yue, and, Siyi. They enriched the day by day in the office talking at lunch, having a coffee, jumping on the eisbach, or enjoying Mai Tais or ice cream. Thanks to my friends from GSN Jaime, Fernando, Aurore, Oriol, Elena, Giorgiana, who made even better all the social activities we had the opportunity to share.

As well, I would like to thank immensely to my family, especially to my parents Rosa and Irenarco, who together with my brother Javi and my cousin Patty, were there for me, always making me feel loved in the distance. Thanks for understand my decision of moving so far away, for support my dreams, for believe in me. Muchas gracias. Also many thanks to my aunt Myriam and my friend Linita, without them I would have never been able to follow my aspirations. I would like to thank to the friends who became my family in this side of the world Mafe & Anibal, Efsun, Idil, Kacha, Lee, Anja, Mariam, amazing persons that shared their love and care and make my life better. Finally, I present my absolute gratitude to my beloved Ebru who has been patiently beside me, helping me to stand up straight more than once, and encouraging me everyday to be a better person, scientist and friend. Her constant advice, support and contributions were absolutely necessary to achieve many goals including this dissertation. I feel very lucky to have found you.

TABLE OF CONTENTS

ABSTRACT

The human brain has developed mechanisms to estimate future internal and external events utilizing previous experiences. Based on predictive processing of information, appropriate cognitive resources can be available for goal-oriented behavior, to expedite the recognition and interpretation of critical stimuli and events, generate and select appropriate action alternatives, and adjust behavior to the environmental demands. A large body of evidence supports the predictive nature of cognitive processing across different domains: vision (Friston, 2005) and attention (Mehta & Schaal, 2002; Enns & Lleras, 2008), motor control (Wolpert & Flanagan, 2001), action understanding (Jeannerod, 2001; Kilner, Friston & Frith, 2007; Kilner, Vargas, Duval, Blakemore, & Sirigu, 2007), language (DeLong, Urbach, & Kutas, 2005), music (Keller and Koch, 2008), executive functions (Fuster, 2001; Wylie et al., 2006), and theory of mind (Frith and Frith, 2006). Evidence from studies on sensory and sensory-motor activity suggests that prediction modulates neuronal activity through reducing activation thresholds and/or increasing the signal-to-noise ratios, facilitating subsequent stimulus processing (for review, see Gomez et al 2004). Similarly, prediction is crucial to create a stable representation of the environment (Kveraga Boshyan, & Bar, 2007; Rolfs, Jonikaitis, Deubel, & Cavanagh, 2011), to identify action sequences (Schubotz & von Cramon, 2002), and to determine others' intentions (Frith & Frith, 2006).

The aim of the three studies was to examine how higher-order predictions about others' actions influence participants' orienting of attention to gaze cues in a relatively complex social scenario. To define the framework for the series of studies, a review of the ubiquity of prediction mechanisms in the human brain will be presented; along with the basic models proposed to account for predictive processing in sensory, motor, and – importantly – social domains. Additionally, details about the relevance of the gaze in the social context, the orienting by the gaze and the relationship between gaze following and action sequences will be discussed. Subsequently, two chapters present three studies designed to disclose predictive processing of others' actions (or action goals) being at work in a naturalistic social scenario. The studies showed that expectations regarding action goals affected shared attention, as evidenced by gaze-cueing effects: when the gaze behavior of an observed actor violated observers' expectations, which were based on their understanding of the action context, gaze cueing effects were reduced or even completely eliminated. These effects are due to covert (rather than overt) attentional orienting to goal-congruent vs. -incongruent objects. Further evidence from eye tracking and EEG/ERPs reveals the brain mechanism involved in these processes. In the General Discussion, these findings are described in the context of prediction mechanisms. Further theoretical implications regarding social attention, theory of mind, and general interactions between low-level and higher-order cognitive mechanisms are considered.

1. INTRODUCTION

1.1 Predictive processing of the brain

The idea of predictive brain has gaining strength in the recent years. Increasing evidence suggest that predictive processing is inherent to all levels of our hierarchically organized neural system (Bubic, von Cramon, & Schubotz, 2010). For the first time in history of cognitive psychology predictive processing models promise to bring cognition, perception, action, and attention together within a common framework (Clark, 2013). In general, this approach suggests the probability-density distributions induced by hierarchical generative models as our basic means of representing the world, and prediction-error minimization as the driving force behind learning, action-selection, recognition, and inference.

Various models have been proposed for the mechanisms underlying predictive processing in the brain. However, two main models have been widely investigated: forward models, and the theory of predictive coding. Forward models aimed initially to explain how the internal prediction mechanisms contribute significantly to motor control (Miall & Wolpert, 1996; Jaennerod, 2001; Wolpert & Flanagan, 2001; Wolpert, Doya, & Kawato, 2003; Schubotz, 2007). These models suggest that predictions about the future states of the body are generated by emulations of musculoskeletal system dynamics using internal models. In detail, the motor system uses an ‘efference copy’ of the formulated motor command to produce (internal model-based) predictions of the sensory feedback; the sensory effects of the motor action. When the action is executed, predictions are compared with the sensory inputs to determine whether they match the anticipated effects. Comparison of results updates the model to increase the accuracy of predictions. The extensive application of the models have been proved for explaining prediction in other cognitive domains, including higher-order mechanisms, such as social cognition and theory of mind (Wolpert et al., 2003). Furthermore, Schubotz and von Cramon (2003) suggested a joint sensorimotor forward model which unifies the perceptual and motor models. Motivated by findings that show links between the motor system and perceptual processing, this model postulates that the brain can predict both motor and perceptual processes, regardless of whether such constitute sensory consequences of one’s own actions or expected sensory stimuli. The model assumes that mechanisms that emulate future states in both domains use the same computations, without considering both models as identical. In detail, motor processing requires lower accuracy predictions compared to perceptual processing, given that the systems rely on relational properties of external events. Thus, forward models have proven a considerable explanatory power that transcend its original purpose, offering valuable insights into the fundamental mechanisms of the Bayesian brain.

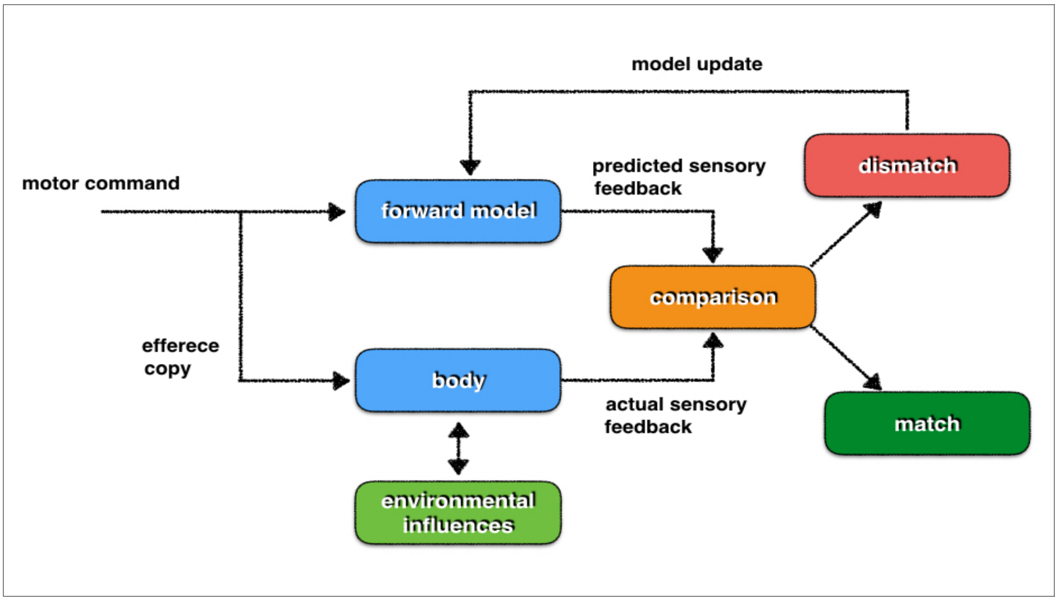

Figure 1A. Prediction in motor control.

Similarly, initially conceived for explaining the visual processing, the ‘predictive coding’ framework proposes that information flowing forward through a hierarchy of sensory regions is met at each stage by a top-down ‘prediction’ projected back from the layer immediately above (Friston, 2005). Sensory information flows through different levels in the visual cortex, a hierarchically organized structure (see Figure A2). Each level processes and compares sensory information with predictions transmitted through backward connections. The model postulates that when the sensory information and prediction do not converge, a ‘prediction error’ is detected and fed forward. After several repetitions of this process, such prediction errors update expectations across the different levels, leading to a progressive decrease in the prediction error at subsequent levels. Prediction errors might increase neural activity, reflecting attempts of the brain to update predictions and maintain stability in the representation of the environment (Friston & Stephan, 2007). This is made by integrating top-down expectancies with bottom-up information across multiple levels of processing (Friston, 2005).

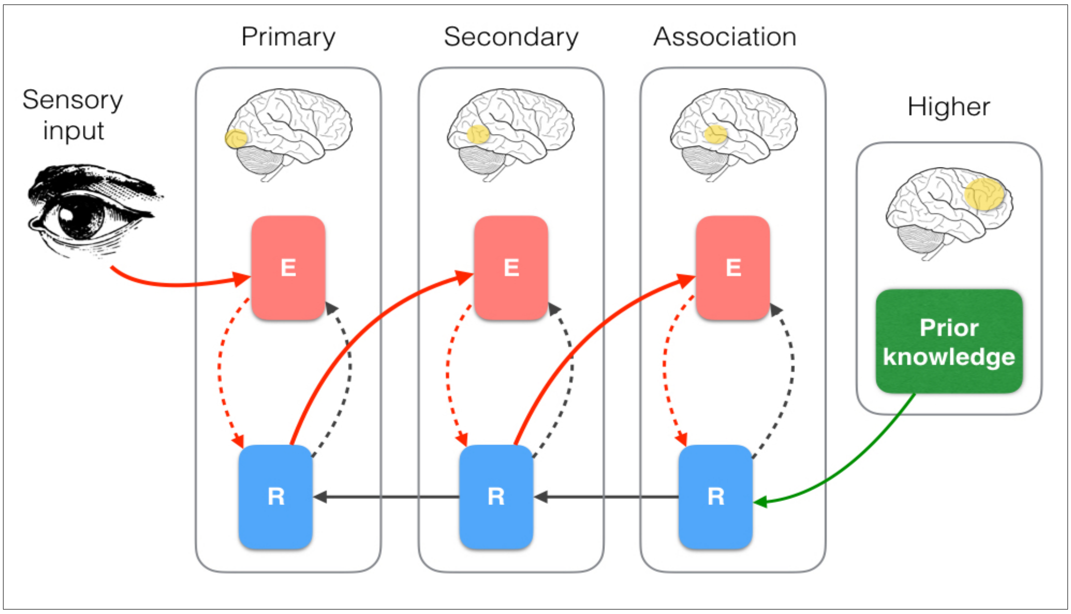

Figure A2. Schematic presentation of predictive coding in visual perception. Via iterative interactions, expected and observed information are reconciled within a hierarchically organized sensory system. Red arrows represent forward prediction error, black arrows represent backward predictions. Information travels from primary to higher association cortex (yellow circles). The main elements are two main types of cells, representation units (R) which encode expectations about what is possible or likely in the coming sensory world; and error units (E) driven by sensory input from the preceding layer of the hierarchy (red solid). Predictions are 'subtracted' from error unit responses –surprising events elicit larger neuronal responses. Representation unit predictions are updated on the basis of the mismatch between expected and observed information. Local circuit interactions (dashed arrows) occurring via many interactions between individual nodes subtract and update predictions. Resultant predictions in turn are used as priors for the preceding layer (black arrows).

Importantly, the proposal of co-existing representational and error neuron populations provides a natural account for the observation that cortical responses scale with sensory surprise: when stimulation confirms expectations, error neuron activity will be minimal, but when sensory evidence diverges from predictions, error units will respond vigorously. In conclusion, predictive coding and forward models postulate that the brain is regarded as an inference machine that follows Bayesian principles, constantly building models of internal and external events so as to predict future states (Knill and Pouget, 2004; Friston and Stephan, 2007). Collectively, predictions play a crucial role in the information processing, not only guide cognition and behavior actively –matching expected incoming events and evaluating the accuracy of our expectations– but also actively updating those predictions by preferentially selecting corresponding features in the environment.

1.1.1 Prediction brain structures

Different parts of the brain have been associated with predictive processing, specifically different sensory cortices (occipital, parietal and dorso-lateral), the motor system (premotor cortex), the cerebellum, the frontal cortex (orbitofrontal, medial frontal and dorsolateral cortex), insula, and subcortical structures (like thalamus, basal ganglia, and amygdala) (See Table 1 for functions associated with different regions). This shows that different aspects of prediction can be localized across the whole brain or nervous system, each of the structures playing a distinct role. Therefore, it is important not only to associate brain areas with predictive processing, but also to specidy the type of predictive processing they are related to.

Table 1A. Predictive brain. Areas involved with prediction following reported findings.

Prediction related function	Brain structure
Prediction on longer timescales (Eichenbaum and Fortin, 2009; Lisman and Redish, 2009)	Prefrontal cortex: medial temporal regions, especially the hippocampus
Imagining the future as well as remembering the past (Schacter et al., 2007; Schacter and Addis, 2009)	Posterior cerebral cortices: including the lateral parietal and temporal regions, the precuneus and the retrosplenial cortex.
Planning (Fuster, 2001); formulating temporal expectations (Coull and Nobre, 2008; Coull, 2009)	Parietal cortices together with premotor regions
Predictive motor processing (Schultz and Dickinson, 2000; Fleischer, 2007; Kotz et al., 2009)	Basal ganglia
Reward prediction (Knutson and Cooper, 2005)	Ventral striatum
Pain or emotional processing (Ploghaus et al., 1999; Porro et al., 2003; Ueda et al., 2003)	Amygdala, insula and the anterior cingulate cortex
Inferring others' intentions and mental states (e.g., Baron-Cohen, 1995; Frith & Frith, 2006; Adolphs, 1999; Brothers, 1990; Frith & Frith, 2006).	A network consisting of medial prefrontal cortex (mPFC), superior temporal sulcus (STS), fusiform gyrus (FG), amygdala and intraparietal sulcus (IPS)

1.2 Prediction of others' behavior

From having a conversation to playing sports, we attend to people's movements, facial expressions, and language, to anticipate others' mental states and intentions. These interactions are all embedded in an environment, usually involving objects, spatial arrangements and temporal restrictions, that should be taken into account for successful social interactions. For instance, background knowledge about the context of an observed action enables understanding and predicting others' action goals, intentions, as well as bodily and mental states (Frith & Frith, 2006). The following paragraphs will explore in more detail how predictive mechanisms in the brain are deployed (1) action understanding and prediction, (2) inferring mental states, and (3) spatial perspective taking.

1.2.1 Action understanding and prediction

Predicting where and how people are going to move is highly relevant in social interactions, allowing to anticipate bodily postures based on current states (Verfaillie & Daems, 2002), and to determine whether an action is intended or unintended (Grezes, Frith, & Passingham, 2004; Wolpert et al., 2003). Realizing all these functions involves rapid prediction of successive steps in action sequences. In order to successfully predict events and action sequences, actions need to be correctly identified. Recognition of action sequences is thought to be based on matching perceived actions onto corresponding action representations stored in semantic memory (Jeannerod, 2006). Kilner, Vargas, Duval, Blakemore, and Sirigu (2004) showed that an electrocortical readiness potential (RP)[1], commonly taken to reflect motor preparation processes, also manifested during mere action observation – and, importantly, before the action was actually performed by the observed actor. This indicates that motor preparation can be elicited by the mere anticipation of successive action events, rather than being based solely on the active preparation of a motor (re-)action. These results shed light on the mechanisms involved in action understanding, that is: mechanisms mapping an observed action onto an action representation in the observer's cognitive system.

According to Schubotz and von Cramon (2002), each action sequence has a 'syntax': a basic schedule that is fixed and mandatory (though tolerating some level of flexibility). There is evidence suggesting that observing actions triggers a corresponding action schema in the

---

[1] RP is observed at electrodes over motor cortex, typically C3/4, characterized by a more negative deflection over sites contralateral vs. ipsilateral to a manual response (see Eimer, 1998; Eimer & Coles, 2003; Gratton, Coles, Sirevaag, Eriksen, & Donchin, 1988; Libet, Gleason, Wright, & Pearl, 1986).

observer, including a goodness-of-fit evaluation between the observed action and the action schema (Rizzolatti, Fogassi, & Gallese, 2001). An action schema can be described on two levels: the goal of the action (see, e.g., Hamilton & Grafton, 2007) and its implementation, with the latter defined by the actor's movements and the objects involved (Manthey, Schubotz, & von Cramon, 2003). For an action schema to be triggered, at least one of the two components of implementation must be observable: the relevant objects and/or movements. Moreover, one more element is crucial for the activation of an action schema: the aim of the action. Goal-directed actions follow a largely predefined pattern: a coherent sequence of steps, which makes actions relatively predictable (Schubotz & von Cramon, 2002). Wurm & Schubotz (2012) showed how the compatibility between actions and action context modulates action recognition. When participants were asked to identify an action performed either in a compatible (i.e. squeezing lemons in the kitchen), incompatible (i.e. squeezing lemons in the bathroom) or neutral context (i.e. squeezing lemons with a empty white background). They found that participants were 100 ms slower recognizing incompatible contexts relative to compatible. Interestingly, action compatible contexts did not facilitate action recognition, as neutral and compatible conditions showed no differences. Additionally, neural responses recorded with fMRI showed an increased activation in the left ventrolateral prefrontal cortex only for incompatible contexts. Altogether, authors suggests that response delays may reflect the attempt to resolve a conflict between the observed action and context. Identify an action requires to embed the action step into an overarching action compatible with the provided context, possible by the activation of the left prefrontal cortex.

During action observation humans implicitly generate inferences about the intentions and goals that are associated with the movements (Wolpert et al., 2003). Such predictions are produced at the beginning of the movements and are tested by predicting how the movement will proceed. Violations of expectations regarding other's actions elicit increased activity in posterior superior temporal sulcus (pSTS) consistent with the notion that this region has a special role in reading intentions from movements. STS codes for biological movements, such as head direction, pointing gestures and in particular gaze direction, with eye gaze being the most important source for inferring the mental states of others (Baron-Cohen, 1995; Frith & Frith, 2006, for review). Numerous studies coincide with the crucial role of pSTS in the detection of unexpected movements (Pelphrey et al., 2003 and 2005; Saxe et al., 2004; Grezes et al., 2004; Senju et al., 2005, 2008; Tipples et al., 2012). In addition, the STS is also involved in face perception via its connections with the fusiform gyrus (Haxby, Hoffman & Gobbini, 2000, 2002; Hoffman & Haxby, 2000), in addition to the 'mirror-neuron system' (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Umiltà et al., 2001; see Rizzolatti & Craighero, 2004, for review). Reciprocal projections have also been found between the STS and the amygdala, a structure of the limbic system involved in the processing of facial expressions and emotional content of perceived information (Aggleton, Burton, & Passingham, 1980; Aggleton, 1993; Thomas et al., 2001). In sum, evidence suggest that specialized brain mechanisms generate expectations about other's behavior, and in addition, there are highly sensitive mechanisms that confirm or revoke such expectations.

1.2.2 Inferring mental states

Consider another person's mental perspective and predict what they can know contributes to humans ability to anticipate others' behavior. This requires to understand that people's actions are the result of the interaction between their own desires, knowledge, goals, and beliefs with the actual state of the world. This ability is a powerful skill and is referred as mentalising (Frith and Frith, 2003). Mentalising can be deployed in real or imaginary scenarios, experienced or novel situations, when behavior is observed, read about or illustrated (Frith & Frith, 2006). Nonetheless, this ability has been investigated with hypothetical situations. Participants are exposed to enacted, read or depicted stories and asked to either explain protagonists' actions or infer their thoughts, feelings and intentions.

Using this type of paradigm, different studies aimed to examine and characterize the cognitive mechanisms that play a role in mentalising. Firstly, increased activation in pSTS has been associated with bottom-up effects in mentalising tasks. Viewers seem to readily attribute emotions, desires and false beliefs to animated cartoons and abstract shapes with human like movements, in contrast to animations showing random or mechanical movements

(Heider and Simmel, 1944). Activation recorded during these tasks, revealed robust increases in activity in pSTS along with lesser increases in a number of brain areas including medial prefrontal cortex (MPFC) (Castelli, Happé, Frith & Frith, 2000; Martin and Weisberg, 2003). Elicited activation remained the same even when participants were informed that the movements could be intentional or randomized. Secondly, increased activation in MPFC has been associated with top–down effects in mentalising tasks. MPFC elicited higher activation when participants believed that they were interacting with a person relative to when they believed they were interacting with a pre-programmed agent. This pattern has been elicited consistently during competitive (Gallagher, Jack, Roepstorff, & Frith, 2002), trust (McCabe et al., 2001) or economic games (Rilling et al.,2004). Taken together, studies consistently reported activation in the MPFC and the pSTS (Frith and Frith, 2003; Saxe et al., 2004). These brain areas seem to be closely related with mentalising.

Importantly, humans, among other primates, carefully attend counterparts' faces to read out information about others' interest and intentions. A special skill of remembering, identifying and interpreting faces might be facilitated for different physical and environmental conditions (for review see Emery, 2000). Attending to gaze direction and facial expressions (Haxby, Hoffman, & Gobbini, 2000) provides important clues for inferring others' mood, interests, and, most importantly, for predicting others' intentions (e.g., Emery, 2000; Frith and Frith, 2006; Itier & Batty, 2009). This skill is acquired in the first months of life and play a crucial role in establishing social bonding and attachment. Studies revealed that even infants can identify and follow faces (Hood, Willen, & Driver, 1998; Vecera & Johnson, 1995). Also, along the lifespan, most people spend more time looking at faces than at any other type of objects (Haxby, Hoffman, & Gobbini, 2000). For social interaction perceiving faces is critical: it provides information about others, as well as hints to infer others' mood, interests and intentions.

1.2.3 Perspective taking

Another important factor that helps to predict behavior is the ability to appreciate another person's point of view. Once the other's perspective is reached we are able to respond to the actions of others and therefore able to predict goals, interpret other's intentions and beliefs, and even pursue mutual aims (Baron-Cohen, 1995), and this is essential to reason about other's mental states (Premack & Woodruff, 1978). Allocentric perception is required for social interactions and it is easily assumed when required. This was showed in a study of Tversky & Hard (2009). Participants were asked to describe the spatial relations of items in a video which included (or not) a person who could be presented as a passive element of the scene or performing an action. They found that percentages of responses using allocentric perception increased when there was a person in the scene, and he increase was significantly higher when the person was performing an action. In other words, once there is another human included in the context, observers spontaneously took other's perspective. Taking the other's perspective was boosted by observing an intended action. Such evidence is consistent with a another study that reports that actions at the disposal of another agent had an impact on one's own actions, even when the task at hand did not require taking the actions of another person into account (Sebanz et al., 2003). Both studies revealed that humans are prone to take others' perspective, which might facilitate understanding other's intentions and goals.

Interestingly, a recent study (Bukowski, Hietanen, & Samson, 2015) showed that the mere presence of the other person is not sufficient to trigger the computation of where/what that person is looking at. Participants were asked to perform a gaze-cueing task or a visual perspective taking (VPT) task, both including an avatar in the screen. They found that whether participants' attention is drawn to the person in the first place is context-dependent. Participants attended to the other person in the VPT; whereas they attended to external factors in the gaze cueing task. This has two main implications: first, deployment of attention is partially stimulus-driven and hence not fully automatic; and, second, the social mind-set created by the task goal determines where the attentional resources are allocated. Therefore, authors suggested that dependent on the task, narrowing down attention to the other person prevents competition from other potential sources of attentional capture or orienting. Similarly, non-social (attentional) factors might help prioritizing attention to the other person and trigger the computation of where and what the other person is looking at.

1.2.4 Gaze-based shared attention and prediction

Importantly, social interaction is embedded in an environment, objects and events are located in a particular location and are relevant in a particular situation. Therefore, monitoring where others attend is a crucial means for understanding and predicting events in social interactions. Based on the ability to attend where others attend, shared attention is a rudimentary mechanism underlying social cognition (Baron-Cohen, 1995). There is ample evidence that humans follow the gaze direction of others (e.g., Driver et al., 1999; Friesen & Kingstone, 1998; Hietanen, 1999; Langton & Bruce, 1999; Vuilleumier, George, Lister, Armony, & Driver, 2005), so as to obtain information relevant for interaction, for instance, about significant objects or events in the environment. Furthermore, the ability to discern and follow others' gaze direction constitutes an essential component of the ability to infer their current mental states, and helps establish a common social context (e.g., Baron-Cohen, 1995, 2005; Baron-Cohen, Jolliffe, Mortimore, & Robertson, 1997; Emery, 2000; Sebanz et al., 2006).

Experimentally, shared attention based on gaze following has been studied using the gaze-cueing paradigm. This is a modification of the classical spatial cueing paradigm. In attentional cuing experiments like Posner's spatial cuing task (Figure A3), participants detect a visual cue that appears at a peripheral location before the target appearance is showed. Cues that correctly predict a target's subsequent location are called "valid" and conversely, non-predictive cues are "invalid". Reaction times (RTs) are shorter to detect validly cued targets than invalidly cued targets, indicating that people shift attention to the cued location. Two types of attentional orienting have been identified in this task, exogenous orienting (in response to events in the environment) and endogenous orienting (in response to internal factors such as motivations, expectancies, and goals) (Klein, Kingstone, & Pontefract, 1992; Klein & Shore, 2000). Each of the attentional orienting types corresponds usually to different cue types: Peripheral cues (e.g. a brief flash at the predicted target location) direct attention exogenously, or centrally presented (symbolic) cues (e.g. target location informed using an arrow or a word) direct attention endogenously. Peripheral cues tend to orient spatial attention automatically or reflexively independent of their validity in a so-called bottom-up manner. These cues often cannot be ignored and do not interfere with symbolic cues (Jonides, 1981; Müller & Rabbit, 1989). On the contrary, central symbolic cues require observers to shift attention to the cued location in a top-down manner. In turn, these can more easily be ignored, and peripheral cues can interfere with central cues resulting in slower RTs in the valid condition, as compared to exogenous cues (Jonides, 1981; Müller & Rabbit, 1989).

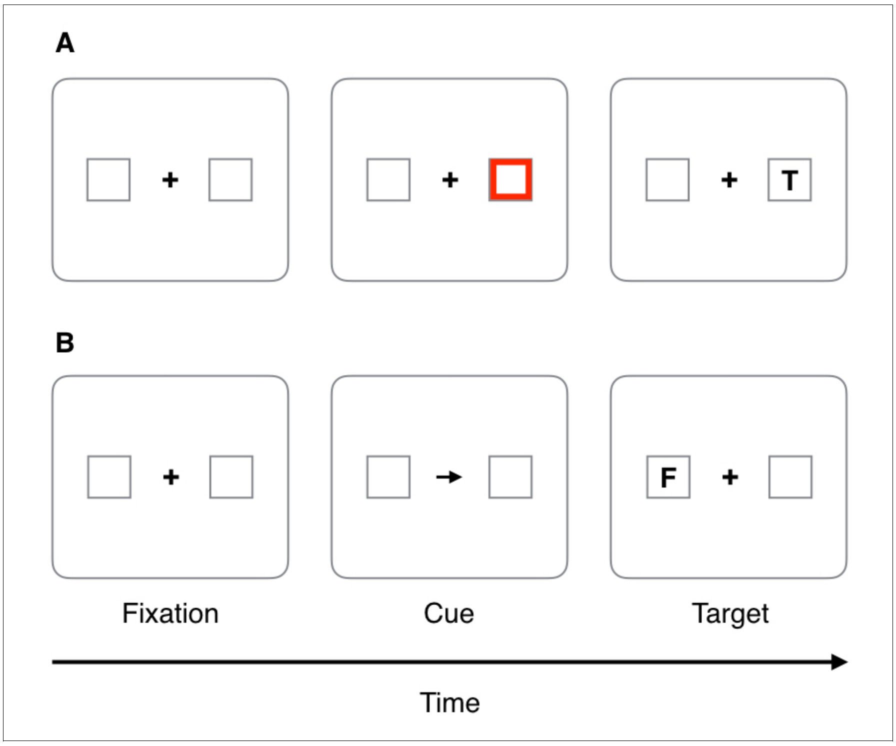

Figure A3. Classical spatial cueing paradigm. Using a peripheral sudden-onset cue (Panel A) or a central symbolic cue (Panel B) in a discrimination task. In Panel A, the target appears in the previously cued location (valid trial), whereas Panel B shows an invalid trial in which the target appears in the uncued location.

Studies on attention orienting have provided substantial evidence that cognitive resources are driven to peripheral locations by different kinds of cues which appear before a target stimulus onset (Posner, 1980; Posner, Snyder, & Davidson, 1980).

In the case of a standard gaze-cueing paradigm (Figure A4), gaze shift of a face (or face-like) stimulus constitutes the cue. Typically, a face (real or schematic) in canonical view is presented centrally prior to the onset of a target in the periphery. Subsequently, the face's eyes are directed towards one of the sides of the visual field – a potential target position. In a typical gaze-cueing study, processing of the target (detection, localization, or discrimination) is facilitated when the gaze direction and target position coincide, relative to when the gaze is directed elsewhere – *gaze-cueing effect*. Similar to traditional attentional cuing results, faster RT for the cued targets are reported indicating that attention has been directed to where the eyes were looking (e.g. Friesen & Kingstone, 1998; Langton & Bruce, 1999). Cuing effects are reliably observed with Stimulus Onset Asynchronies (SOA) between 250 and 750 ms (Driver et al., 1999; Kingstone, Tipper, Ristic & Ngan, 2004), and disappear around 1000 ms (Friesen & Kingstone, 1998). The effect has been reported as early as 3 months of age (Farroni, Johnson, Brockbank, & Simion, 2000; Hood et al., 1998) and is present even for simple schematic drawings of faces and eyes (Friesen and Kingstone, 1998). Likewise, the magnitude of this orienting effect is similar regardless of the identity of the cue, whether it is a human face, an animal face (e.g. ape or tiger), an object such as an apple or a glove with eyes (Quadflieg, Mason, & Macrae, 2004, see however Wiese, Wykowska et al., 2012 for different findings), or an inverted face (Tipples, 2005).

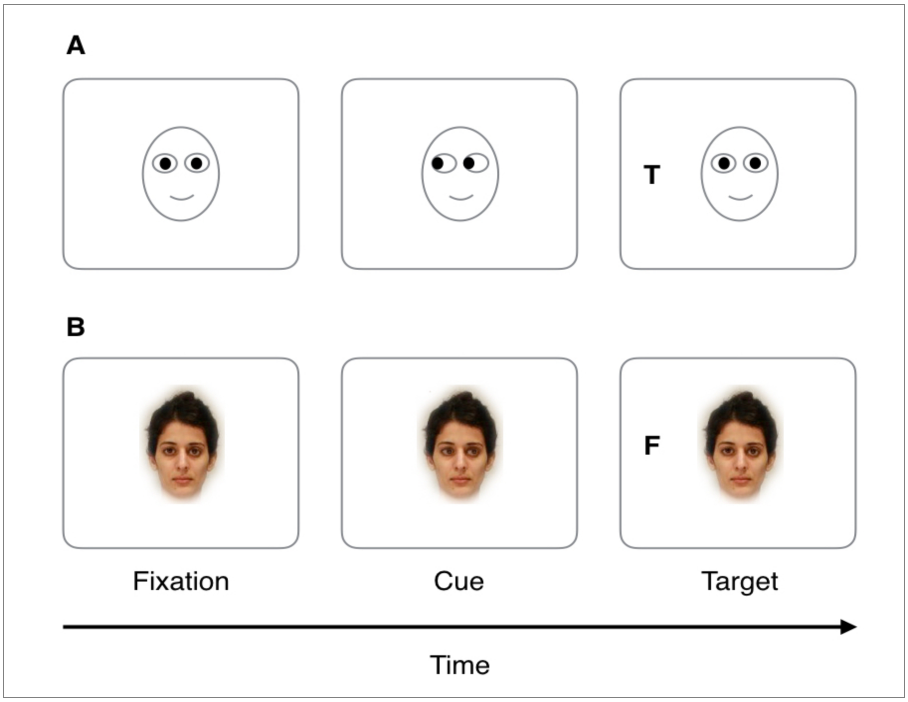

Figure A4. Basic gaze-cueing paradigm. (A) using a schematic drawing (valid trial) or (B) a real-life photograph of a face (invalid trial) in a discrimination task.

Following the attentional cueing literature, gaze cue being a symbolic centered cue might then be described as endogenous orienting of attention. However, many studies (e.g. Friesen & Kingstone, 1998) report that the effect is evoked in a reflexive (exogenous) way, independent of validity and volitional control. Collected evidence point out the reflexive character of gaze-cueing effect, which Driver et al. (1999) summarized, as follows: "*operating without intention; operating contrary to intention; independent of set-size; unconscious; innate; highly practiced; informationally encapsulated; cognitively impenetrable; modularized; and dependent on dedicated neural systems*" (pp. 531)

More recently, it has been suggested that attentional orienting in response to gaze direction is susceptible to top-down modulation (e.g., Hoffman & Haxby, 2000; Kanwisher, 2000; Ristic & Kingstone, 2005). For example, Teufel and colleagues (Teufel, Alexis, Clayton, & Davis, 2010; Teufel, Fletcher, & Davis, 2010) proposed that information about others' mental states influences automatic components of the gaze-following response. Similarly, Wiese, Wykowska, Zwickel, and Müller (2012) investigated whether the mere belief that the observed agent is an intentional system influences gaze following. Interestingly, they found the magnitude of the gaze-cueing effect to be dependent on whether or not the gazer was construed as an intentional system, independently of the gazer's physical appearance. Moreover, Wykowska, Wiese, Prosser, and Müller (2014) showed that the modulation of the gaze-cueing effect by beliefs about the gazer was mirrored by a modulation of the target-related P1 component of the EEG, indicating that early processes of perceptual selection are prone to top-down biasing from higher-order cognition. Taken together, these findings reveal that social perception is the result of an interactive process that involves the integration of bottom-up information provided by the stimulus and top-down influences by various social context variables. Similarly, Ristic & Kingstone (2005) addressed the differentiation between a pure bottom-up reflexive and a top-down control approaches reaching a middle ground. The authors claim that even though the response to the gaze is automatic, the face recognition is the result of a top down process (e.g. Hoffman & Haxby, 2000; Kanwisher, 2000). Once a stimulus activates inferior temporal cortex and is perceived as a face with features such as eyes, the attentional effect of this stimulus appears to be insensitive to top-down modulation. To sum up, social orienting of attention operates in complex dynamics that requires top-down control to evoke the bottom-up mechanisms broadly reported in the literature.

1.2.5 Neural mechanisms that support gaze perception

Several sources point to the specialized cerebral mechanisms directly related in the perception of gaze. Single cell recording (Perrett et al., 1985, 1990, 1991, 1992, 1994; Emery, Lorincz, Perrett, Oram, & Baker, 1997) and lesion studies in monkeys (Campbell, Heywood, Cowey, Regard, & Landis, 1990; Heywood and Cowey, 1992; Eacott, Heywood, Gross, & Cowey, 1993) have shown that neurons in the STS are specifically sensitive to the orientation of the eyes and recognition of others' gaze location. Furthermore, electrophysiological studies have revealed that: a) isolated human eyes evoke particularly large and early visual responses as compared to whole face stimuli or other isolated facial features (e.g. Bentin, Allison, Puce, Perez, & McCarthy, 1996; Taylor, Itier, Allison, & Edmonds, 2001); b) in infants, the occipito-temporal N170 evoked by isolated eyes, is present earlier than the same component elicited by whole face stimuli, indicating a faster maturation of eye processing system (Taylor et al., 2001; Farroni, Johnson, & Csibra, 2004; Farroni, Mansfield, Lai, & Johnson, 2004); c) eyes evoke the earliest and largest face-sensitive ERP responses (Schyns, Jentzsch, Johnson, Schweinberger, & Gosselin, 2003; Eimer, Lorincz, Perrett, Oram & Baker, 1997). Also, studies revealed a close relation between STS and amygdala. Namely, the STS region is likely to be essential for recognizing facial expressions as stimuli used in social communication, whereas the amygdala is likely to be essential for attaching socio-emotional significance to these stimuli (for review see, Bickart, Dickerson, & Barrett, 2014). And d) enhanced early visual evoked potentials (P1 and N1) were reported. Larger amplitudes and earlier onsets of these components were found in a gaze-cueing paradigm in the valid condition (Schuller & Rossion, 2001). These unveiled that reflexive attention to gaze increases visual activity and speeds up the processing of visual information –at least as soon as it reaches the extrastriate visual cortex. In addition, authors reported an enhanced P3 component for invalid trials, which might be related to decreasing target probability.

Similarly, neuroimaging studies in humans show that the superior temporal sulcus (STS) region is involved in gaze and face processing (Hooker et al., 2003; Pelphrey, Singerman, Allison, & McCarthy, 2003; Puce, Allison, Bentin, Gore, & McCarthy, 1998; for a review see Allison, Puce, & McCarthy, 2000). The intraparietal sulcus (IPS), which participates in covert shifts of attention (e.g. Corbetta, 1998; Nobre et al., 1997), is also significantly activated during averted gaze perception (George et al., 2001; Hoffman and Haxby, 2000; Hooker et al., 2003; Pelphrey et al., 2003; Puce et al., 1998; Wicker, Michel, Henaff, & Decety, 1998). Cells in the STS tuned to detect gaze direction, were found to modulate subcortical structures that are specialized in reflexive spatial orienting –and might doing so in a relatively automatic fashion. An fMRI study by Hoffmann and Haxby (2000) found that passive viewing of faces that had averted gazes, elicited a significantly stronger response in the IPS bilaterally and in the left STS, as compared to passive viewing of faces with direct gazes. This results are consistent with Puce et al. (1998) that reported activation of the inferior temporal sulcus (ITS) only to eye movements. Collectively, evidence suggests that even though STS and lateral fusiform area are closely connected and work together in face perception, the ITS is particularly sensitive to gaze. Activity in the IPS might be associated preferably with the spatial aspects of perceived eye gaze and its role in directing attention.

A more integrative approach postulates that two different mechanisms mediate the gaze cueing effect. Evidence from a split-brain patient study (Kingstone, Friesen, & Gazzaniga, 2000) found that non-predictive gaze direction produces a rapid and short-lived attention effect that is lateralized to patient's face/gaze-processing right hemisphere. Interestingly, non-predictive arrow direction elicited a cuing effect in both hemispheres. Furthermore, Friesen, Ristic, & Kingstone (2004) reported that in healthy observers eyes cues elicited a reflexive orienting of attention absent for arrows cues. Findings from both studies suggest that attentional mechanisms involved in gaze following are indeed 'special'. That is, reflexive orienting of attention by gaze-cueing depends on two different neural systems, one mediated subcortically and triggered by abrupt onsets (that supports reflexive attention to biologically relevant stimuli); and second one mediated cortically and triggered by perceived gaze direction (Friesen, Moore & Kingstone, 2005). Consistently, Greene and Zeidel (2011), reported a right hemisphere bias for attentional orienting cued by gaze (facilitation effect for targets in the left visual field), but not for attentional orienting cued by non-social stimuli. Taken together, authors suggest a theory of a separate neural system for orienting of attention induced by social cues, as well as a theory of separate parallel and simultaneous neural systems for attention in the two cerebral hemispheres.

Similarly, Hietanen et al. (2006) reported differences in the brain activity for gaze-cued compared to arrow-cued orienting. Mainly, in gaze-cued orienting elicited only three relatively small foci of activation in the left inferior occipital gyrus and right medial and inferior occipital gyri. Meanwhile for arrow-cued orienting, elicited large bilateral post-central activations in areas including the medial/inferior occipital gyri and medial temporal gyri, and in the left intraparietal area. Arrow cueing also activated the right frontal eye and supplementary eye fields. Taken together, evidence suggest that orienting of attention by gaze and following of arrow cues are supported by different cortical networks. In summary, evidence from several sources agree with the importance of the gaze-cueing effect and remark that very specific brain areas and networks are related to this phenomenon observed behaviorally.

1.3 Gaze behavior in action sequences

If gaze direction provides important clues regarding an observed agent's intentions, it is highly plausible that humans use gaze direction to anticipate successive (action) steps in complex action sequences, facilitating prediction of upcoming events in social environment. Indeed, there is evidence that eye movements provide useful hints for understanding actions and predicting successive action steps: Examining participants' oculomotor behavior in a block-stacking task, Flanagan and Johansson (2003) found that eye fixations invariably preceded the landing points of manual movements during task execution. Importantly, participants showed similar eye movement patterns merely by observing an actor performing the same task. They concluded that during action observation, humans implement similar oculomotor programs to those employed in action production – indicative of anticipatory mechanisms operating during action observation. Similarly, Hayhoe, Shrivastava, Mruczek, & Pelz

(2003) recorded eye movements of participants in natural situations (i.e. making a sandwich). Interestingly, eye fixations preceded action steps, that is, eye movements were strongly coupled to the task-relevant objects and anticipated their use. That leads to the conclusion that fixations pick up critical information to perform a task and support high precision movements. In sum, both Flanagan and Johansson (2003) and Hayhoe et al. (2003) showed that gaze behavior serves as a good hint regarding successive action steps of others. Furthermore, Sebanz et al. (2006) found STS to be particularly involved in updating predictions after a violation of an expected action sequence. Interestingly, as mentioned before the STS region is also involved in gaze direction detection and gaze cueing (Haxby et al., 2000). Hence, gaze following appears to be closely linked to anticipatory mechanisms in action.

1.4 Aim of the project

The aim of the present PhD project was to investigate whether high-level expectations concerning others' action goals and predictions regarding their successive action steps affect the fundamental mechanism of gaze following. To date, most studies have examined separately either gaze following or prediction mechanisms related to actions. The present project was, to our knowledge, the first one designed to investigate how attention to objects in a naturalistic scene is guided by two factors: the (perceived) gaze direction of an actor and background expectations regarding the action context. We designed a paradigm in which a gaze-cueing procedure was embedded within a complex action context, where the context was introduced at the start of each trial by an image that represented a certain social situation. Subsequently, we presented to participants photographs of a (female) human agent who gazed either at an action-congruent or an incongruent object. Participants' task was to discriminate the level of liquid (high vs. low) in a cup (target) situated next to an object (bottle) that could be either (1) action congruent or not: the action-congruency factor; and (2) gazed-at or not (gaze validity factor). The type of liquid in the (target) glass would always correspond to the liquid in the bottle placed next to it. Participants knew which glass was the target on a given trial, as the critical picture frame in the trial sequence contained only one glass filled with liquid. The crucial question was whether discrimination performance would depend on whether the observed agent's gaze was directed to the object that was congruent with the action context, and whether the target was gazed-at by the observed agent. The paradigm was thus expected to engender two types potential conflict: one related to gaze direction being incongruent with the action context, and one related to discriminating a target located on the side opposite to the (observed) agent's gaze direction (validity of the cue).

We expected performance in the target discrimination task to be affected by gaze validity, that is, observe the typical gaze-cueing effect. Furthermore, and importantly, we hypothesized that gaze-cueing effects would be modulated by whether the (observed) agent directed her gaze to an action-congruent or an incongruent object. The latter hypothesis was predicated on two ideas: (1) humans typically expect gaze direction to precede successive action steps (Flanagan & Johansson, 2003; Hayhoe et al., 2003), and (2) gaze-cueing effects reflect a combination of bottom-up reflexive mechanisms (Driver et al., 1999; Friesen & Kingstone, 1998) and top-down modulatory influences (Teufel et al., 2010; Wiese et al., 2012; Wiese et al., 2012; Wiese et al., 2012; Wykowska et al., 2014). Thus, in the present paradigm, gaze-cueing effects are likely to be modulated by the observers' action expectancies: participants would expect the (observed) agent to direct her gaze to the object that was to be manipulated according to the action context (i.e., action-congruent object), and violation of this expectation would influence the mechanism of gaze following, possibly attenuating the validity effects.

Behavioral studies

In Experiment 1, participants in general followed the gaze of the observed agent (showed gaze cueing effects), though the gaze-cueing effect was larger when the actor looked at an action-congruent object relative to an incongruent object. Experiment 2 examined whether the pattern of effects observed in Experiment 1 was due to covert, rather than overt, attentional orienting, by requiring participants to maintain eye fixation throughout the sequence of critical photographs (corroborated by monitoring eye movements). The essential pattern of results of Experiment 1 was replicated, with the gaze-cueing effect being completely eliminated when the observed agent gazed at an action-incongruent object. These findings show that covert attentional orienting in response to gaze can be modulated by expectations that humans hold regarding successive steps of the action performed by an observed agent.

EEG/ERPs Study

In the final study we examined –using the event-related potentials (ERPs) of the EEG signal– which stages of processing are influenced by expectations about others' action steps. We used a modified paradigm that was used in the behavioral studies. Similarly to the behavioral studies, a gaze-cueing procedure was embedded in successively presented naturalistic photographs composing an action sequence. Our findings: (1) replicated behavioral gaze-cueing effects modulated by whether the observed agent gazed at an object which was expected to be gazed-at, according to the action sequence; (2) showed modulatory effects on the P1/N1 components locked to the onset of a target: while P1 was modulated by gaze validity with respect to target location, N1 was modulated both by gaze validity and participants' expectations about where the agent would gaze at to perform an action; (3) revealed a more positive amplitude in the range of the N300 component –locked to the gaze direction shift– when the gaze was congruent with the action sequence, relative to incongruent and neutral conditions. Taken together, these findings revealed that confirmation or violation of expectations concerning others' goal-oriented actions can modulate attentional selection processes, as indexed by early ERP components.

2. Study 1

Gaze Following Is Modulated by Expectations Regarding Others' Action Goals

Jairo Pérez-Osorio [1,4], Hermann J. Müller[1,2], Eva Wiese[1,3] & Agnieszka Wykowska[1,4]

[1] Department of Psychology, Ludwig-Maximilians University, Munich, Germany

[2] Department of Psychological Sciences, Birkbeck College, University of London, London, UK

[3] Department of Psychology, George Mason University, Fairfax, United States of America

[4] Engineering Psychology Unit, Division of Human Work Sciences, Luleå University of Technology

[5] Chair for Cognitive Systems, Technical University of Munich

Author contributions: AW JP HJM, Conceived and designed the experiments. JP, Performed the experiments. JP AW, Analyzed the data. JP AW HJM EW, Wrote the paper.

Perez-Osorio, J., Müller, H.J., Wiese, E., & Wykowska, A. (2015) Gaze Following Is Modulated by Expectations Regarding Others' Action Goals. *PLoS ONE 10(11)*: e0143614.
doi:10.1371/journal. Pone.0143614

2.1 Abstract

Humans attend to social cues in order to understand and predict others' behavior. Facial expressions and gaze direction provide valuable information to infer others' mental states and intentions. The present study examined the mechanism of gaze following in the context of participants' expectations about successive action steps of an observed actor. We embedded a gaze-cueing manipulation within an action scenario consisting of a sequence of naturalistic photographs. Gaze-induced orienting of attention (gaze following) was analyzed with respect to whether the gaze behavior of the observed actor was in line or not with the action-related expectations of participants (i.e., whether the actor gazed at an object that was congruent or incongruent with an overarching action goal). In Experiment 1, participants followed the gaze of the observed agent, though the gaze-cueing effect was larger when the actor looked at an action-congruent object relative to an incongruent object. Experiment 2 examined whether the pattern of effects observed in Experiment 1 was due to covert, rather than overt, attentional orienting, by requiring participants to maintain eye fixation throughout the sequence of critical photographs (corroborated by monitoring eye movements). The essential pattern of results of Experiment 1 was replicated, with the gaze-cueing effect being completely eliminated when the observed agent gazed at an action-incongruent object. Thus, our findings show that covert gaze following can be modulated by expectations that humans hold regarding successive steps of the action performed by an observed agent.

Keywords: gaze cueing, expectations, action prediction.

2.2 Introduction

Social interactions require the ability to predict and understand others' behavior and its underlying intentions. To infer intentions and action goals, humans pick up various social signals, such as the others' gestures or gaze direction, providing information about their focus of attention or intended action steps. There is ample evidence showing that humans attend to facial expressions and gaze direction of others (e.g., Emery, 2000; Langton & Bruce, 1999; Haxby, Hoffman, & Gobbini, 2000; Vuilleumier, George, Lister, Armony, & Driver, 2005). The capacity for discerning and following others' gaze direction is an essential component of the ability to infer their current mental states, and helps establishing a common social context (e.g., Emery, 2000, Baron-Cohen, 1995; Baron-Cohen, Jolliffe, Mortimore, & Robertson, 1997; Sebanz, Bekkering, & Knoblich, 2006; Itier & Batty, 2009). Gaze following has been extensively studied using the gaze-cueing paradigm (e.g. Friesen & Kingstone, 1998; Hietanen, 1999), in which a face, in canonical view, is typically presented centrally prior to the onset of a target in the periphery. Subsequently, the face's eyes are directed towards one of the sides of the visual field – a potential target position. In a typical gaze-cueing study, processing of the target (detection, localization, or discrimination) is facilitated when the gaze direction and target position coincide, relative to when the gaze is directed elsewhere – the *gaze-cueing effect*. The gaze-cueing effect has been considered to rely on a reflexive mechanism (for review, see Hietanen, 1999), though more recently, it has been suggested that attentional orienting in response to gaze direction is susceptible to top-down modulation (e.g., Hoffman & Haxby, 2000; Kanwisher, 2000; Ristic & Kingstone, 2005; Bayliss, Schuch, & Tipper, 2010; Wiese, Wykowska, Zwickel, & Müller, 2012; Wykowska, Wiese, Prosser, & Müller, 2014). For example, Teufel and colleagues (Teufel, Alexis, Clayton, & Davis G, 2010; Teufel, Fletcher, & Davis, 2010) proposed that information about others' mental states influences automatic components of the gaze cueing effect. Similarly, Wiese, et al. (2012) examined whether the mere belief that the observed agent is an intentional system influences gaze cueing. They manipulated the likelihood of adopting the intentional stance by instruction (in some conditions, participants were told that they were observing a human or a robot, in others, that they were observing a human-like mannequin or a robot whose eyes were controlled by a human). Interestingly, the authors found the magnitude of the gaze-cueing effect to be dependent on whether or not the gazer was construed as intentional, independently of the gazer's physical appearance. Moreover, Wykowska et al. (2014) analyzed the event-related potentials (ERPs) of the EEG signal recorded during the same task and found that the impact of beliefs about the gazer on the gaze-cueing effect was mirrored by a modulation of the target-locked P1 component at posterior-occipital electrode sites, indicating that already early processes of perceptual selection are prone to a top-down bias from higher-order cognition. Taken together, previous findings reveal that social perception is the result of an interactive process that involves the integration of bottom-up information provided by the stimulus and top-down influences by contextual variables.

If gaze direction provides important clues regarding the intentions of an observed agent, it is plausible that humans also use gaze direction to infer the subsequent (action) steps in complex action sequences, facilitating prediction of what others are going to do next and of crucial upcoming events in social interactions. Thus, arguably, observing others' gaze behavior might elicit expectations about unfolding action sequences. Indeed, there is evidence that eye movements provide useful hints for understanding actions and predicting successive action steps: examining participants oculomotor behavior in a block-stacking task, Flanagan & Johansson (2003) found that eye fixations invariably preceded proactively the landing points of manual movements during task execution. Importantly, Flanagan and Johansson observed similar eye movement patterns when participants merely observed an actor performing the same task. From this, they concluded that during action observation, humans implement similar oculomotor programs to those employed in action production. Similarly, Hayhoe, Shrivastava, Mruczek, & Pelz (2003) recorded eye movements of participants in natural situations, such as when making a sandwich. The results indicated that eye fixations predicted action steps: eye movements were strongly coupled to the task-relevant objects and preceded their use. The authors concluded that fixations serve to pick up critical information for performing the task and support high-precision movements. In summary, both studies reveal that gaze behavior provides good hints regarding successive action steps of others.

Similarly to gaze-induced expectations regarding successive action steps, humans also develop expectations regarding the way actions themselves unfold. For example, Wurm & Schubotz (2012) showed that videos of action sequences incongruent with an action context produced longer recognition times, as compared to action sequences congruent with the context. Furthermore, several authors have claimed that observing actions triggers a corresponding action schema in the observer, including a goodness-of-fit evaluation between the observed action and the action schema (Rizzolatti, Fogassi, & Gallese, 2001). An action schema can be described on two levels: the goal of the action (see, e.g., Hamilton & Grafton, 2007) and its implementation, with the latter defined by the actor's movements and the objects involved (Manthey, Schubotz, & von Cramon, 2002). In sum, evidence suggests that people have expectations regarding subsequent action steps, as goal-directed actions follow a largely predefined pattern: a coherent sequence of steps, which makes actions relatively predictable (Schubotz & von Cramon, 2002) – a notion also supported by electrophysiological evidence (Kilner, Vargas, Duval, Blakemore, & Sirigu 2004).

The aim of the present study was to examine the interplay between expectations about an observed action and gaze-cueing effects. Consistent with the notion that, in daily life, gaze is informative with respect to subsequent action steps of an observed agent, and with empirical evidence in support of this notion [21, 22], we hypothesized that participants would have certain expectations regarding where an observed agent should gaze, given the action sequence the agent is performing. This, in turn, might affect gaze following (gaze-cueing effects), as gaze-cueing effects have been shown to be affected by how much 'social sense' is involved in the gaze behavior (Wiese et al., 2012; Wykowska et al., 2014; Teufel et al., 2010a). With regard to expected action sequences, if the observed gaze behavior is in line with the expected pattern, it would make more social sense to the observer – who might therefore more readily follow the other's gaze relative to when the observed gaze behavior contravene expectations. To implement these ideas in an experimental study, we designed a paradigm in which a gaze-cueing protocol was embedded in a scenario that would evoke expectations regarding action sequences and gaze behavior of an observed agent. We were interested in examining how attention would be guided by gaze direction (gaze-cueing effects) when the expectations regarding action sequences would be either confirmed or violated. In our paradigm, a gaze-cueing procedure was embedded in a series of naturalistic photographs depicting a person (a woman named 'April') completing a goal-oriented task. At the beginning of each trial, an image introduced an action goal: it depicted either a guest asking her to bring her something to drink, or her flat mate asking her to fetch fabric softener to do the laundry. Afterwards, April was depicted in the kitchen with two bottles located to her left and right, respectively – each containing one of the liquids: either orange juice or fabric softener. Beside each bottle, there was a plastic cup. Subsequently, April gazed at either the action-congruent or action-incongruent bottle (e.g., in the context of bringing a drink to her friend, the congruent bottle would be the one with the orange juice, while the incongruent one would be that containing pink softener). In the final frame, some of the liquid (either orange juice or softener) appeared in one of the plastic cups (the target), and participants' task was to discriminate whether the level of liquid in the plastic cup was high or low. Only one cup contained liquid and this always corresponded to the adjacent bottle. The crucial question was whether (liquid-level) discrimination performance would depends on whether the target was, or was not, gazed-at by the observed agent (*validity* of the gaze with respect to subsequent target presentation – the classical *gaze-cueing* manipulation) and how the validity effect would be modulated by whether the observed agent's gaze was directed to the object congruent or incongruent with the action context (*congruency* of the actor's gaze with respect to the action). We expected performance of the discrimination task to be affected by gaze validity, that is, to show the typical gaze-cueing effect. Importantly, we additionally hypothesized that the gaze-cueing effect would be modulated by whether the observed agent directed her gaze to an action-congruent or an action-incongruent object, in accordance with the ideas sketched above.

2.3 Experiment 1

Method

Participants. To determine the sufficient sample size for Experiment 1, we conducted an a-priori power analysis for the effect of congruency on gaze cueing, using: (i) a moderate effect size ($d_z = .6$), (ii) an α-error equal to .05, and (iii) a power level of .80 (as recommended by Cohen, 1992). This analysis yielded an adequate sample size of 24. A total of 27 participants were recruited for the experiment to obtain 24 useable data sets (three of the initial 27 had to be excluded due to error rates higher than 15%). All 24 participants included in the analyses (age range 21–35 years, M = 24.36 years; 16 women; all right-handed) reported normal or corrected-to-normal vision, and normal color vision. None of the participants had previously taken part in an experiment with a similar design.

Ethical statement. Experiment 1 (as well as Experiment 2; see below) was conducted at the Department of Experimental Psychology, LMU Munich, where all experimental procedures involving the collection of purely behavioral data (e.g., reaction times and error rates) with healthy adult participants (i.e., procedures that do not involve any invasive or potentially dangerous methods) are approved by the Department's ethics committee in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). Data were stored and analyzed anonymously. Participants gave their informed, prior consent and were either paid or received course credit for participating. Finally, the individuals depicted in the photographic images in this article (cf. Figures. 1, 2, 4) gave written consent (in conformity with the PLoS ONE guidelines) for this material to be published.

Apparatus and Materials. Participants performed the task seated in a dimly lit experimental cabin; looking at a 17" standard CRT monitor (100-Hz refresh rate, 1024 x 768 pixels screen resolution) positioned approximately 85 cm from their eyes. Stimulus presentation on the CRT was controlled by a Pentium IV PC using the E-Prime software (Psychology Software Tools, Pittsburgh, USA). The stimuli consisted of a series of color photographs taken for the purposes of this study – see Figure A1. The photographic images covered a screen area of 13.75° (width) x 10.35° (height) of visual angle; images were presented centrally, 6.7° from the screen borders.

Procedure and design. At the beginning of the experiment, participants received written instructions describing that a woman, April, would find herself in one of two situations: either a guest asks for something to drink, or a person who is living with her asks her to fetch fabric softener to do the laundry (Figure 1). Sometimes, there was no social situation and the context image was replaced by a picture of the sky with clouds, which meant that there was no specific task to be performed by April (Figure 2). Next, April goes to the kitchen (not explicitly presented in the trial sequence) where both potentially action-relevant items are to be found. Accordingly, the next image presented on the trial depicts a 'kitchen-counter' scene with two bottles, one positioned on the left and one on the right side, with a plastic cup next to each; one of the bottles contains orange juice (yellow), the other fabric softener (pink). Subsequently, the next image shows April standing between the bottles, gazing straight-ahead (in the direction of the observer).

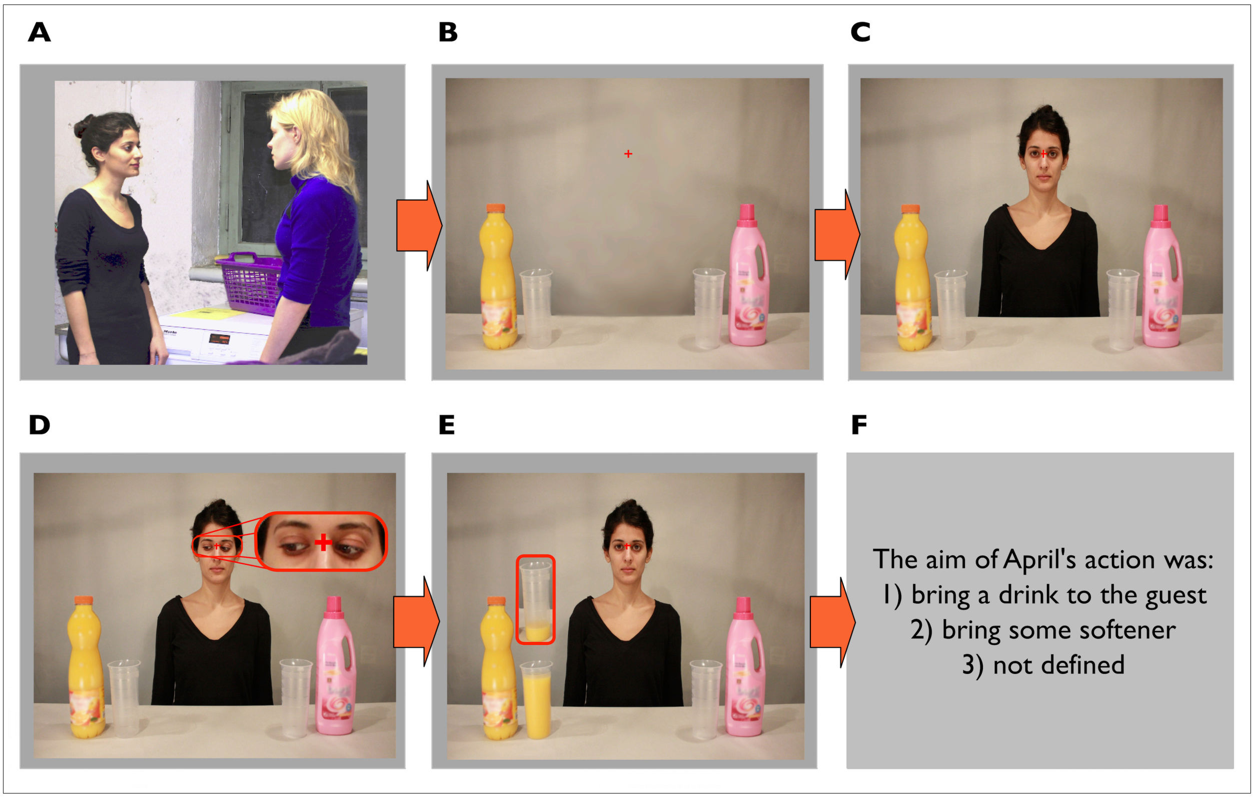

Figure 1. Schematic representation of an example trial in Experiment 1, depicting a 'laundry' context with an incongruent gaze direction and a validly cued target. Gaze direction in Frame D is zoomed-in only for the purpose of illustration. Frame E shows both the target with a low level of liquid as well as the target with high level of liquid. These targets are presented together only for illustration purposes; in the experiment proper, only one of the two targets was presented.

Figure 2. The two other action context images. The 'drink' context (left) and the 'neutral' context (right).

The next image shows April either making a gaze shift towards one of the bottles or maintaining straight-ahead gaze direction. In the final image, one of the cups (target) positioned next to one of the bottles appeared already containing a certain level of liquid – implying that

April, in the meantime, had poured liquid into it. The sequence of actually lifting the bottle and pouring the liquid was not shown, in order to prevent the introduction of additional directional cues (arm extension, body posture) over and above April's gaze direction. The participants' task was then to determine whether the level of liquid (in the target cup) was either low or high (target discrimination task).

The most important manipulation was that before the frame containing the target (Figure 1, D) the actor's gaze was averted to either the bottle congruent or the bottle incongruent with the action context. That is, her gaze could be directed to the orange juice (yellow) in the 'bring-a-drink' scenario, or to the softener (pink) in the 'laundry' scenario (congruent conditions); or her gaze could be directed to the softener in the 'drink' context, or orange juice in the 'laundry' context (incongruent conditions). In the *congruency-neutral condition*, the image presented at the start of the trial depicted a sky with clouds, rather than a social scene (see Figure 2). Therefore, although the actor's gaze was directed to one of the bottles, this had no relation to an action context (because there was no action context specified in the congruency-neutral condition). This condition was introduced as a baseline for the gaze-cueing effects. All three gaze-congruency conditions were distributed equally across the experiment.

Additionally, the gaze was either valid or invalid with respect to the target position (Figure 1, E). That is, the gaze direction could either coincide with the position at which the target would subsequently appear (valid trials) or not coincide (invalid trials). In the *neutral-validity* condition, the actor's gaze remained looking straight ahead. The neutral validity condition was introduced to test whether compatibility of the *target* itself with respect to the action scenario had an impact on performance (independently of gaze direction). All three validity conditions we distributed equally across the experiment.

All nine conditions were pseudo-randomized across trials; also, the side on which each bottle was presented, the target type (orange juice or softener), and the level of liquid (low or high) were pseudo-randomized across trials – yielding a total of 48 trials per condition. The total number of trials was (9 x 48 =) 432, presented in 6 blocks of 72 trials each; an experimental session, including training, took some 80 minutes to complete.

Individual trials consisted of the following sequence of images (cf. Figure 1): First, a fixation point appeared at the center of the screen for 1000 ms (Figure 1, A) Next, a centered context picture was presented together with an explanatory sentence (i.e., “Could you bring me a drink, please?” or “Could you fetch some softener, please?”) for 2500 ms; there were different images for the “drink”, “softener”, and “neutral” contexts – see Figure 2. (Figure 1, B) This was followed by a picture displaying a kitchen counter with two bottles on it: yellow orange juice and pink softener, on opposite sides, equidistant from the center (5.71°); an empty transparent plastic cup (“glass”) was located next to each bottle; this picture remained on the screen for 600 ms (Figure 1, C) April appeared between the two bottles looking straight ahead for 2000 ms (Figure 1, D) For another 600 ms, she looked to one of the sides, or she kept looking straight ahead (neutral validity trials). (Figure 1, E) An image was displayed (until response) in which April was presented again looking straight ahead, and which contained the response-relevant target: one of the glasses was partially filled with one of the two types of liquid to a high or a low level. Participants were asked to respond as fast and as accurately as possible by pressing the left mouse key for a low level and the right button for a high level of liquid. Finally, after the target response, an action context question was presented (Figure 1, F) with three possible response options: ‘the aim of April’s action was: bring a drink to the guest, bring some softener, not defined’. We introduced the action context question in 2/3 of trials to ensure that the participants correctly encoded and kept in mind the context throughout the trial. The response was given by pressing the 1, 2, or 3 key on a standard computer keyboard, with accuracy (rather than speed) being stressed. The location of the response alternatives was randomized for each trial. Feedback regarding accuracy was given (1000 ms) right after the action context response: the word “correct” or “incorrect” in the center of the screen. Consecutive trials were separated by an intertrial interval of 500 ms. Feedback about accuracy and reaction time in the target discrimination task for each entire block was provided in the breaks between blocks. Participants were asked to fixate in the middle of each frame and not move their eyes. They were explicitly informed that the direction of April’s gaze was not predictive with regard the location of the target.

Analysis. Consistent with our hypotheses, our main analysis focused on reaction times (RTs) in the target discrimination task as a function of gaze validity (valid, invalid) and gaze congruency with respect to action context (congruent, incongruent, neutral). RTs, measured as the time between target appearance and key press, were analyzed as follows: First, trials on which the action context question had not been queried were excluded from analysis, as well as trials on which the context probe was answered incorrectly (M = 6.66%, SD = 3.57). This was done to ensure that participants had actually attended the action scenario on the analyzed trials. Finally, trials with incorrect target responses (liquid-level discrimination) were eliminated (error rate M = 3.82%, SD = 2.64). Individual participants' median RTs for each condition were calculated and subjected to a 3 × 2 repeated-measures ANOVA with the factors congruency (congruent/ incongruent/ neutral) and validity (valid/invalid). In all analyses, degrees of freedom were adjusted according to Greenhouse-Geisser's procedure when the sphericity assumption was violated.

Results

Average median RTs and standard errors (in brackets) as well as error rates for each condition are presented in the Table 1.

TABLE 1: Group average RTs and error rates Experiment 1. Group average of individual median RTs and associated standard errors of the means (SEMs, in ms), and a group average of error rates, as a function of cue validity and gaze congruency.

	Validity			
Gaze Congruency	Valid		Invalid	
	RT	Error rate	RT	Error rate
Congruent	441 (15)	3.65%	480 (18)	4.25%
Incongruent	447 (15)	2.52%	462 (14)	3.91%
Neutral[a]	446 (14)	3.30%	464 (16)	3.73%

[a] 'Neutral' refers to the 'neutral congruency' condition, in which observers were present with a neutral context image; in this condition, the observed gaze shift could still validly or invalidly cue the location of the target.

Reaction Times. The ANOVA on median RTs with the factors congruency (congruent, incongruent, neutral) and validity (valid, invalid) yielded a significant main effect of validity [F (1, 23) = 31.698, $p = .00001$, $\eta_p^2 = .580$]: discrimination RTs were shorter with valid (M = 444 ms; SEM = 14.6) than with invalid gaze cues (M = 468 ms; SEM = 15.94). The main effect of gaze congruency was not significant [F (2, 46) = .861, p = .429, η_p^2 = .036]. Importantly, the congruency × validity interaction was significant [F (2, 46) = 4.439, p = .017, η_p^2 = .162]. Although the validity effect was reliable in all three congruency conditions (valid vs. invalid for congruent gaze: ΔRT = 38.98 ms, t (23)= 5.028 , 95% CI [22.94, 55.01],p = .00004, d_z = 1.02; for incongruent gaze: ΔRT = 14.66 ms, t (23)= 2.932, p = .007, 95% CI [4.32, 25.01], d_z = .60; and for neutral gaze: ΔRT= 18.16 ms, t (23)= 2.651, p = .014, 95% CI [3.99, 32.34], d_z = .54), it was larger in the congruent relative to the incongruent and the neutral condition. To assess the differences in gaze cueing effects as a function of gaze congruency, we calculated gaze-cueing effects ($\Delta RT = M\ RT_{invalid} - M\ RT_{valid}$) and subjected them to planned comparisons across the three congruency conditions; see Figure 3.

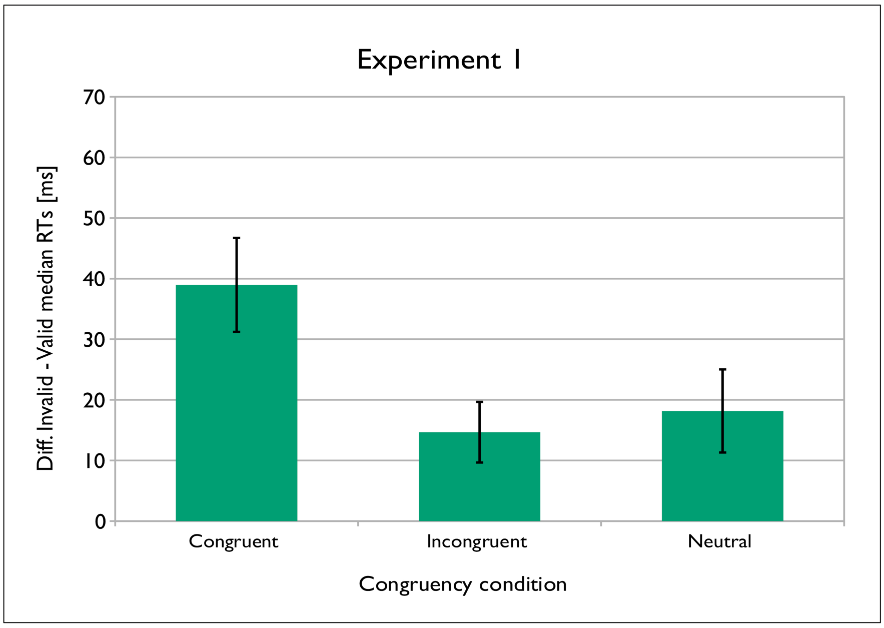

Figure 3. Gaze-cueing effect as a function of gaze congruency in Experiment 1. Error bars represent the confidence intervals (95% CIs) adapted for within-participants designs according to Cousineau's (2005) procedure.

The planned comparisons (two-tailed t-tests, Bonferroni-corrected) of the gaze-cueing effects revealed ΔRT to be significantly larger only for the congruent vs. incongruent conditions [t (23)= 3.153, p = .013, 95% CI [8.35, 40.26], d_z = .64], but not the congruent vs. neutral conditions [t (23)= 2.090, p = .144, 95% CI [.20, 41.41], d_z = .42]; the small difference between the neutral and incongruent conditions was not reliable [*t* (23)= .405, *p* = 1.00, 95% CI [-21.39, 14.39], d_z = .08], see Figure 3.

Note that we also examined whether compatibility of the target itself with the action context (e.g., orange juice in the 'drinking' scenario vs. orange juice in the 'softener scenario') influenced performance. For this purpose, we analyzed separately trials on which April's gaze remained straight-ahead. We compared the target compatible-, incompatible-, and neutral-context conditions for gaze-neutral (straight-ahead) trials. The ANOVA yielded no significant target-compatibility effect [*F* (2, 46) = 1.587, *p* = .216, η_p^2 = .065]: compatible (M = 503 ms; SEM = 18.48), incompatible (M = 514 ms; SEM = 19.25), and neutral (M = 514 ms; SEM = 19.41). That is, in the gaze-neutral condition, target discrimination responses were not significantly affected by the compatibility of the target with respect to the action context.

Error rates. Participants' error rates in the experiment were low overall (M = 3.56%, SD = 3.62). An ANOVA conducted on the error rates (analogous in design to the RT ANOVA) revealed none of the effects to be significant, all *p*s > .1. Nevertheless, in order to examine for potential speed-accuracy trade-offs (SATOs), we carried out an ANOVA analogous to the analysis on median RTs using so-called 'inverse efficiency scores' (IES) [(Townsend & Ashby, 1978; 1983). IE scores are calculated by dividing individual (in our case: median) RTs for a particular experimental condition by an index of response accuracy: RT/(1 – p(E)), where p(E) is the error probability. That is, the RT value is increased the more the lower the accuracy associated with responses in this condition, effectively correcting for a speed-accuracy trade-off. Thus, examining the IE scores was designed to establish whether the pattern of RT results would remain unchanged when correcting for response errors. This analysis revealed essentially the same pattern of effects as the RT ANOVA: a significant validity effect, F (2, 23) = 34.130, p < .001, ηp2 = .597, and a marginally significant interaction, F (2, 46) = 3.172, p = .051, ηp2 = .121. That is, the pattern of RT results is reasonably robust, holding up even if when taking SATO effects into account.

Discussion Experiment 1

The aim of Experiment 1 was to examine whether gaze cueing is affected by expectations regarding others' gaze behavior in complex action sequences. To this end, we embedded a gaze-cueing protocol within an action context. The question of interest was whether participants' performance of a discrimination task would be affected by the actor gazing (or not gazing) at the location of the subsequently appearing target object (gaze-cueing effect); and if the gaze-cueing effect would be influenced by whether the actor's gaze behavior was in line with participants' expectations induced by the action context in which the actor was embedded. The results revealed a main effect of validity, indicating that participants followed the observed agent's gaze even though it was uninformative with respect to the position of the target. This is consistent with the idea that attentional orienting to gaze direction cues relies, to some extent, on a reflexive mechanism (e.g., Hietanen, 1999, Friesen & Kingstone, 1998). Importantly, however, the gaze-cueing effects were significantly modulated by the congruency of the gaze shift with respect to the action context: the gaze-cueing effect was larger when the observed agent directed her gaze to the context-congruent object, as compared to the context-incongruent object. In other words, the congruency of the observed agent's gaze direction with respect to the action context played a significant role in the extent to which the agent's gaze was followed: when the agent's gaze shift confirmed participants' prior expectations as to the object she would gaze at (in accordance with the action context), her gaze was followed to a larger extent compared to the action-incongruent gaze-shift condition.

In sum, Experiment 1 indicates that expectations regarding gaze behavior with respect to action plans influence the degree to which the gaze of an observed actor is followed. However, Experiment 1 did not permit us to determine whether the observed effects were indeed due to covert attention, or rather to overt attention. That is, alternatively, the slower responses in the invalid conditions might simply be due to participants having made (overt) saccades to the side opposite to the target location (despite the instruction to maintain fixation), rather than being attributable to violation of expectations (as assumed in the interpretation). Given this, the aim of Experiment 2 was to decide this issue by replicating the results of Experiment 1 while monitoring participants' eye movements.

2.4 Experiment 2

Experiment 2 was designed to isolate the influence of gaze direction cues and expectations regarding the observed gaze behavior on *covert* attentional orienting in a naturalistic scene. The paradigm and procedure were essentially the same as in Experiment 1, with the addition of monitoring participants' eye fixation during critical frames of the trials using eye tracking. Also, to reinforce the instruction and make it easier for participants to maintain fixation during presentation of the critical stimulus frames (Figure 4, Frames A-E), a fixation cross was presented and remained on screen throughout frames A-E (see Figure 4). If participants' gaze deviated from the fixation cross by more than ±2° of visual angle (Figure 4, red dotted circle), the trial was aborted and repeated at the end of the block. This ensured that any effects obtained in Experiment 2 would not be attributable to shifts of overt attention. Note that the monitoring of participants' eye movements in a gaze-cueing paradigm embedded in a naturalistic action scenario is a novel feature. While studies of simple, non-naturalistic scenarios have shown that observing gaze shifts evokes both covert and overt orienting (e.g., Mansfield, Farroni, & Johnson, 2003; Ricciardelli, Bricolo, Aglioti, & Chelazzi 2002; Friesen & Kingstone, 2003), to our knowledge, no studies on orienting of covert attention in response to gaze cues in complex natural scenes have monitored eye movements.

Method

Participants. In order to determine the sufficient sample size for Experiment 2, we conducted an a-priori power analysis for the effect of congruency (congruent vs. incongruent) on the size of gaze cueing, based on: (a) the effect size of Experiment 1 (d_z = .64); (b) an α-error equal to .05; and (c) a recommended power level of .80. This analysis yielded an adequate sample size of 22 participants. Accordingly, 22 healthy volunteers took part in Experiment 2 (age range 21–31 years, M = 24.8 years; 18 women; all right-handed), receiving monetary compensation or course credits for their participation. All participants had normal or corrected-to-normal vision, reported normal color vision, and provided written consent regarding participation in the experiment. None of the participants had taken part in Experiment 1 or in any other experiment with a similar design. Note that only participants who were able to maintain fixation during practice session (see Procedure section below) were admitted to the proper experiment.

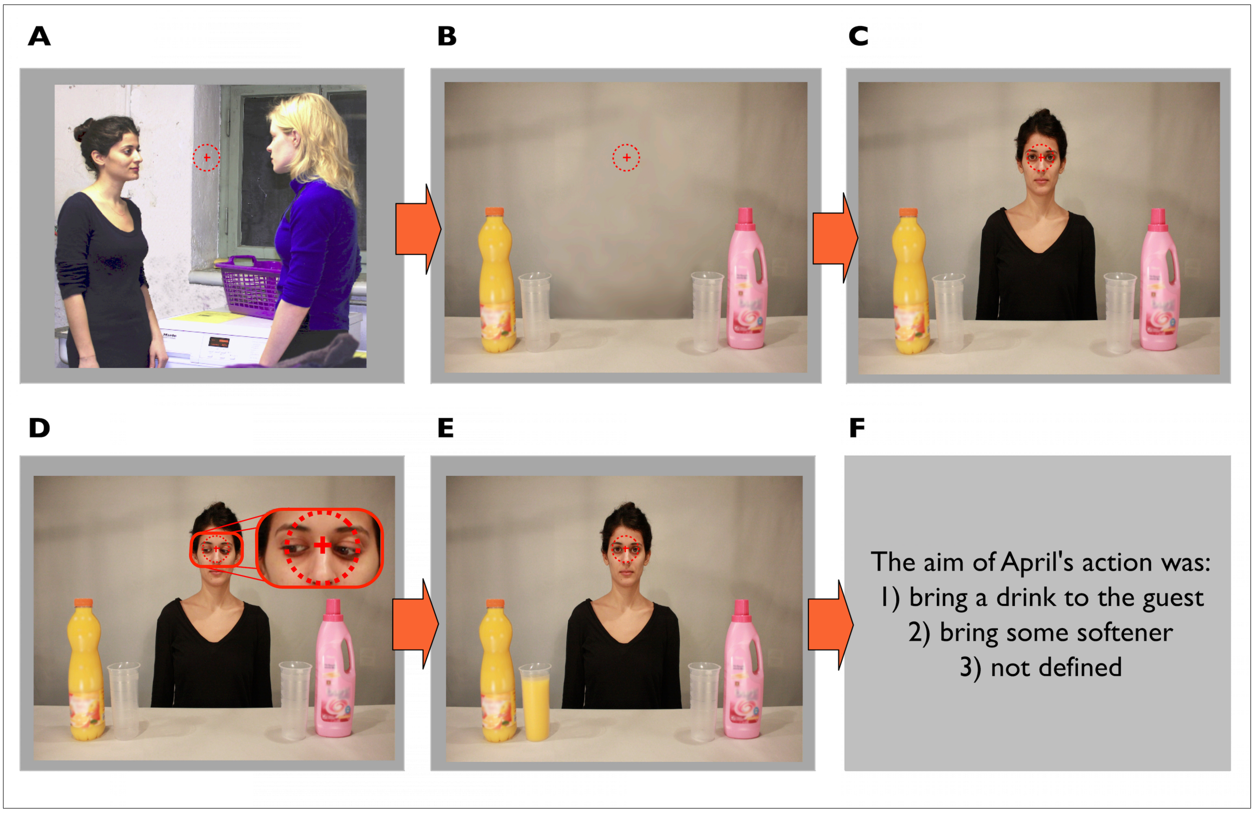

Figure 4. Schematic representation of an example trial in Experiment 2, with the same sequence of image frames as in Experiment 1 (Figure 1). The main difference between the experiments was that a fixation marker was presented throughout the image sequence. Participants' eye movements were restricted to a circular area (represented by the dotted circle) of a radius of 2° around the (center of the) fixation cross.

Apparatus and Procedure. The apparatus and procedure (as well as the design) were essentially the same as in Experiment 1. However, there were a number of differences relating to the monitoring of participants' eye movements. Participants' head position and thus their eye-to-screen distance was 'stabilized' by means of a desk-mounted chin-and-headrest device positioned 60 cm in front of the CRT monitor. Eye movements were recorded monocularly (right eye) using an Eyelink 1000 tower-mounted eye-tracking system (SR-research Inc.; sampling rate: 2000 Hz; monocular accuracy: 0.25°–0.5°; resolution: 0.01°RMS (root mean square), related to the absolute sensor performance, the smaller the better). The apparatus was controlled using PsychToolbox 3 (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007) based on Matlab 2008a. Stimulus presentation on the screen was controlled by an Apple Mac Mini 2.3 using the same Matlab software. The images presented covered a screen area of 23.53° (width) x 17.06° (height) of visual angle, and each image was presented centrally, 7.9° from the borders of the screen. Note that while the images were of the same screen size as in Experiment 1, their perceived size (in terms of degrees of visual angle) was larger in Experiment 2 owing to a reduction of the eye-to-screen distance (according to the EyeLink user manual, the camera and illuminator should be placed at a distance of 40 to 70 cm from the observer, with the ideal tracking distance being 50 to 55 cm; SR Research Ltd. Mississauga, Ontario, Canada, 2005-2008). In order to facilitate maintenance of fixation, the sentences that had been presented under the context image in Experiment 1 were eliminated in Experiment 2.

After receiving written instructions, participants performed two practice blocks without eye tracking, and two additional blocks with gaze monitored by the eye tracker. The eye tracker was calibrated using a 13-point calibration procedure, which was immediately followed by a validation procedure. Calibrations were accepted if the mean error was less than 1.5°. In the experiment proper, when participants' gaze diverted from central fixation by more than 2°, the trial was aborted and repeated at the end of the block (a trial block was finished only once participants had correctly completed all 24 trials within a block). In all other respects, the procedure was similar to Experiment 1, except that (i) the neutral-gaze condition was dropped (after having confirmed that target congruency per se did not affect participants' responses [Experiment 1], there was no need to include the neutral gaze condition in the design of Experiment 2; thus, including a neutral condition in Experiment 2 would not have yielded any benefits for the design); and (ii) all trials included a question at the end regarding the action context (rather than only 66% of trials, as in Experiment 1). This was done in order to ensure that participants would encode and maintain in memory the action context under conditions that were more demanding than in Experiment 1 (in Experiment 2, participants were instructed to maintain fixation on the fixation cross for an extended period of time, which can be considered an additional task). Finally, (iii) trials with the 'drink' and 'laundry' scenarios had a second probe question (presented after the action context probe) regarding the correctness of the liquid with respect to the action context: “Did April take the correct liquid?” (Answers: number key 1 = yes, number key 2 = no); this additional question was introduced to reinforce the relevance of the action context to the entire task.

All six conditions were pseudo-randomized across trials; also, the side on which each bottle was presented, the target type (orange juice or softener), and the level of liquid (low or high) were pseudo-randomized across trials – yielding a total of 48 trials per condition (as in Experiment 1). The total number of trials was (6 x 48 =) 288, presented in 12 blocks of 24 trials each; an experimental session, including training, took some 60 minutes to complete. Feedback about accuracy and reaction time in the target discrimination task for each entire block was provided in the breaks between blocks. Participants were asked to fixate in the middle of each frame and not move their eyes and were explicitly informed that the direction of April’s gaze was not predictive with regard the location of the target.

Analysis. The same preprocessing steps and criteria for exclusion of trials were used as in Experiment 1. All participants maintained error rates lower than 15% in the action-context responses (M = 4.53%, SD = 3.19%). Median RTs were calculated for each participant and each condition and were subjected to a 3 x 2 repeated-measures ANOVA with the factors gaze congruency (congruent, incongruent, neutral) and validity (valid, invalid).

Results

Average median RTs and standard errors (in brackets) as well as error rates for each condition are presented in the Table 2.

Reaction Times. The 3 x 2 repeated-measures ANOVA of the median RTs with the factors congruency (congruent, incongruent, neutral) and validity (valid, invalid) revealed both main effects to be significant. The congruency effect [F (2, 42) = 4.058, p = .024, η_p^2 = .162] was due to RTs being faster in the neutral condition (M = 529 ms, SEM = 20 ms) as compared to the congruent and incongruent conditions (M = 545 ms, SEM = 24.17 ms; and, M = 549 ms, SEM = 23 ms, respectively), though planned comparisons (two-tailed t-test) revealed only the difference between the neutral and incongruent conditions to be significant [t(21)= 2.964, p = .007, d_z

= .63]. The effect of validity [*F* (1,21) = 28.307, *p* = .00003, η_p^2 = .574] was owing to RTs being faster with valid (M = 525 ms, SEM = 21.27 ms) than with invalid gaze cues (M = 557 ms, SEM = 23.17 ms).

TABLE 2. Group average RTs and error rates Experiment 2. Group average of individual median RTs and associated standard errors of the means (SEMs, in ms), and a group average of error rates, as a function of cue validity and gaze congruency.

Gaze Congruency	Validity			
	Valid		Invalid	
	RT	Error rate	RT	Error rate
Congruent	515 (24)	12.59%	574 (25)	12.97%
Incongruent	543 (23)	13.26%	554 (24)	13.35%
Neutral[a]	515 (19)	12.03%	543 (22)	12.03%

[a] 'Neutral' refers to the 'neutral congruency' condition, in which observers were present with a neutral context image; in this condition, the observed gaze shift could still validly or invalidly cue the location of the target.

Similarly to the Experiment 1, and importantly for the purposes of the study, the congruency x validity interaction was significant [*F* (2, 42) = 11.875, *p* = .000082, η_p^2 = .361], with the validity effect being significant only in the congruent and neutral gaze conditions, but not in the incongruent condition (valid vs. invalid for congruent gaze: ΔRT= 58.37 ms, *t* (21)= 6.803, *p* = .000001, 95% CI [40.53, 76.21], d_z = 1.44; for incongruent gaze: ΔRT= 11.65 ms, t (21)= 1.651, 95% CI [-3.03, 26.33], *p* = .114, d_z = 0.33; and for neutral gaze: ΔRT= 27.42 ms, *t* (21)= 3.001, *p* = .007, 95% CI [8.42, 46.43], d_z = .63). Planned comparisons (two-tailed t-test, Bonferroni-corrected) of the gaze-cueing effects ($\Delta RT = M\ RT_{invalid} - M\ RT_{valid}$) revealed a significant difference in ΔRT between the congruent and incongruent gaze conditions [*t* (21)= 5.080, *p* = .00005, 95% CI [27.6, 65.84], d_z = 1.08] and between the congruent and neutral conditions, [*t* (21)= 3.272, *p* = .011, 95% CI [11.27, 50.61], d_z = 0.70], but no difference between the incongruent and neutral conditions [*t* (21)= 1.495, *p* = .450, 95% CI [-6.17, 37.73], d_z = .32]. In other words, the gaze-cueing effect was enhanced in the gaze-congruent condition as compared to the gaze-incongruent and gaze-neutral conditions (see Figure 5).

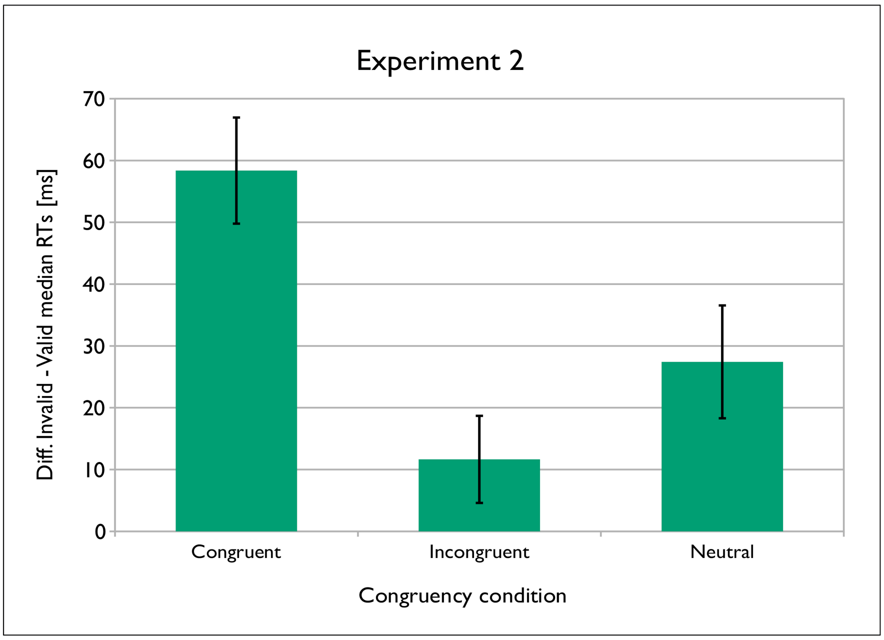

Figure 5. Gaze-cueing effect as a function of gaze congruency in Experiment 2. Error bars represent the confidence intervals (95% CIs) adapted for within-participants designs according to Cousineau's (2005) procedure.

Error rates. An ANOVA conducted on the error rates (analogous in design to the RT ANOVAs) revealed no significant effects (*ps* > .75, except for the main effect of congruency). Only the main effect of congruency approached significance [F (2, 42) = 2.870, p = .068, η_p^2 = .120], with error rates being slightly larger for the incongruent (M = 13.3%, SEM = 3%) as compared to the congruent and neutral conditions (M = 12.78%, SEM = 2.9% and M = 12.1%, SEM = 2.8%, respectively). Post-hoc comparisons (two-tailed t-test, Bonferroni corrected) revealed no significant difference between the three congruency conditions (*ps* > .118). Note that the congruency effect on the error rates is of the same direction, and thus reinforces, the effect obtained in the RTs; that is, it is not indicative of a speed-accuracy trade-off. Similarly to Experiment 1, we conducted an ANOVA on inverse efficiency scores (IES) to account for potential speed-accuracy trade-off (SATO). The analysis showed an identical pattern of results as those obtained with uncorrected median RTs: a significant validity effect, F (1, 21) = 23.481, p < .001, ηp2 = .528; a significant congruency effect, F (1.528, 32.095) = 7.813, p = .004, ηp2 = .271; and a significant interaction, F (2, 42) = 5.748, p = .006, ηp2 = .215 –again confirming that the pattern of RT effects cannot be explained by SATO influences.

Comparison across Experiments 1 and 2

To examine whether the congruency factor had a similar influence on the effects of gaze cueing in both experiments, we conducted a mixed-design ANOVA on the gaze-cueing effects with *congruency* as within-participants factor and *experiment* as between-participants factor. This ANOVA yielded a significant main effect of gaze congruency [F (2, 88) = 15.681, p = .000001, η_p^2 = .263], but no significant interaction between gaze congruency and experiment (p= .237). Post-hoc comparisons (Bonferroni corrected) revealed significant differences in cueing effects between both the congruent and incongruent (p= .000001) and the congruent and neutral conditions (p=.002). Furthermore, independent-samples t-tests (with the assumption of homogeneity of variances ensured via Levene's F test: p = .972, p = .233, and p = .150 for the gaze-congruent, incongruent, and neutral conditions, respectively) revealed no significant differences in the gaze-cueing effects between experiments for any of the congruency conditions (p = .100, p = .726, and p = .417 for the gaze-congruent, incongruent, and neutral conditions respectively). Thus, the two experiments yielded same pattern of results regarding the congruency factor.

Discussion Experiment 2

To examine whether the findings of Experiment 1 were due to covert attention (rather than eye movement artifacts), Experiment 2 used essentially the same design as Experiment 1, including however the monitoring of participants' eye fixation (trials with eye position shifts deviating by more than 2° from the fixation marker were excluded). Importantly, Experiment 2 replicated the results of Experiment 1 in all critical respects (in fact, without there being any significant differences in the cueing effects between the two experiments; see Comparison across Experiments 1 and 2), thus ruling out that the effect pattern was owing to overt, oculomotor orienting responses to the observed gaze shift. As in Experiment 1, the main effect of gaze cueing was significant, even with participants fixating the fixation cross throughout the critical trial frames A-E (Figure 4). This verifies that participants covertly attended the object that the actor gazed at, even though the actor's gaze direction was spatially uninformative with respect to the target object. Likewise, the modulation of the gaze-cueing effect by the congruency of the actor's gaze (the gazed-at object) with the action context cannot be attributed to a congruency-dependent modulation of overt eye movements, but instead reflects a modulation of covert attentional orienting.

Importantly, no gaze-cueing effect was obtained in Experiment 2 when the actor's gaze was incongruent with the action context. That is, the facilitation normally engendered by gaze cues was eliminated when expectations regarding which object would be gazed at (given the overarching action context) were violated. Furthermore, the congruent gaze condition yielded stronger gaze-cueing effects relative to the neutral action-context baseline. Restated, relative to the baseline, the gaze-cueing effect was actually enhanced when the actor's gaze 'complied' with observers' expectations regarding the object that would be gazed at in order to achieve the ultimate action goal.

2.5 General Discussion

The present study was designed to examine – using a naturalistic scenario – how the gaze direction of an observed actor and the observer's expectations, induced by the action context, would influence the gaze-cueing effects. The results of both experiments showed that task performance (i.e., discrimination of the liquid level in the target) depended on gaze-cue validity with respect to the target side, with attention following the gaze direction of the observed agent ('April') – the typical gaze-cueing effect. Importantly for the purposes of this study, the gaze-cueing effect was modulated by whether the actor had gazed at an action-congruent or at an incongruent object: specifically, the cueing effect was significantly enhanced when the actor gazed at an object congruent with the action context, relative to an incongruent object (Experiments 1 and 2), and it was entirely eliminated when the actor gazed at an object incongruent with the action context, relative to a congruent object (Experiment 2).

Consequently, gaze cueing appeared to be modulated by participants' expectations with regard to the gaze behavior in the context of upcoming action steps within an overarching action sequence. This is in line with the notion that humans activate a certain action schema (Manthey, Schubotz, & von Cramon, 2002) when observing others in action and (possibly implicitly) expect the gaze of the observed agent to precede (and, thus, provide a pointer to) successive action steps (Flanagan & Johansson, 2003; Hayhoe et al., 2003). Hence, when the observed gaze behavior confirms participants' expectations concerning the action sequence, the gaze cueing effects seem to be enhanced relative to when the gaze behavior violates the expectations. Moreover, gaze-cueing effects can even be entirely suppressed when the gaze behavior violates action-related expectations. The complete lack of a validity effect seen in Experiment 2 when the actor looked at an object that was action-incongruent further attests to the impact of top-down control on bottom-up-driven attentional orienting in response to gaze direction.

This set of findings is in line with a wide body of literature suggesting that higher-order cognitive mechanisms can modulate orienting of attention in response to spatial cues (e.g., Posner & Cohen, 1984; Jonides, 1981; Müller & Rabbitt, 1989). Teufel et al. (2010a), Wiese et al. (2012), and Wykowska et al. (2014) extended this evidence to gaze following, demonstrating that gaze-cueing effects can be reduced (Wiese et al., 2012) or even eliminated (Wykowska et al., 2014) by top-down regulation of the mechanisms involved. In particular, Wiese et al. (2012) found that the context in which a gaze shift is performed – in their study: the presence versus absence of physical reference objects to which the gaze would refer – modulates the degree to which attentional resources are deployed to the cued location. On this basis, they proposed that this modulation is mediated via a top-down mechanism which binds the gaze shift (in central vision) to a referred-to object (in peripheral vision). Subsequently, Wiese, Wykowska, & Müller (2014) extended the notion of 'context' to also include social factors, such as knowledge of the reliability of the cue provider, to account for their finding that gaze following was modulated by whether participants perceived the gaze behavior displayed by the observed gazer as reliable and highly predictive or not. Applied to the results of the present experiment, the top-down component can engender both an enhancement of the gaze-cueing effect (namely, when the actor's gaze is seen to be shifted to the action-congruent object) and a suppression of the default, presumably 'reflexive', gaze following (when the actor's gaze is shifted to an action-incongruent object).

Finally, the present results extend earlier findings in that they suggest a link between action prediction and gaze following. This is in accordance with neuroimaging results showing activation of the superior temporal sulcus (STS) during action prediction (Saxe, Xiao, Kovacs, Perrett, & Kanwisher, 2004) and updating predictions after a violation of an expected action sequence (Sebanz, Bekkering, & Knoblich, 2006). Interestingly, the STS region is also involved in gaze-direction detection and gaze cueing (Haxby, Hoffman, & Gobbini, 2000). Hence, also at the neural level, following gaze seems to be closely linked to anticipatory mechanisms in action observation. In more general terms, the close link between gaze following and action prediction seems to be very adaptive: In order to interact with others, we need to know what the others are going to do next (Frith & Frith, 2006). As people tend to look at objects they are planning to manipulate, gaze direction is informative about the identity (*what?*) and spatial location of attended objects (*where?*). Together with knowledge about people's preferences (*who?*), which can be acquired directly by interacting with them or indirectly by either observing someone interacting with another person or receiving information about a person, we make inferences about their internal states (*why?*) (Sebanz & Knoblich, 2009) –so as to predict which action is most likely to be performed next under the given circumstances. Thus, in our paradigm, knowing that April was asked to bring a glass of orange juice to a guest, one would predict that she would be looking for an empty cup and a bottle of orange juice (in order to pour some juice and bring the filled glass to the guest). Therefore, through linking processing of gaze direction with action prediction within a single paradigm, the current study demonstrates that spatial information derived from gaze direction and context information about the action goal can be combined in order to predict consecutive steps in an action sequence. Taken together, the present findings have implications concerning the actual function of the gaze-following mechanism and the role it plays in natural daily-life scenarios: arguably, gaze following has developed not just to pick up signals that others convey regarding potentially relevant events in the environment, but also, and conceivably foremost, to enable us to infer what others are going to do next (Cohen, 1995; Frith & Frith, 2006).

Concerning methodological implications, using a sequence of naturalistic photographs brought our paradigm closer to more ecologically valid, real-life social scenarios. Admittedly, though, it is still a rather artificial protocol in which participants are just observing a series of static images. Arguably, however, this step needed to be made between entirely artificial stimuli and completely realistic protocols. Having a sequence of images allowed us to maintain experimental control over factors of interest and to circumvent certain confounds (such as involuntary attentional capture or motion-related effects). With the present paradigm providing a first step into more naturalistic scenarios, future research should take the design even closer to real life (e.g., by using video or virtual-reality technology) and to a more interactive protocol (rather than merely involving an observational stance). We contend that when the protocol is made more naturalistic, the effects observed in the current study might turn out even stronger.

Conclusion

The present study is, to our knowledge, the first to examine how attention is deployed within a naturalistic visual scene as a result of the gaze direction displayed by an observed actor and the observer’s expectations regarding the unfolding of the action sequence in a complex action scenario. With the use of a novel paradigm that uses naturalistic images, we show that the gaze-cueing effects, reflecting covert shifts of attention, can be modulated –either enhanced or (even entirely) suppressed– dependent on whether the gaze behavior of the observed agent (gazing at action-congruent or incongruent objects) does or does not fit with the expectations that participants hold with regard to the unfolding action sequence. In summary, our findings indicate that one of the key functions of gaze following is to monitor and predict others’ actions.

Acknowledgements

We thank Ebru Baykara and Nona Gergova for help with data collection, and Ella Bahry and Lee Simon-Vermot for their help with developing the photographic images used in this study. We also thank Marius 't Hart for his tremendous support with programming Experiment 2.

2.1.6 References

Baron Cohen, S. (1995). *Mindblindness. Learning, development, and conceptual change.*

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: evidence from very high functioning adults with autism or asperger syndrome. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *38*(7), 813–822. http://doi.org/10.1111/j.1469-7610.1997.tb01599.x

Baron-cohen, S., Baron-cohen, S., Centre, A. R., & Centre, A. R. (2005). The Empathizing System: a revision of the 1994 model of the Mindreading System. *Mind*, 1–44.

Bayliss, A. P., Schuch, S., & Tipper, S. P. (2010). Gaze cueing elicited by emotional faces is influenced by affective context. *Visual Cognition*, *18*(8), 1214–1232. http://doi.org/10.1080/13506285.2010.484657

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. http://doi.org/10.1037/0033-2909.112.1.155

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*(1), 42–45. http://doi.org/no DOI found

de C. Hamilton, A. F., & Grafton, S. T. (2007). The motor hierarchy: from kinematics to goals and intentions. *Sensorimotor Foundations of Higher Cognition*, pp. 381. http://doi.org/DOI:10.1093/acprof:oso/9780199231447.003.0018

Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze Perception Triggers Reflexive Visuospatial Orienting. *Visual Cognition*, *6*(5), 509–540. http://doi.org/10.1080/135062899394920

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience and Biobehavioral Reviews*, *24*(6), 581–604. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10940436

Flanagan, J. R., & Johansson, R. S. (2003). Action plans used in action observation. *Nature*, *424*(6950), 769–771. http://doi.org/10.1038/nature01861

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review*, *5*(3), 490–495. http://doi.org/10.3758/BF03208827

Friesen, C. K., & Kingstone, A. (2003). Covert and overt orienting to gaze direction cues and the effects of fixation offset. *Neuroreport*, *14*(3), 489–493. http://doi.org/10.1097/00001756-200303030-00039

Frith, C. D., & Frith, U. (2006). How we predict what other people are going to do. *Brain Research*, *1079*(1), 36–46. http://doi.org/10.1016/j.brainres.2005.12.126

Haxby, J., Hoffman, E., & Gobbini, M. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*(6), 223–233. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10827445

Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision*, *3*(1), 49–63. http://doi.org/10.1167/3.1.6

Hietanen, J. K. (1999). *Does your gaze direction and head orientation shift my visual attention? Neuroreport* (Vol. 10). http://doi.org/10.1097/00001756-199911080-00033

Hoffman, E. A., & Haxby, J. V. (2000). *Distinct representations of eye gaze and identity in the distributed human neural system for face perception. Nature neuroscience* (Vol. 3). http://doi.org/10.1038/71152

Itier, R. J., & Batty, M. (2009). Neural bases of eye and gaze processing: the core of social cognition. *Neuroscience and Biobehavioral Reviews*, *33*(6), 843–63. http://doi.org/10.1016/j.neubiorev.2009.02.004

Kanwisher, N. G. (2000). Domain specificity in face perception. *Nature Neuroscience*, *3*(8), 759–763. http://doi.org/10.1038/77664

Kilner, J. M., Vargas, C., Duval, S., Blakemore, S.-J., & Sirigu, A. (2004). Motor activation prior to observation of a predicted movement. *Nature Neuroscience*, *7*(12), 1299–1301. http://doi.org/10.1038/nn1355

Langton, S. R. H., & Bruce, V. (1999). Reflexive Visual Orienting in Response to the Social Attention of Others. *Visual Cognition*. http://doi.org/10.1080/135062899394939

Mansfield, E., Farroni, T., & Johnson, M. (2003). Does gaze perception facilitate overt orienting? *Visual Cognition*. http://doi.org/10.1080/713756671

Manthey, S., Schubotz, R. I., & von Cramon, D. Y. (2003). Premotor cortex in observing erroneous action: an fMRI study. *Brain Research. Cognitive Brain Research*, *15*(3), 296–307. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12527103

Ricciardelli, P., Bricolo, E., Aglioti, S. M., & Chelazzi, L. (2002). My eyes want to look where your eyes are looking: exploring the tendency to imitate another individual's gaze. *Neuroreport*, *13*(17), 2259–2264. http://doi.org/10.1097/00001756-200212030-00018

Ristic, J., & Kingstone, A. (2005). Taking control of reflexive social attention. *Cognition*, *94*(3), B55–65. http://doi.org/10.1016/j.cognition.2004.04.005

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews. Neuroscience*, *2*(9), 661–670. http://doi.org/10.1038/35090060

Saxe, R., Xiao, D. K., Kovacs, G., Perrett, D. I., & Kanwisher, N. (2004). A region of right posterior superior temporal sulcus responds to observed intentional actions. *Neuropsychologia*, *42*(11), 1435–1446. http://doi.org/10.1016/j.neuropsychologia.2004.04.015

Schubotz, R. I., & von Cramon, D. Y. (2002). Predicting perceptual events activates corresponding motor schemes in lateral premotor cortex: an fMRI study. *NeuroImage*, *15*(4), 787–96. http://doi.org/10.1006/nimg.2001.1043

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in Cognitive Sciences*, *10*(2), 70–6. http://doi.org/10.1016/j.tics.2005.12.009

Sebanz, N., & Knoblich, G. (2009). Prediction in Joint Action: What, When, and Where. *Topics in Cognitive Science*, *1*(2), 353–367. http://doi.org/10.1111/j.1756-8765.2009.01024.x

Teufel, C., Alexis, D. M., Clayton, N. S., & Davis, G. (2010). Mental-state attribution drives rapid, reflexive gaze following. *Attention, Perception & Psychophysics*, *72*(3), 695–705. http://doi.org/10.3758/APP.72.3.695

Teufel, C., Fletcher, P. C., & Davis, G. (2010). Seeing other minds: attributed mental states influence perception. *Trends in Cognitive Sciences*, *14*(8), 376–82. http://doi.org/10.1016/j.tics.2010.05.005

Vuilleumier, P., George, N., Lister, V., Armony, J., & Driver, J. (2005). Effects of perceived mutual gaze and gender on face processing and recognition memory, *12*(1), 85–102. http://doi.org/10.1080/13506280444000120

Wiese, E., Wykowska, A., & Müller, H. J. (2014). What we observe is biased by what other people tell us: Beliefs about the reliability of gaze behavior modulate attentional orienting to gaze cues. *PLoS ONE*, *9*(4). http://doi.org/10.1371/journal.pone.0094529

Wiese, E., Wykowska, A., Zwickel, J., & Müller, H. J. (2012). I See What You Mean: How Attentional Selection Is Shaped by Ascribing Intentions to Others. *PLoS ONE*, *7*(9). http://doi.org/10.1371/journal.pone.0045391

Wurm, M. F., & Schubotz, R. I. (2012). Squeezing lemons in the bathroom: contextual information modulates action recognition. *NeuroImage*, *59*(2), 1551–9. http://doi.org/10.1016/j.neuroimage.2011.08.038

Wykowska, A., Wiese, E., Prosser, A., & Müller, H. J. (2014). Beliefs about the minds of others influence how we process sensory information. *PloS One*, *9*(4), e94339. http://doi.org/10.1371/journal.pone.0094339

3. Study 2

Expectations regarding action sequences modulate electrophysiological correlates of the gaze-cueing effect

Jairo Pérez-Osorio[1,4], Hermann J. Müller[1,2], & Agnieszka Wykowska[3,4]

[1] Department of Psychology, Ludwig Maximilian University, Munich Germany.

[2] Department of Psychological Sciences, Birkbeck College, University of London, London, UK

[3] Chair for Cognitive Systems, Technische Universität München, Munich Germany

[4] Division of Human Work Sciences, Luleå University of Technology, Luleå, Sweden

Author contributions: AW JP HJM, Conceived and designed the experiments. JP, Performed the experiments. JP AW, Analyzed the data. JP AW HJM, Wrote the paper.

Perez-Osorio, J., Müller, H.J., & Wykowska, A. (under review) “Expectations regarding action sequences modulate electrophysiological correlates of the gaze-cueing effect”. *Psychophysiology*.

3.1 Abstract

Predictive mechanisms of the brain are important for social cognition, as they enable inferences about others' goals and intentions, thereby allowing for generation of expectations regarding what will happen next in the social environment. Therefore, attentional selection is modulated by expectations regarding behavior of others (Perez-Osorio, Müller, Wiese, & Wykowska, 2015). In this paper we examined –using the event-related potentials (ERPs) of the EEG signal– which stages of processing are influenced by expectations about others' action steps. We used a paradigm in which a gaze-cueing procedure was embedded in successively presented naturalistic photographs composing an action sequence. Our results showed: (1) behavioral gaze-cueing effects modulated by whether the observed agent gazed at an object which was expected to be gazed-at, according to the action sequence; (2) modulatory effects on the P1/N1 components locked to the onset of a target: while P1 was modulated by gaze validity with respect to target location, N1 was modulated both by gaze validity and participants' expectations about where the agent would gaze at to perform an action; (3) a more positive amplitude in the range of an N300 component, locked to the gaze direction shift for gaze congruent with the action sequence, relative to incongruent and neutral conditions. Taken together, these findings revealed that confirmation or violation of expectations concerning others' goal-oriented actions can modulate attentional selection processes, as indexed by early ERP components.

Keywords: gaze cueing, prediction, action sequences, EEG/ERP, ERP correlates of gaze cueing.

3.2 Introduction

Gaze direction, besides being informative about others' intentions, goals and interests, also indicates their focus of attention, pointing to potentially relevant objects or events in the environment. People tend to follow others' gaze direction, as it has been widely demonstrated in laboratory settings by means of a gaze-cueing paradigm in which participants detect or discriminate targets at gazed-at (i.e., validly cued) locations versus other (invalidly cued) locations. The typical pattern of results shows faster responses for validly (vs. invalidly) cued locations (e.g. Driver et al., 1999; Friesen & Kingstone, 1998) – the *gaze-cueing* effect. This effect has been taken to indicate that visual attention is allocated to a spatial location where others gaze, thereby facilitating visual processing of stimuli (subsequently) presented at that location. While initial findings tended to support the notion that gaze following is a purely reflexive process (Friesen & Kingstone, 1998; Driver et al., 1999), more recent evidence indicates that gaze-cueing effects can be modulated by top-down control (Teufel, Alexis, Clayton, & Davis, 2010; Bayliss, Schuch, & Tipper, 2010; Perez-Osorio, Müller, Wiese & Wykowska, 2015; Wiese, Wykowska, Zwickel, & Müller, 2012; Wykowska, Wiese, Prosser, & Müller, 2014; Wiese, Wykowska & Müller, 2014). Factors such as knowledge about others' mental states (Teufel et al., 2010; Teufel, Fletcher, & Davis, 2010; Wiese, Zwickel, & Müller, 2013; Wiese, Wykowska, & Müller 2014), reliability of gaze 'cues' (Wiese, Wykowska & Müller, 2014), background context in general (Wiese, Zwickel & Müller, 2012) or expectations regarding others' behavior (Perez-Osorio, Müller, Wiese & Wykowska, 2015) modulate a relatively automatic component of spatial-attentional orienting in response to gaze direction cues.

While gaze direction is in general informative regarding others' mental states, it particularly facilitates understanding and anticipation of people's next action steps. Predictions about subsequent action steps based on gaze direction are possible due to the strong coupling between action planning and eye movements. Flanagan and Johansson (2003) examined participants' oculomotor behavior in a block-stacking task. The authors found that eye fixations proactively preceded manual actions related to moving the blocks, and to the landing positions of those manual actions. Furthermore, and importantly, simply observing another agent performing the block-stacking task produced similar eye movement patterns. Flanagan and Johansson concluded that action observation (alone) evokes similar oculomotor programs to those employed in action production, in particular, for object-oriented goal-directed task. Similarly, Hayhoe, Shrivastava, Mruczek, and Pelz (2003), recording eye movements of participants in natural task situations (such as making a sandwich), found that gaze was directed to task-relevant objects that were about to be manipulated. Hayhoe et al. argued that eye fixations provide stable representations of the visual task environment, allowing the extraction of critical information required for high-precision movements. Taken together, both studies show that gaze behavior can offer reliable information regarding successive steps in complex action sequences, in controlled as well as natural action scenarios.

In this context, it is plausible to assume that expectations about forthcoming actions modulate gaze following. Indeed, a recent study (Perez-Osorio, Müller, Wiese, & Wykowska, 2015) showed that gaze-guided orienting of attention can be modulated by expectations about observed action sequences. In that study, a gaze-cueing manipulation was embedded in an action scenario consisting of a sequence of naturalistic photographs depicting a person (a woman named, e.g., Anna) completing a goal-oriented action. For each trial, an image defined a task for Anna: either a guest ask her to bring something to drink, or her friend ask her to fetch some fabric softener to do the laundry. In the subsequent image, she was depicted in the kitchen, in the center of the image, behind a counter and looking straight ahead. On the counter, she would see a bottle with orange juice on one side and a bottle with fabric softener on the other, together with a plastic cup next to each bottle. Subsequently, Anna gazed at either the action-congruent or the action-incongruent bottle (e.g., within the action context of bringing a drink, the congruent bottle is the one containing the orange juice, while the incongruent bottle is the one with the softener). In the final frame, one of the plastic cups (target) was filled with some liquid (either orange juice or softener). Only one cup had liquid in it, and this always corresponded to the adjacent bottle (e.g., yellow juice in the cup next to the orange-juice bottle). Participants' task was to discriminate whether the level of liquid in the plastic cup was high or low. Gaze cueing effects were analyzed with respect to whether the observed gaze behavior was in line with participants' expectations regarding an action sequence (i.e., whether the actor, Anna, directed her gaze to an object congruent or incongruent with the action goal specified at the beginning of the trial). Results showed that when the actor's gaze behavior violated the observer's expectations, the gaze cueing effects were strongly attenuated or even completely suppressed, relative to both gaze behavior confirming expectations and gaze behavior in a neutral baseline conditions (in which no goal was previously specified).

3.2.1 Aim of the study

The aim of the present study was to identify electrophysiological correlates of expectation-related modulations of gaze-cueing effects. ERP correlates of the gaze cueing effects have previously been observed on the early sensory P1 and/or N1 components, locked to the onset of the target (Schuller & Rossion, 2001). As the gaze-cueing protocol is a modified version of a standard Posner-type cueing paradigm (e.g., Posner & Cohen, 1984), the P1/N1 validity effects can be interpreted analogously to validity effects observed on the P1/N1 components in other attention-related cueing paradigms (Doalloa et al., 2003, Anllo-Vento, 1995; Eimer, 1994, 1998, 2000; Fu, Fan, Chen, & Zhuo, 2001; Hillyard, Luck, & Magnun, 1994; Hopf, Vogel, Woodman, Heinze & Luck, 2002; Hopfinger & Mangun, 1998; Mangun & Hillyard, 1991; Mangun et al., 1997; Magnun, Hillyard, & Luck, 1993), that is, as indicators of the *sensory-gain-control* mechanism (Mangun & Hillyard, 1990; Mangun et al., 1997; Luck, Woodman, & Vogel, 2000). This mechanism is thought to bring about an increase in the signal-to-noise ratio (SNR) for stimuli at attended, relative to stimuli at other locations (Hawkins et al., 1990; Müller & Findlay, 1988). In more detail, the P1 and N1 components observed at parieto-occipital sites show an earlier onset and increased amplitudes for stimuli at cued, relative to other, locations (Mangun, Hillyard, & Luck, 1993). The P1 component is thought to reflect a perceptual suppression for ignored locations, whereas the N1 indexes enhanced discriminative processing of stimuli within the focus of attention (for reviews, see Hillyard, Vogel, & Luck, 1998, and Woodman, Vogel & Luck, 2000). Recent findings suggest that the N1 effect reflects top-down modulation of discriminative processing in areas of the ventral visual stream (Pedota et al., 2012). In a more social context, these ERP indices of the sensory-gain-control mechanism (P1 and N1) have also been shown to be modulated by attribution of mind to the observed agent (Wykowska, Wiese et al., 2014).

Therefore, within the context of the literature on P1/N1 effects in cueing paradigms, we predicted the P1/N1 complex to be modulated not only by gaze-cue validity, but also by participants' expectations regarding the gaze behavior of the observed agent in relation to her action goals. To examine this prediction, we carried out an EEG/ERP experiment in which participants performed the same task as in Perez-Osorio et al. (2015). We expected to find a validity-related modulation of the P1/N1 ERP components related (i.e., locked) to target onset in the discrimination task. Furthermore, we aimed at examining ERP effects related to gaze shifts themselves, and whether they confirm or violate participants' expectations regarding the action sequence of the observed agent.

Here, we based our hypotheses regarding the ERP effects on previous literature regarding neural activity related to expectations: For example, Pelphrey, Singerman, Allison, and McCarthy (2003) examined the degree of brain activation elicited by perceived gaze shifts directed either towards (congruent) or away from (incongruent) a location that contained a small checkerboard (visual stimulus). Authors reported that all observed gaze shifts consistently elicited larger activation in the posterior superior temporal sulcus (STS), intra parietal sulcus (IPS), and fusiform gyrus (FFG) compared to no shifts. Importantly, however, the activation was increased when perceived gaze shifts were incongruent with the location of the visual stimulus, relative to congruent locations. Corresponding findings have been reported in analogous studies that examined event-related potentials (ERPs): an enhanced posterior occipital component (N330) locked to gaze shift was found to be more negative when the observed gaze shifts were incompatible with the location of a visual stimulus, relative to compatible conditions (Senju, Johnson, & Csibra, 2006; Senju, Csibra, & Johnson 2008). Authors argued that the observed activation was attributed to operation of the STS in response to violations of expectation for human action, in line with Pelphrey and colleagues (2003). Using a similar paradigm, Tipples, Johnston, and Mayes (2013) examined ERPs when participants were observing two types of central cues – either a human agent's gaze or an arrow symbol – directed to a location that could contain a visual target (participants task was to determine whether the direction of the cue and the visual stimulus matched or not). Results revealed a congruency effect for both cues gaze shift and arrow onsets on the N300 component in the time window between 310-350 ms in the parieto-occipital areas similar to those reported by Senju and colleagues (2006, 2008). Importantly, though, the effect was more pronounced for the gaze shifts, relative to the arrows. Taken together, Tipples and colleagues suggested that activation observed in STS reflects a more general brain mechanism related to detection of unexpected events for both social and nonsocial stimuli. Therefore, in our design we also expected an ERP component elicited around 300 ms after the shift of the gaze to be modulated by expectations that participants held regarding the actor's gaze behavior Importantly, our study provided an additional unique contribution to the state-of-the-art literature: we included a neutral baseline which allowed for interpreting the results in a univocal manner. In the studies of Pelphrey, et al., 2003, Senju et al., 2003, 2006, as well as Tipples et al., 2012, no neutral baseline condition was used, restricting analysis to comparisons between incongruent and -congruent conditions. Therefore, the observed effects could either be due to violations of observer's expectations (larger STS activity and ERP amplitude); or confirmation of observers' expectations (smaller activity and ERP amplitudes). Thus, including a neutral baseline in the present study allowed for univocal interpretation of the effects related to expectancies regarding behavior of the observed agent. In sum, in line with findings of an expectancy-related N300 ERP component in paradigms with gaze-direction shifts (Senju et al., 2006, 2008; Tipples et al., 2013), we expected a modulation of the N300 locked to the gaze shift (i.e., an event preceding target presentation), which would be related to participants' expectations regarding the behavior of the observed agent.

3.3 Methods

Participants. A total of twenty-two volunteers (taking part in the experiment for course credits or for honorarium) were recruited. Out of these twenty-two, eighteen participants provided usable data sets (artifact-free trials >70%) (Age: M = 23.33 years, SD = 1.84; 13 women, 1 left-handed). All participants had normal or corrected-to-normal vision, normal color vision, and provided written informed consent. None of the participants included in the analyses had reported history of neurological diseases, or had previously taken part in an experiment with a similar design.

Apparatus and Stimuli. The experiment was performed in a dimly lit cabin. Participants looked at a 17" standard CRT monitor (100-Hz refresh rate, 1024 x 768 pixels screen resolution) positioned approximately 85 cm from their eyes. Stimulus presentation was controlled by a Pentium-IV PC using the software package E-Prime (Psychology Software Tools, Pittsburgh, USA). The stimuli consisted of a series of color photographs taken for the purposes of this study – see Figure 1. The photographic images covered a screen area of 13.75° (width) x 10.35° (height) of visual angle; images were presented centrally, 6.7° from the screen borders.

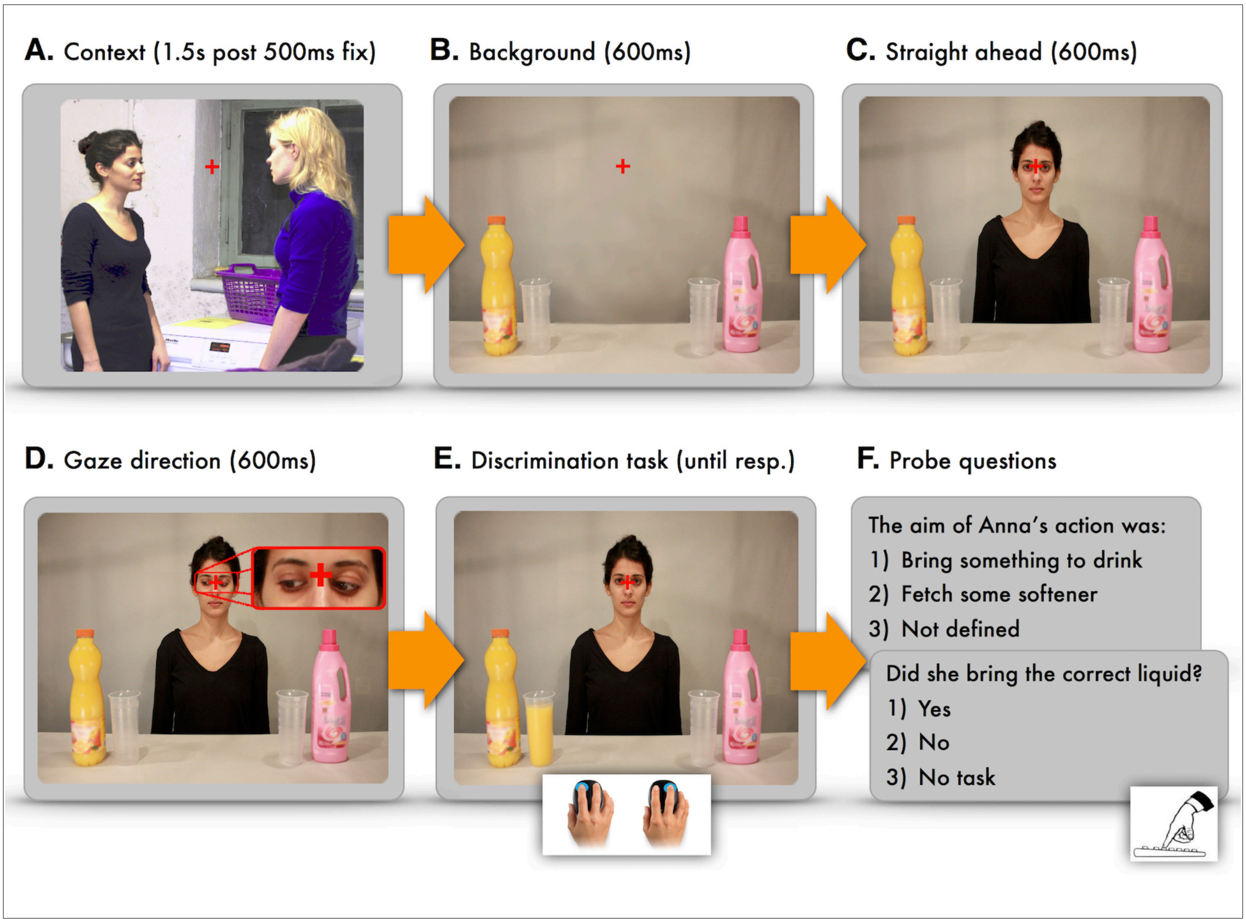

Figure 1. Schematic representation of an example trial, depicting a 'laundry' context with an incongruent gaze direction and a validly cued target. Gaze shift related ERPs were recorded in the Frame D, meanwhile target related ERPs were recorded in the Frame E. Gaze direction is zoomed-in only for the purpose of illustration.

Design and Procedure: The present study used a 3 x 2 factorial design (gaze congruency: congruent/incongruent/neutral; gaze validity: valid/invalid). Gaze congruency was manipulated by embedding the trial sequence within an action scenario. The action scenario was introduced in an image presented at the beginning of each trial (randomized across trials). The scenario could either be (1) "bringing a drink to a friend", (2) "bringing softener to a flat-mate", or (3) neutral (see Figure 2).

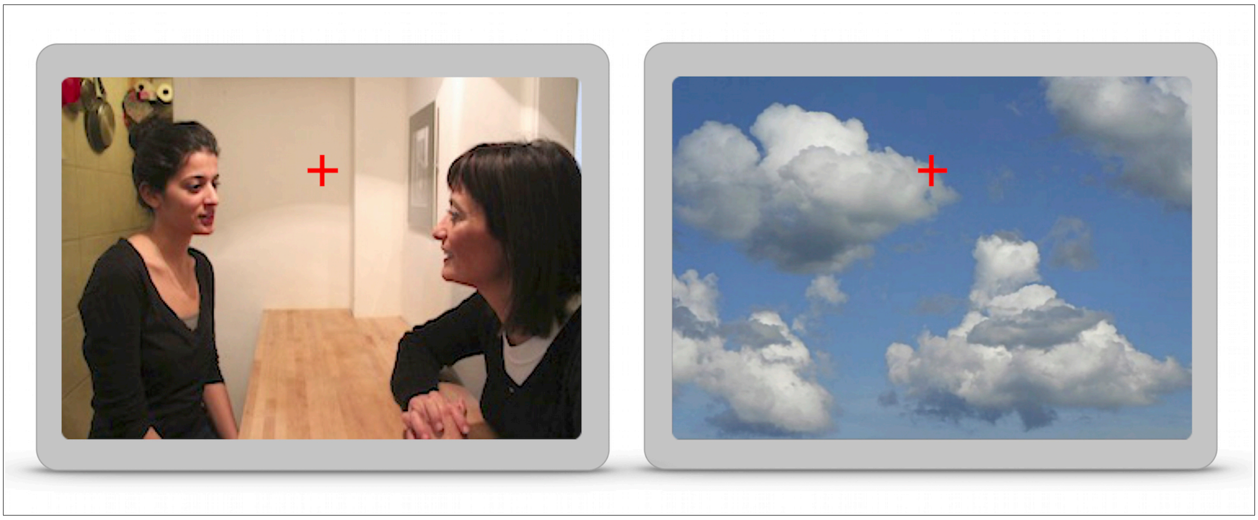

Figure 2. The two other action context images. The 'drink' context (left) and the 'neutral' context (right).

After the scenario was set, the main character, "Anna", was presented between two objects (see frame C of Figure 1), one of which was congruent with the action scenario, and the other incongruent. At this point, participants would form an expectation regarding where Anna would direct her gaze to (i.e., to the action-congruent object) in order to perform her task. The next frame depicted Anna gazing at one of the objects (either the one congruent or that incongruent with the action context). Finally, in the last frame (frame E of Figure 1), a target appeared in one of the cups next to one of the objects. As Anna had gazed at one of the objects in the preceding frame, the target could be either validly cued or invalidly cued by Anna's gaze (independently of, and orthogonal to, whether Anna had gazed at the action-congruent or the incongruent object). Thus, the two factors (congruency and validity) were operationalized as: (1) Anna's gaze landing on an object congruent or incongruent with the action context, and (2) the target object being validly or invalidly cued (in terms of location) by Anna's gaze direction. The most important manipulation was that before the frame containing the target (Figure 1D) the actor's gaze was directed to either the bottle congruent or the bottle incongruent with the action context. That is, her gaze could be directed to the orange juice (yellow) in the 'bring-a-drink' scenario, or to the softener (pink) in the 'laundry' scenario (congruent conditions); or her gaze could be averted to the softener in the 'drink' context, or orange juice in the 'laundry' context (incongruent conditions). In the *congruency-neutral condition*, the image presented at the start of the trial depicted a sky with clouds, rather than a social scene (see Figure 2). Therefore, although the actor's gaze was directed to one of the bottles, this had no relation to an action context (because there was no action context specified in the congruency-neutral condition). This condition was introduced as a baseline for the gaze-cueing effects.

The conditions were distributed evenly across the experiment (33% for each gaze congruency condition and 50% for each validity condition). All conditions, including position of the bottles (i.e., orange juice/softener on the left/right), the target type (orange juice or softener), and the level of liquid (low/high) were pseudo-randomized across trials. In total, the experiment consisted of 576 trials divided into 24 blocks. Participants had two additional blocks for practice before the actual experiment. Feedback about accuracy and reaction time in the target discrimination task for each trial block was provided in the breaks between blocks. Participants were asked to fixate on the fixation cross in the center of the screen throughout the trial (see Figure 1). Additionally, they were explicitly informed that the direction of Anna's gaze was not predictive with regard the location of the target.

Individual trials consisted of the following sequence of images (cf. Figure 1): First, a fixation point appeared at the center of the screen for 500 ms (Figure 1A). Next, a centered context picture was presented for 1500 ms; there were different images for the "drink", "softener", and "neutral" contexts – see Figure 2. (Figure 1B) This was followed by a picture displaying a kitchen counter with two bottles on it: yellow orange juice and pink softener, on opposite sides, equidistant from the center (5.71°); an empty transparent plastic cup ("glass") was located next to each bottle; this picture remained on the screen for 600 ms (Figure 1C) Anna appeared between the two bottles looking straight ahead for 600 ms (Figure 1D). For another 600 ms, she looked to one of the sides, or she kept looking straight ahead (neutral validity trials). (Figure 1E) An image was displayed (until response) in which Anna was presented again looking straight ahead, and which contained the response-relevant target: one of the glasses was partially filled with one of the two types of liquid to a high or a low level. The task was to determine whether the level of liquid in the target cup was low or high (target discrimination task). Participants were asked to respond as fast and as accurately as possible by pressing the left mouse key for a low level and the right button for a high level of liquid. After the target response, two questions were presented (Figure 1F): (i) an action-context question ("The aim of Anna's action was: bring a drink to the guest/bring softener to her flat-mate/not defined"; the arrangement of the three possible answers was randomized across trials), and (ii) a question regarding the liquid (Did Anna pour the correct liquid?: Yes/No/No task) with three possible response options: 'the aim of Anna's action was: bring a drink to the guest, bring some softener, not defined'. Responses to both questions were given by pressing the 1, 2, or 3 key on a standard computer keyboard, with response accuracy (rather than speed) being stressed. Feedback about accuracy and reaction time in the target discrimination task for each entire block was provided in the breaks between blocks.

Behavioral analysis

Behavioral analyses were conducted as follows: first, trials with incorrect responses to the action and liquid questions were excluded (this was intended to ensure that the analyses were performed on trials in which participants had actually encoded and maintained the action-scenario information). Second, trials with incorrect target-discrimination responses were excluded. Observers' error rates were less than 3% for all tasks: liquid-level discrimination (target response), M = 1.29%, SD = 0.01; action question, M = 2.60%, SD = 0.03; and liquid question, M = 2.84%, SD = 0.03. Next, individual participants' median reaction times (RTs) were calculated for each condition and subjected to a 3 × 2 repeated-measures ANOVA with the factors gaze congruency (congruent/ incongruent/ neutral) and gaze cue validity (valid/ invalid). In all analyses, when the sphericity assumption was violated, degrees of freedom were adjusted according to Greenhouse-Geisser's procedure.

EEG recording and analysis

EEG was recorded with Ag-AgCl electrodes from 64 electrodes of an active electrode system (ActiCap, Brain Products, GmbH, Munich, Germany). Horizontal and vertical EOG were recorded bipolar from the outer canthi of the eyes and from above and below the observer's left eye, respectively. All electrodes were referenced to Cz and re-referenced offline to the average of all electrodes. Electrode impedances were kept below 5 kΩ. Sampling rate was 500 Hz, and the

EEG activity was amplified with a band-pass filter of 0.1 to 250 Hz using BrainAmp amplifiers (Brain Products, Munich).

Raw data was filtered offline using a 30-Hz high-cutoff filter. The data was averaged over 800-ms epochs including a 200-ms pre-stimulus baseline, time-locked to (1) target onset and to (2) gaze shift onset. For eye-movement artifacts, we inspected the F10, F9, Fp1, and vEOG channels using an automatic artifact-rejection procedure. Trials with eye movements and blinks on either of the channels specified above (indicated by any absolute voltage difference in a segment exceeding 50 µV or voltage steps between two sampling points exceeding 80 µV) were excluded from analyses prior to averaging. Moreover, channels with other artifacts (all channels considered) were excluded if amplitude exceeded ± 80 µV or any other voltage that was lower than 0.10 µV for a 100-ms interval. The epochs were baseline-corrected with the 200-ms baseline period prior to stimulus onset.

Target-locked ERPs

To examine the ERP correlates of the behavioral gaze-cueing effect, we focused on the P1/N1 components locked to the target onset. The time windows for the P1 and N1 were selected based on the latency of the grand-average peak amplitude in the gaze-neutral condition for the O1, O2, PO3, PO4, PO7, and PO8 channels (pooled). For the P1 peak at 100 ms, a time window between 80–120 ms (±20 ms, relative to the peak latency) was selected; and for the N1 peak at 183 ms, a time window between 143–223 ms (±40 ms relative to the peak latency). The mean amplitudes within the respective time windows were subjected to separate two-way ANOVAs with gaze congruency (congruent, incongruent, vs. neutral) and validity (valid vs. invalid) as within-subject factors. Planned comparisons (two-tailed t-tests) were conducted for the valid versus invalid conditions in each gaze congruency conditions. Where appropriate, statistics were corrected according to Greenhouse-Geisser for potential nonsphericity. The experiment had 96 repetitions per condition. After rejection of eye movement artifacts and incorrect-response trials, 74 trials remained on average in each experimental condition (gaze congruent valid: 74; and invalid: 74; gaze incongruent valid: 70; and invalid: 69, gaze congruency neutral: valid 76; and invalid: 78, trials). Level of liquid (high/low) and side of presentation (left/right) were averaged together.

Gaze-shift-locked ERPs

To examine the brain responses to violations of (action-context-related) expectations, we focused on ERPs locked to onset of Anna's gaze shift, in particular, the N300, with the analyses focusing on the comparison between the gaze-congruency conditions. The gaze-locked activity was measured at Left and Right Lateral (electrode) Regions of Interest (following Senju et al., 2006, and Tipples et al., 2012), including the electrodes TP7, TP8, P7, P8, T7, and T8. As visual inspection of the grand average revealed the N300 component to be relatively broad (which is also in line with the findings of Tipples et al., 2013), we selected a time window of ±100 ms centered on 300 ms following the onset of the gaze shift. The mean amplitudes of the EEG signal in this 200–400-ms time window were subjected to a one-way ANOVA with congruency (congruent, incongruent, vs. neutral) as a within-subject factor. Where appropriate, statistics were corrected according to Greenhouse-Geisser for potential nonsphericity.

The experiment included 192 repetitions per gaze congruency condition. After rejection of eye movement artifacts and incorrect response trials, 160 repetitions remained on average (congruent: 157; incongruent: 158; and neutral: 165 trials). Orange-juice and softener bottles, as being the gazed-at objects, were averaged together.

3.4 Results

Behavioral Data

Average median RTs and standard errors for each condition are presented in Table 1.

Table 1. Average Median RTs and SEM [in ms] as a function of cue validity and gaze congruency.

Gaze congruency	Validity	
	Valid	Invalid
Congruent	485 [21]	540 [28]
Incongruent	506 [25]	513 [23]
Neutral	479 [24]	499 [24]

The 3 x 2 repeated-measures ANOVA of the median RTs with the factors congruency (congruent, incongruent, neutral) and validity (valid, invalid) revealed both main effects to be significant. The congruency effect [F (1.24, 21.19) = 5.899, p = .019, η_p^2 = .258] was due to RTs being faster in the neutral condition (M [SEM] = 489 [24] ms) as compared to both the congruent and the incongruent condition (M [SEM] = 512 [24] ms and, 509 [24] ms, respectively), though planned comparisons (two-tailed t-test, Bonferroni corrected) revealed that only the difference between the neutral and congruent to be significant [t (17)= 2.722, p = .043, d_z = .64]. The main effect of validity [F (1,17) = 34.209, p = .00002, η_p^2 = .668] was due to RTs being faster with valid (M [SEM] = 490 [22] ms) than with invalid gaze cues (M [SEM] = 517 [25] ms).

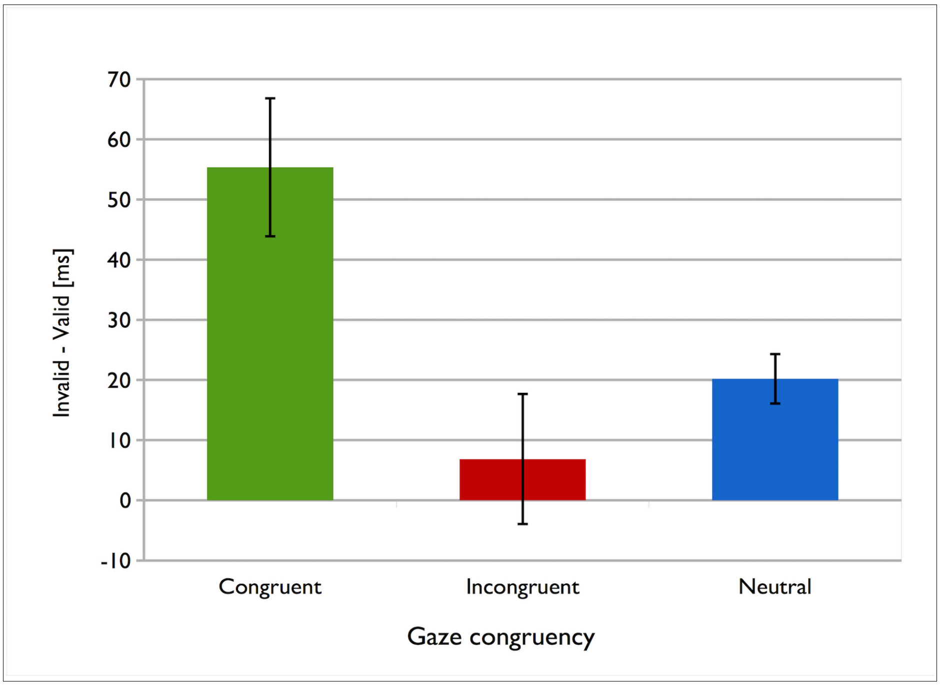

Figure 3. Gaze-cueing effect as a function of gaze congruency in the behavioral, RT data. Error bars represent 95%-confidence intervals (CIs) adapted to within-participants designs according to Cousineau's (2005) procedure.

Most importantly for the purposes of this experiment, and replicating the results of Perez-Osorio et al. (2015), the interaction between congruency and validity was significant [F (1.202, 20.423) = 4.547, p = .039, η_p^2 = .211], with the validity effect being significant only in the congruent and neutral gaze conditions, but not in the incongruent condition (valid vs. invalid for congruent gaze: ΔRT= 55.36 ms, t(17)= 3.778, p = .002, d_z = 0.89; for incongruent gaze: ΔRT

= 6.86 ms, *t* (17)= .651, *p* = .524, d_z = 0.15; and for neutral gaze: ΔRT = 20.20 ms, *t* (17)= 4.999, *p* = .0001, d_z = 1.16). Planned comparisons (two-tailed t-tests) of the gaze-cueing effects (ΔRT = M RTinvalid – M RTvalid) revealed a significant difference in ΔRT between the congruent and incongruent gaze conditions [*t* (17)= 2.170, *p* = .044, d_z = 0.51] and between the congruent and neutral conditions, [*t* (17)= 2.570, *p* = .020, d_z = 0.60], but no difference between the incongruent and neutral conditions [*t* (17)= -1.121, *p* = .278, d_z = 0.26]. In other words, the gaze-cueing effect was significantly enhanced in the gaze-congruent condition as compared to the gaze-incongruent and gaze-neutral conditions (see Figure 3).

EEG results

Target-locked ERPs

Grand-average ERPs over the 18 subjects elicited by the onset of the target display are illustrated in Figure 4.

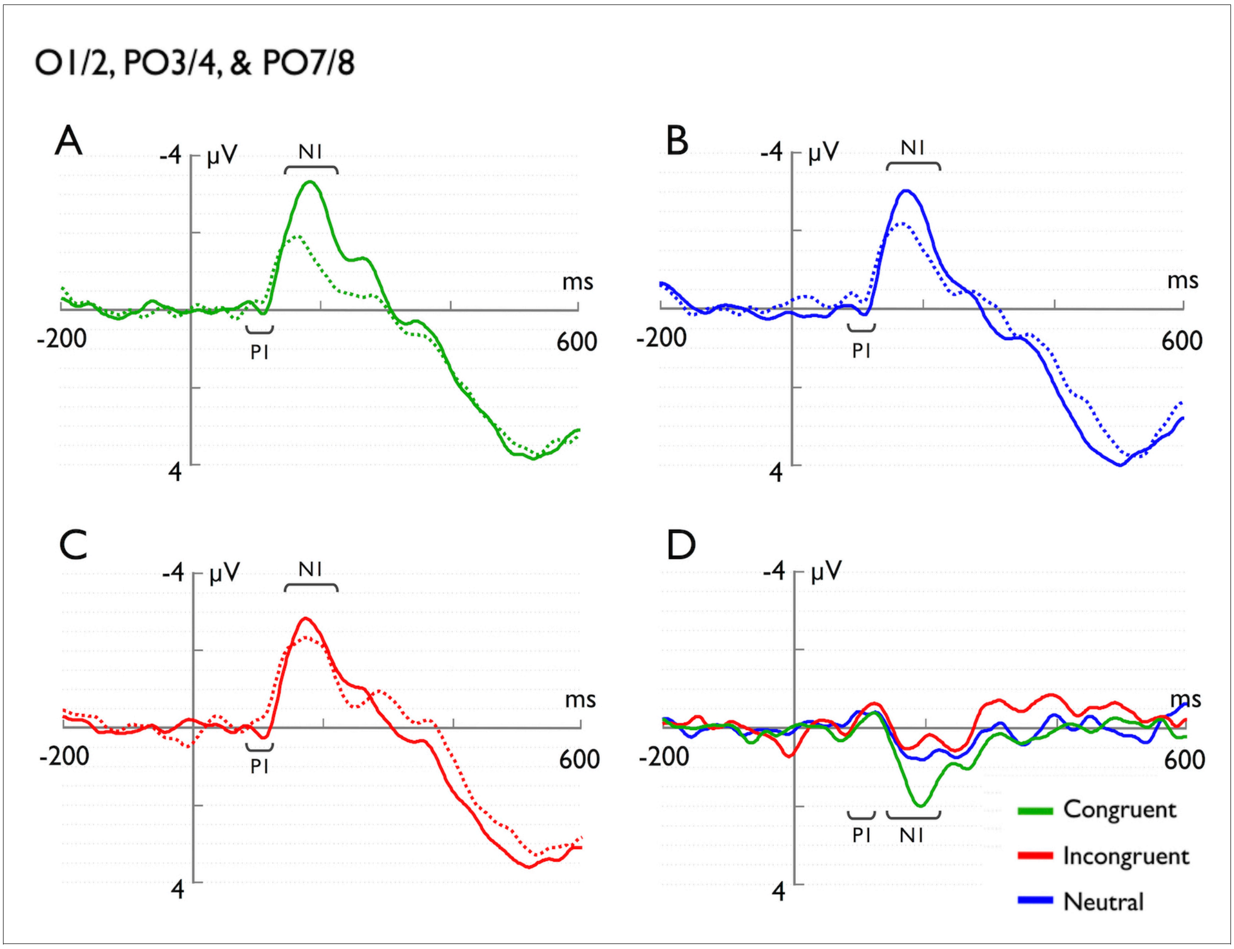

Figure 4. Grand average ERP waveforms time-locked to target onset. Solid lines represent valid trials, dashed lines invalid trials. (A) Congruent, (B) incongruent, and (C) neutral gaze-congruency conditions. (D) Difference waves valid vs. invalid trials. P1 time window: 80–120 ms; N1 time window: 143–223 ms.

P1 component

Analysis of amplitudes in the P1 time window 80–120 ms post-target onset revealed a main effect of gaze-direction validity [F (1, 17) = 6.141, p = .024, η_p^2 = .265], due to more positive amplitudes for trials with valid (M [SEM] = .049 [.276] μV) as compared to invalid gaze cues (M [SEM] = -.230 [.261]μV). However, neither the main effect of congruency [F (2, 34) = .123, p = . 884, η_p^2 = .007] nor the interaction between congruency and validity were significant [F (2, 34) = .102, p = .806, η_p^2 = .013].

N1 component

Analysis of amplitudes in the N1 time window between 143–223 ms post-target onset revealed a main effect of validity [F (1, 17) = 10.670, p = .005, η_p^2 = .386], due to more negative amplitudes for valid (M [SEM] = -2.48 [.418] μV) relative to invalid gaze cues (-1.70 [.425] μV); the main effect of congruency was non-significant [F (2, 34) = .365, p = .697 η_p^2 = .021]. Importantly, and in line with the behavioral results, the interaction between congruency and validity was significant [F (2, 34) = 11.755, p = .0001, η_p^2 = .406], with the validity effect being reliable only in the congruent- and neutral-gaze conditions, but not in the incongruent condition (valid vs. invalid for congruent gaze: -1.36 μV, t(17)= 4.409, p = .0003, d_z = 1.03; for incongruent gaze: -.263 μV, t (17)= .930, p = .365, d_z = 0.22; and for neutral gaze: -.708 μV, t (17)= 3.287, p = . 004, d_z = 0.77). Similarly, analyses of the validity effect (mean amplitudes valid minus mean amplitudes invalid) between gaze-congruency conditions revealed significant differences for congruent vs. incongruent [t (17)= 5.755, p = .00007, d_z = 1.36], and congruent vs. neutral [t (17) = 3.026, p = .023, d_z= .71], but not incongruent vs. neutral [t (17) = 1.652, p = .35, d_z = .39] (planned comparisons, Bonferroni corrected).

Gaze-shift-locked N300

Grand-average ERPs elicited by the onset of the gaze-shift display are illustrated in Figure 5. The one-way ANOVA (with the within-subject factor congruency: congruent, incongruent, neutral) on the mean amplitudes of the N300 component (in the time window 200–400 ms) in the region of interest (covering the channels TP7, TP8, P7, P8, T7, and T8), locked to the onset of the gaze shift, revealed a significant congruency effect, F (2, 34) = 6.203, p = .005, η_p^2 = .267.

As corroborated by subsequent planned comparisons, this was due to a significantly more negative amplitude for incongruent (M [SEM] = -1.09 [.38] μV) compared to congruent trials (-.48 [.33] μV), *t* (17) = 3.242, *p* = .005, d_z = .76; see Figure 5. Importantly, however, there was no significant difference in the mean amplitudes (*p* = .57) between the incongruent and the neutral condition (M [SEM]= -.98 [.41] μV). The neutral condition differed significantly only from the more positive amplitude of the congruent condition, *t* (17) = 2.760, *p* = .013, d_z = .65.

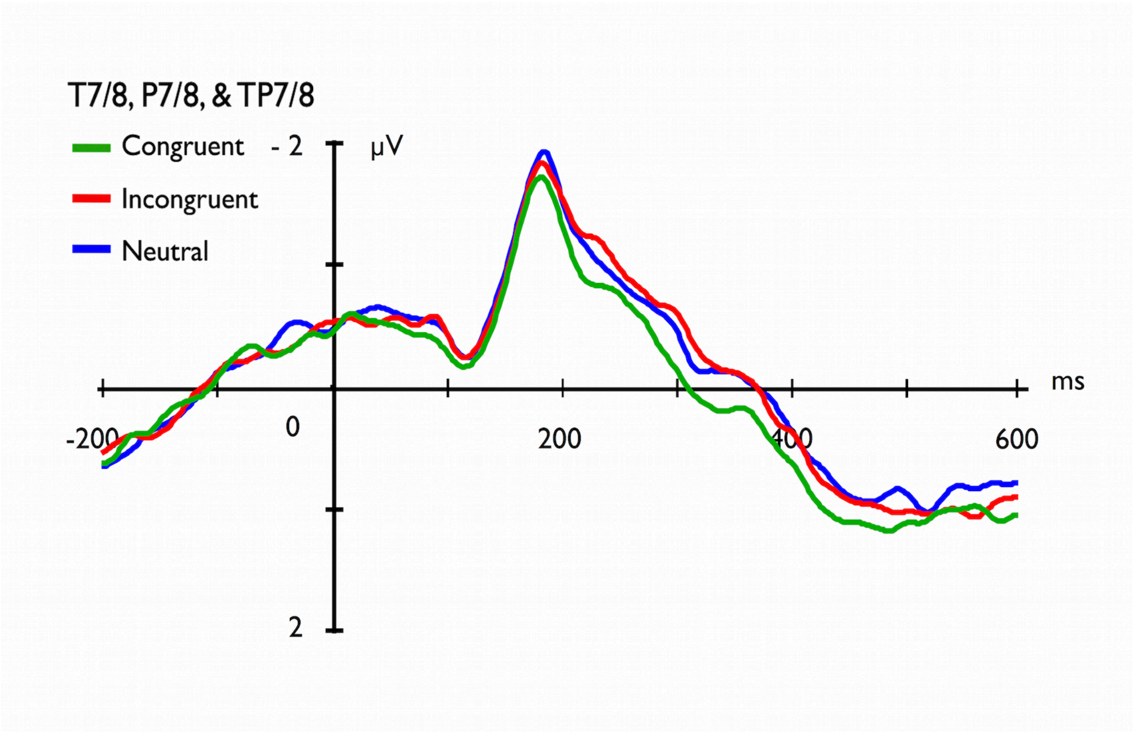

Figure 5. Grand average ERP waveforms time-locked to gaze-shift onset. N300 time window: 200–400 ms.

3.5 Discussion

The aim of the present study was to investigate the stages of processing at which expectations regarding gaze behavior of others' modulate gaze cueing effects. To this end, we used a paradigm that embedded a gaze-cueing protocol within a complex action sequence. A series of naturalistic photographs depicted a woman (Anna) completing a goal-oriented action. In the critical frame within the trial sequence, Anna could gaze either at an object that the observer expected to be manipulated by Anna, based on the action Anna 'instructed' to perform; i.e., the action-congruent object. Alternatively, Anna could gaze at the other, action-incongruent object.

Participants were asked to make a discrimination of the level of liquid positioned in a cup next to the object gazed-at by Anna (validly cued) or, alternatively, a cup next to the other object (invalidly cued by Anna's gaze). Importantly, gaze congruency with respect to the action scenario and gaze validity were manipulated orthogonally. Gaze-cueing effects were calculated as the difference in RTs for validly vs. invalidly cued objects (in terms of location), and analyzed in relation to whether Anna gazed at a congruent or an incongruent object (in terms of action goal). We expected to replicate behavioral results previously reported by Perez-Osorio et al. (2015), that is, attenuation or complete abolishment of the gaze-cueing effects when the actor's gaze behavior violated participants' expectations regarding the action sequence. In terms of ERPs, we expected the P1/N1 complex (target-locked) to show validity effects being modulated by gaze congruency. Target-locked ERPs were analyzed for the parieto-occipital region (O1, O2, PO3, PO4, PO7, and PO8) for the P1/N1 components (time windows of 80–120 ms and, respectively, 143–223 ms),following Schuller & Rossion (2001), Magnun, Hillyard, & Luck (1993) Hopf, Vogel, Woodman, Heinze, & Luck (2002). We also expected to find effects of expectations on ERPs locked to the gaze shifts of the observed agent. These ERPs were analyzed in the posterior lateral regions (T7,T8, P7, P8, TP7, and TP8) in the time window of the N300 component (following Senju et al., 2006, 2008, and Tipples et al., 2012)

We found that the behavioral results followed the same pattern as previously reported (Perez-Osorio, et al, 2015): (1) faster target discrimination responses to targets at validly cued, relative to invalidly cued, locations – the typical gaze-cueing effect; and (2) modulation of the gaze-cueing effect by action expectation: an enhanced gaze-cueing effect when the actor gazed at objects congruent, rather than incongruent or neutral, with respect to the action context. This implies that participants' expectations with regard to the gaze behavior displayed by the observed agent modulated gaze following. In other words, gaze-cueing effects involve (at least to some degree) a top-down component that modulates spatial-attentional orienting in response to shifts of gaze direction.

Importantly, we found that the target-locked ERPs mirrored the behavioral results: larger validity effects (increased amplitude in the occipital-temporal N1 component for validly cued locations vs. invalidly cued locations) in congruent, relative to incongruent, gaze conditions. Interestingly, modulatory effect was observed on the N1 component, N1 is thought to reflect perceptual discrimination processes for stimuli at attended locations (Ritter, Simson, & Vaughan, 1983; Ritter, Simson, Vaughan, & Macht 1982; Senkowski & Hermann, 2002; Vogel & Luck, 2000; Fort, Besle, Giard, & Pemier, 2005; Luck & Hillyard, 1995; Mangun, 1995; Mangun & Hillyard, 1991; Parasuraman, 1980). Previous studies have shown that a number of stimulus properties can modulate N1 amplitudes (e.g.: color or shape: Vogel & Luck, 2000; task difficulty, defined in terms of the similarity of the target to the distractors: Fedota, McDonald, Roberts, & Parasuraman, 2012). This modulation has been interpreted in terms of an influence of top-down control over early discrimination processes. Hence, in our paradigm – in which the top-down influence originated in expectations about the forthcoming action sequence – the N1 effects parallel previous findings. Importantly, this is the first study that extends previous results to more naturalistic, social situations.

Furthermore, apart from the target-related ERP effects and behavioral results, we found that the N300 component –locked to gaze-shift onset– yielded a more negative amplitude in the incongruent condition relative to the congruent condition, which is consistent with Senju et al. (2006) and Tipples et al. (2013). However, and importantly, our paradigm also included a baseline (i.e., an action-neutral) condition. Our results showed that the incongruent condition did not elicit an ERP amplitude significantly different from the neutral condition. Instead, it was the congruent condition that yielded a (significantly) more positive ERP amplitude compared to the neutral baseline. Accordingly, the difference between the incongruent and congruent conditions (also reported by Senju et al., 2006, 2008; and Tipples et al., 2013) represents a more positive amplitude related to congruent gaze, rather than an increased negativity related to incongruent gaze. Hence, our findings support the notion that the modulation of the N300 component might be better understood as a congruency-related positivity, rather than incongruency-related negativity, as proposed by Senju et al. (2006) and Tipples et al. (2013). Moreover, this effect pattern is paralleled in our behavioral data, as the gaze-cueing effects were larger for the congruent condition, relative to both the incongruent and the neutral condition (the two latter did not differ from each other).

The congruency-related positivity is perhaps best interpreted in line with attention-related effects on the P1/N1 components in spatial cueing paradigms (Hillyard, Vogel, & Luck, 1998; Woodman, Vogel, & Luck, 2000): the typical pattern of results (see Hillyard et al., 1998) is that of an enhanced N1 for validly cued targets, relative to invalidly cued targets and neutral cueing conditions. Participants in spatial cueing also expect a stimulus to appear at the cued location, relative to other locations. Therefore, expected stimulus events yield more enhanced amplitudes, relative to unexpected or neutral conditions. Similarly, in our study, the expected gaze event yielded more enhanced amplitude, relative to the neutral and unexpected conditions. Even though this difference was in the 'negativity' range, the more positive amplitude in the expected condition can be seen as enhanced amplitude, as this was the only condition that differed from the neutral baseline. Moreover, both the behavioral effects and the target-locked ERPs showed this as an enhancement of processing in the congruent condition (larger gaze-cueing effects for congruent vs. incongruent and neutral, and no difference between incongruent and neutral). Hence, in the present study expectancy yielded a more positive amplitude of the N300 component, in accordance with attention-related studies showing also more enhanced amplitudes for expected stimuli, relative to unexpected or neutral conditions.

It is plausible that the effect of gaze congruency (with respect to the action context) on the N300 component is related to activation of STS, which has been found to be associated with the processing of expectations (Sebanz et al., 2006; Tipples et al., 2013) and with observed gaze shifts (Pelphrey et al., 2003, 2004, 2005; Senju et al., 2006, Tipples et al., 2013), as well as other types of (observed) biological motion (Hirai et al., 2005; Jokisch et al., 2005). Our results are also in line with the proposed function of the STS, namely, the encoding of the relations between perceived biological motion and environmental context (Pelphrey, Morris, & McCarthy, 2004; Saxe, Xiao, Kovacs, Perrett, & Kanwisher, 2004). Previous studies also indicate that the STS is the main area of a network involved in understanding others' intentions, deployment of anticipatory mechanisms in action observation, and prediction of the successive action steps produced by (observed) others (Frith & Frith, 2006). This area has been shown to be specialized in processing of human non-verbal cues (for review, see Puce & Perrett, 2003), determining gaze direction (Haxby, Hoffman, & Gobbini, 2000), and detecting unexpected human body movements (Grezes, Frith, & Passingham, 2004; Pelphrey et al., 2004; Morris, Pelphrey, & McCarthy, 2005; Saxe et al., 2004). In addition to the STS, the IPS (Pelphrey et al., 2003, 2005) and the FFG (Pelphrey et al., 2003) have been reported to show larger activation in response to object-incongruent gaze shifts. Thus, it is conceivable that IPS activation could also contribute to the observed effects. Although one cannot draw conclusions regarding the neural sources of the scalp-recorded EEG without performing source reconstruction analyses, it is probable that the present effects are linked to the activity of the STS/IPS and FFG network.

Overall, the present study provides electrophysiological evidence in support of the idea that gaze following can be modulated by expectations regarding (observed) gaze behavior within the context of complex actions. Gaze following can be enhanced when the observed agent confirms expectations the observers have concerning where the agent would look next, given the action the agent is performing. This finding is not only consistent with previous studies demonstrating the impact of top-down control on gaze-cueing effects (Hoffman & Haxby, 2000; Kanwisher, 2000; Ristic & Kingstone, 2005; Bayliss, Schuch, & Tipper, 2010; Teufel et al., 2010; Wiese, Wykowska et al., 2012; Wykowska, Wiese et al., 2014), but it also shows that gaze direction serves the purpose of informing others about (likely) forthcoming action steps towards reaching the goal of the task (Flanagan & Johannson, 2003). In this context, our findings elucidate the functional role of the gaze following in naturalistic scenarios, while also revealing links between higher-order cognition and lower-level social attention mechanisms. In detail, humans do not simply (passively) observe others' actions; rather, they (actively) generate predictions about subsequent action steps, based on shared action schemata (Schubotz & von Cramon, 2002) and understanding of the underlying action goals.

In conclusion, the present study is the first to provide electrophysiological evidence for the idea that expectations regarding others' actions can modulate gaze-guided orientation of attention. Confirmation, or violation of expectations concerning other's actions can provide signals for the top-down modulation of early sensory processing, as indexed by well-understood ERP components.

Acknowledgements

This work was supported by EXC142 grant (DFG-Excellence Cluster ‘Cognition for Technical Systems’, sub-project #435: SocGaze) awarded to HM, and a DFG grant awarded to AW (WY-122/1-1). We thank Ella Bahry, Mallory Klauning, and Lee Simon-Vermot for their help with developing the photographic images used in this study.

3.6 References

Anllo-Vento, L. (1995). Shifting attention in visual space: the effects of peripheral cueing on brain cortical potentials. *Int J Neurosci*, *80*(1-4), 353–370.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson’s method. *Tutorials in Quantitative Methods for Psychology*, *1*(1), 42–45. http://doi.org/no DOI found

Doallo, S., Lorenzo-López, L., Vizoso, C., Rodríguez Holguín, S., Amenedo, E., Bará, S., & Cadaveira, F. (2005). Modulations of the visual N1 component of event-related potentials by central and peripheral cueing. *Clinical Neurophysiology*, *116*(4), 807–820. http://doi.org/10.1016/j.clinph.2004.11.013

Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze Perception Triggers Reflexive Visuospatial Orienting. *Visual Cognition*, *6*(5), 509–540. http://doi.org/10.1080/135062899394920

Eimer, M. (1994). An ERP study on visual spatial priming with peripheral onsets. *Psychophysiology*, *31*(2), 154–163.

Eimer, M. (1998). Does the face-specific N170 component reflect the activity of a specialized eye processor? *Neuroreport*, *9*(13), 2945–2948. http://doi.org/10.1097/00001756-199809140-00005

Eimer, M. (2000). The time course of spatial orienting elicited by central and peripheral cues: Evidence from event-related brain potentials. *Biological Psychology*, *53*(2-3), 253–258. http://doi.org/10.1016/S0301-0511(00)00049-1

Fedota, J. R., Mcdonald, C. G., Roberts, D. M., & Parasuraman, R. (2012). Contextual task difficulty modulates stimulus discrimination: Electrophysiological evidence for interaction between sensory and executive processes. *Psychophysiology*, *49*(10), 1384–1393. http://doi.org/10.1111/j.1469-8986.2012.01455.x

Flanagan, J. R., & Johansson, R. S. (2003). Action plans used in action observation. *Nature*, *424*(6950), 769–771. http://doi.org/10.1038/nature01861

Fort, A., Besle, J., Giard, M. H., & Pernier, J. (2005). Task-dependent activation latency in human visual extrastriate cortex. *Neuroscience Letters*, *379*(2), 144–148. http://doi.org/10.1016/j.neulet.2004.12.076

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by non-predictive gaze. *Psychonomic Bulletin & Review*, *5*(3), 490–495. http://doi.org/10.3758/BF03208827

Fu, S., Fan, S., Chen, L., & Zhuo, Y. (2001). The attentional effects of peripheral cueing as revealed by two event-related potential studies. *Clinical Neurophysiology*, *112*(1), 172–185. http://doi.org/10.1016/S1388-2457(00)00500-9

Grèzes, J., Frith, C., & Passingham, R. E. (2004). Brain mechanisms for inferring deceit in the actions of others. *The Journal of Neuroscience□: The Official Journal of the Society for Neuroscience*, *24*(24), 5500–5505. http://doi.org/10.1523/JNEUROSCI.0219-04.2004

Hawkins, H. L., Hillyard, S. A., Luck, S. J., Mouloua, M., Downing, C. J., & Woodward, D. P. (1990). Visual attention modulates signal detectability. *Journal of Experimental Psychology. Human Perception and Performance*, *16*(4), 802–811. http://doi.org/10.1037/0096-1523.16.4.802

Haxby, J., Hoffman, E., & Gobbini, M. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*(6), 223–233. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10827445

Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision*, *3*(1), 49–63. http://doi.org/10.1167/3.1.6

Hillyard, S. a, Vogel, E. K., & Luck, S. J. (1998). Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *353*(1373), 1257–1270. http://doi.org/10.1098/rstb.1998.0281

Hillyard, S., Luck, S., & Mangun, G. (1994). The cuing of attention to visual field locations: Analysis with ERP recordings. In *Cognitive electrophysiology* (pp. 1–25). Retrieved from http://books.google.com/books?hl=en&lr=&id=Pmtn2S-vaaEC&oi=fnd&pg=PR5&dq=Cognitive+Electrophysiology&ots=IvKS2ncG9g&sig=vi-u2XMLoXbkrvNmCIdhTL3A1ojM\nhttp://books.google.com/books?hl=en&lr=&id=Pmtn2S-vaaEC&oi=fnd&pg=PR5&dq=Cognitive+electrophysiology&ots=I

Hirai, M., Senju, A., Fukushima, H., & Hiraki, K. (2005). Active processing of biological motion perception: An ERP study. *Cognitive Brain Research*, *23*(2-3), 387–396. http://doi.org/10.1016/j.cogbrainres.2004.11.005

Hopf, J.-M., Vogel, E., Woodman, G., Heinze, H.-J., & Luck, S. J. (2002). Localizing visual discrimination processes in time and space. *Journal of Neurophysiology*, *88*(4), 2088–2095. http://doi.org/10.1152/jn.00860.2001

Hopfinger, J. B., & Mangun, G. R. (1998). Reflexive Attention Modulates Processing of Visual Stimuli in Human Extrastriate Cortex. *Psychological Science*. http://doi.org/10.1111/1467-9280.00083

Jokisch, D., Daum, I., Suchan, B., & Troje, N. F. (2005). Structural encoding and recognition of biological motion: Evidence from event-related potentials and source analysis. *Behavioural Brain Research*, *157*(2), 195–204. http://doi.org/10.1016/j.bbr.2004.06.025

Luck, S. J., & Hillyard, S. A. (1995). The role of attention in feature detection and conjunction discrimination: an electrophysiological analysis. *The International Journal of Neuroscience*, *80*(1-4), 281–297. http://doi.org/10.3109/00207459508986105

Luck, S., Woodman, G., & Vogel, E. (2000). Event-related potential studies of attention. *Trends in Cognitive Sciences*, *4*(11), 432–440. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11058821

Mangun, G. R. (1995). Neural mechanisms of visual selective attention. *Psychophysiology*, *32*(1), 4–18. http://doi.org/10.1111/j.1469-8986.1995.tb03400.x

Mangun, G. R., Hillyard, S. a, & Luck, S. J. (1993). Electrocortical substrates of visual selective attention. *Attention and Performance 14: Synergies in Experimental Psychology, Artificial Intelligence, and Cognitive Neuroscience.* Retrieved from http://search.ebscohost.-com/login.aspx?direct=true&db=psyh&AN=1993-97600-009&site=ehost-live

Mangun, G. R., & Hillyard, S. A. (1991). Modulations of sensory-evoked brain potentials indicate changes in perceptual processing during visual-spatial priming. *Journal of Experimental Psychology. Human Perception and Performance*, *17*(4), 1057–1074. http://doi.org/10.1037/0096-1523.17.4.1057

Mangun, G. R., Hopfinger, J. B., Kussmaul, C. L., Fletcher, E. M., & Heinze, H. J. (1997). Covariations in ERP and PET measures of spatial selective attention in human extrastriate visual cortex. In *Human Brain Mapping* (Vol. 5, pp. 273–279). http://doi.org/10.1002/(SICI)1097-0193(1997)5:4<273::AID-HBM12>3.0.CO;2-F

Morris, J. P., Pelphrey, K. A., & McCarthy, G. (2005). Regional brain activation evoked when approaching a virtual human on a virtual walk. *Journal of Cognitive Neuroscience*, *17*(11), 1744–1752. http://doi.org/10.1162/089892905774589253

Müller, H. J., & Findlay, J. M. (1988). The effect of visual attention on peripheral discrimination thresholds in single and multiple element displays. *Acta Psychologica*, *69*(2), 129–155. http://doi.org/10.1016/0001-6918(88)90003-0

Parasuraman, R. (1980). Effects of information processing demands on slow negative shift latencies and N100 amplitude in selective and divided attention. *Biological Psychology*, *11*(3-4), 217–233. http://doi.org/10.1016/0301-0511(80)90057-5

Pelphrey, K. A., Morris, J. P., & McCarthy, G. (2004). *Grasping the intentions of others: the perceived intentionality of an action influences activity in the superior temporal sulcus during social perception. Journal of cognitive neuroscience* (Vol. 16).

Pelphrey, K. A., Morris, J. P., Michelich, C. R., Allison, T., & McCarthy, G. (2005). Functional anatomy of biological motion perception in posterior temporal cortex: An fMRI study of eye, mouth and hand movements. *Cerebral Cortex*, *15*(12), 1866–1876. http://doi.org/10.1093/cercor/bhi064

Pelphrey, K. A., Singerman, J. D., Allison, T., & McCarthy, G. (2003). Brain activation evoked by perception of gaze shifts: The influence of context. *Neuropsychologia*, *41*(2), 156–170. http://doi.org/10.1016/S0028-3932(02)00146-X

Pelphrey, K., Singerman, J., Allison, T., & McCarthy, G. (2003). Brain activation evoked by perception of gaze shifts: the influence of context. *Neuropsychologia*, *41*, 156–170. Retrieved from http://www.sciencedirect.com/science/article/pii/S002839320200146X

Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In *Attention and performance X: Control of language processes* (pp. 531–556). http://doi.org/10.1162/jocn.1991.3.4.335

Ritter, W., Simson, R., & Vaughan, H. G. (1983). Event-related potential correlates of two stages of information processing in physical and semantic discrimination tasks. *Psychophysiology*, *20*(2), 168–179. http://doi.org/10.1111/j.1469-8986.1983.tb03283.x

Ritter, W., Simson, R., Vaughan, H. G., & Macht, M. (1982). Manipulation of event-related potential manifestations of information processing stages. *Science (New York, N.Y.)*, *218*(4575), 909–911. http://doi.org/10.1126/science.7134983

Saxe, R., Xiao, D. K., Kovacs, G., Perrett, D. I., & Kanwisher, N. (2004). A region of right posterior superior temporal sulcus responds to observed intentional actions. *Neuropsychologia*, *42*(11), 1435–1446. http://doi.org/10.1016/j.neuropsychologia.2004.04.015

Schubotz R.I., von Cramon DY. (2002) Predicting perceptual events activates corresponding motor schemes in lateral premotor cortex: an fMRI study. Neuroimage;15(4):787–96. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11906220

Schuller, a M., & Rossion, B. (2001). Spatial attention triggered by eye gaze increases and speeds up early visual activity. *Neuroreport*, *12*(11), 2381–6. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11496114

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in Cognitive Sciences*, *10*(2), 70–6. http://doi.org/10.1016/j.tics.2005.12.009

Senju, A., Csibra, G., & Johnson, M. H. (2008). Understanding the referential nature of looking: infants' preference for object-directed gaze. *Cognition*, *108*(2), 303–19. http://doi.org/10.1016/j.cognition.2008.02.009

Senju, A., Johnson, M. H., & Csibra, G. (2006). The development and neural basis of referential gaze perception. *Social Neuroscience*, *1*(3-4), 220–34. http://doi.org/10.1080/17470910600989797

Senkowski, D., & Herrmann, C. S. (2002). *Effects of task difficulty on evoked gamma activity and ERPs in a visual discrimination task. Clinical neurophysiology□: official journal of the International Federation of Clinical Neurophysiology* (Vol. 113).

Teufel, C., Alexis, D. M., Clayton, N. S., & Davis, G. (2010). Mental-state attribution drives rapid, reflexive gaze following. *Attention, Perception & Psychophysics*, *72*(3), 695–705. http://doi.org/10.3758/APP.72.3.695

Teufel, C., Fletcher, P. C., & Davis, G. (2010). Seeing other minds: attributed mental states influence perception. *Trends in Cognitive Sciences*, *14*(8), 376–82. http://doi.org/10.1016/j.tics.2010.05.005

Tipples, J., Johnston, P., & Mayes, A. (2013). Electrophysiological responses to violations of expectation from eye gaze and arrow cues. *Social Cognitive and Affective Neuroscience*, *8*(5), 509–14. http://doi.org/10.1093/scan/nss024

Vogel, E. K., & Luck, S. J. (2000a). The visual N1 component as an index of a discrimination process. *Psychophysiology*, *37*(2), 190–203. http://doi.org/10.1111/1469-8986.3720190

Vogel, E. K., & Luck, S. J. (2000b). *The visual N1 component as an index of a discrimination process. Psychophysiology* (Vol. 37).

Wiese, E., Wykowska, A., & Müller, H. J. (2014). What we observe is biased by what other people tell us: Beliefs about the reliability of gaze behavior modulate attentional orienting to gaze cues. *PLoS ONE*, *9*(4). http://doi.org/10.1371/journal.pone.0094529

Wiese, E., Wykowska, A., Zwickel, J., & Müller, H. J. (2012). I See What You Mean: How Attentional Selection Is Shaped by Ascribing Intentions to Others. *PLoS ONE*, *7*(9). http://doi.org/10.1371/journal.pone.0045391

Wiese, E., Zwickel, J., & Müller, H. J. (2013). The importance of context information for the spatial specificity of gaze cueing. *Attention, Perception & Psychophysics*, *75*(5), 967–82. http://doi.org/10.3758/s13414-013-0444-y

Wykowska, A., Wiese, E., Prosser, A., & Müller, H. J. (2014). Beliefs about the minds of others influence how we process sensory information. *PloS One*, *9*(4), e94339. http://doi.org/10.1371/journal.pone.0094339

3. GENERAL DISCUSSION

4.1 Summary of results

The goal of the presented PhD-project was to examine whether higher-order prediction mechanisms of the human brain influence a fundamental process of social cognition, namely, gaze following –or in other words, shared attention. In **Experiments 1 and 2** We investigated how attention to objects in a naturalistic scene is influenced by two factors: the (perceived) gaze direction of an actor, and background expectations regarding the action context. Results across experiments showed consistently (1) faster target discrimination responses to targets at validly cued, relative to invalidly cued, locations –the typical gaze-cueing effect; and (2) modulation of the gaze-cueing effect by action expectation: an enhanced gaze-cueing effect when the actor gazed at objects congruent, rather than incongruent or neutral, with respect to the action context. Additionally, we examined electrophysiological correlates of action-related modulation of gaze cueing effect in **Experiment 3** and found modulatory effects on the P1/N1 components locked to the onset of a target; P1 was modulated by gaze validity with respect to target location, and N1 was modulated both by gaze validity and participants' expectations about where the agent should gaze in order to perform an action. Furthermore, we also found that an ERP component in the range of N300, locked to shift of gaze direction (before the target onset) was more positive when gaze was congruent with the action sequence, relative to incongruent and neutral conditions. Results in Experiment 3 add electrophysiological evidence to the notion that expectations regarding others' actions can modulate gaze-guided orientation of attention. Confirmation or violation of expectations concerning other's actions can provide signals for the top-down modulation of early sensory processing, as indexed by well-understood ERP components. Collectively, our findings suggest that participants' expectations with regard to the gaze behavior displayed by the observed agent modulated gaze following. In other words, gaze-cueing effects involve (at least to some degree) a top-down component that modulates attentional orienting in response to shifts of gaze direction.

4.2 Gaze following is modulated by expectations about other's behavior

The results of experiments presented in this thesis showed that task performance (i.e., discrimination of the liquid level in the target glass) depended on gaze cue validity with respect to the target side, with attention following the gaze direction of the observed agent (gaze cueing effects). However, the gaze-cueing effects were modulated by whether the actor had gazed at an action-congruent or an incongruent object: the cueing effects were significantly reduced (Experiment 1), if not entirely eliminated (Experiment 2 and 3), when the actor gazed at objects incongruent with action context, relative to congruent objects. The differential cueing effect between the congruent and incongruent conditions was due to gaze following being enhanced when the actor's gaze was directed to the action-congruent object (the neutral and incongruent conditions yielded smaller effects compared to the congruent condition, in all Experiments), and possibly also suppression of gaze following, as evidenced by the non-significant cueing effect in the incongruent condition (as compared to significant effects in the neutral and congruent conditions). The complete lack of a validity effect in the latter case would mean that participants did not follow the actor's gaze when she looked at an object that was action-incongruent, thus demonstrating the potency of top-down control over bottom-up-driven orienting of attention in response to gaze direction.

Figure A5 represents the hypothesized dynamics of attentional deployment over the course of an experimental trial. Upon presentation of the first image following the action context, the observer would attend to the action-congruent object. Subsequently, they would follow the observed agent's gaze direction in the scene, which was corroborated by the significant main effect of gaze cue validity. Finally, in the last photograph of the trial sequence, the observer's attention would be summoned to the target glass (filled with liquid to a certain level), as participants were required to discriminate the level of liquid in the target. Not surprisingly, the fastest RTs were found when the observed agent's gaze was directed to the congruent object and the response was to be made to the same (target) object (A5, panel A) –that is, the condition in which no attention shifts were required along the trial sequence. Conceivably, the congruent object being gazed-at by the actor in the scene confirmed and reinforced the observer's prior expectations, strongly summoning attention to the action-congruent location. The costs related to *gaze-incongruent* locations (A5 panel B) were minimal, relative to when the gaze was directed towards action-congruent objects (A5 panel A).

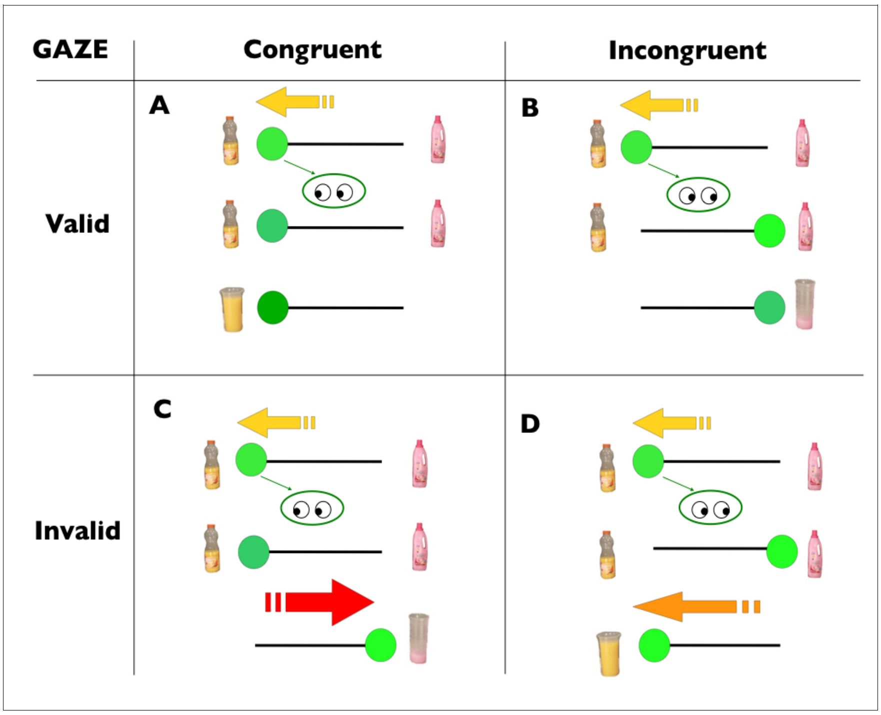

Figure A5. The four main conditions of interest (neutral conditions not represented). The action context selected for the example is "**bring something to drink**". The rows in each panel represent successive events in a trial sequence: the first row depicts the presentation of the kitchen scene with two bottles on the opposite sides of the counter; the second row represents the gaze shift event, in which the actor directed her gaze to one of the two sides; finally, the third row depicts the target event, in which the target (a filled glass) was presented. The fastest RTs were obtained in the congruent gaze, valid condition, in which the gaze was congruent with the context and gaze cue was valid (A). In the incongruent gaze, valid condition (B), the RTs were similar to gaze congruent valid. However, in the incongruent gaze, invalid condition (D), the RTs were increased relative to both conditions with valid gaze cues (A, B), indicative of gaze following. Importantly, the largest increase in RTs relative to the other conditions was observed for incongruent gaze, invalid trials (C), that is, the performance cost was largest when the gaze direction confirmed the initial expectations but the target appeared in the non-cued location. The hypothetical focus of attention is represented by green circles and shifts of the attentional focus by arrows. Arrows of different colors represent hypothesized shifts of attention, with darker colors (orange and red) representing temporal costs of attentional shifts between hemifields.

By contrast, the longest RTs were elicited when the agent's gaze was action-congruent but did not validly cue the target location (Figure A5 panel C). In this condition, the initial assumption of the observer, who would have anticipated the actor's gaze to be directed to the action-congruent object, was actually confirmed by the displayed gaze behavior. As a result, attention would have remained focused on the action-congruent *and* gazed-at object. Hence, when the target did eventually appear on the opposite side, the observer's expectations were violated and a conflict ensued, as the observer was now to produce a response to an invalidly cued target at an unexpected location. This elicited the largest cost in performance –larger RTs–, relative to the other conditions. In the *incongruent-gaze* and *invalidly-cued-target* condition (Figure A5 panel D), the actor's gaze was violating the initial expectation already at the second step of the trial sequence and the costs were smaller relative to when the initial expectations concerning gaze direction were confirmed (A5 panel C). Moreover, in Experiment 2 and 3 when the actor's gaze was incongruent with the action context, no gaze-cueing effect was observed. That is, the facilitation typically associated with gaze cues was diminished or eliminated when expectations regarding which object would be gazed at (given the overarching action context) were violated. Finally, the congruent gaze condition yielded stronger gaze-cueing effects relative to the neutral action context baseline in all experiments.

Our pattern of results is in line with previously observed 'combined-expectancy' effects on selective attention in modality-cueing tasks, in which the response-critical target stimulus varied across sensory modalities (Klein & Hansen, 1990; Kingstone & Klein, 1991; Kingstone, 1992; Mattler, 2003, 2005; Mattler, van der Lugt, & Münte, 2006). In typical modality cueing tasks, participants are cued to expect a stimulus in a given modality (e.g., either vision or audition), with the task being to discriminate a stimulus defined within the (cued or uncued) target modality (e.g., to make a two-alternative forced choice response either to the color of a square or to the pitch of a tone). In a task combining response and modality expectancies (e.g., Mattler, 2005; Mattler et al., 2006), participants were asked to respond to either the color of the stimulus or the pitch of a tone by pressing one of two keys with their left or their right hand (e.g., the target might be a visual stimulus, in which case red color required manual response A and green color response B; or the target might be an auditory stimulus, with a high-pitch stimulus requiring manual response A and a low-pitch stimulus response B). Cues (valid or invalid) presented before the target stimulus indicated the identity of the upcoming stimulus (e.g., "red" or "high"), thus cueing both the modality of the target and the response required. The behavioral results showed larger modality cueing effects (validly cued modality vs. invalidly cued modality) for validly vs. invalidly cued responses. That is, when one of the expectations was confirmed (the validly cued response), violation of the other expectancy (the invalidly cued modality) produced a large impairment (relative to the validly cued modality). However, when both expectations were violated (invalidly cued response and invalidly cued modality), the impairment relative to the validly cued modality was not as large. Our results show a similar pattern: the observed actor's gaze violated the initial expectation regarding her gaze behavior (Figure 5D, second row), violation of the expectation regarding the validity of her gaze with respect to the target object produced a smaller impairment of performance (Figure 5D third row) relative to when the initial expectation was confirmed (Figure 5C, second row) and the later expectation violated (Figure 5C, third row). In a different experimental setup, Hommel (1998, 2004) reported a pattern of results that resembles the present findings: under conditions in which the stimulus and response features could independently change or repeat across trials, the largest impairment of performance was observed when only one but not the other feature (stimulus or response) changed across consecutive trials, in comparison with cases in which both features were either repeated or changed (both stimulus and response were same or both different). Hommel (1998, 2004) attributed these 'partial-repetition costs' to competition between the stimulus-response (S-R) mapping on the current trial and the S-R binding from the previous trial (stored in an episodic 'event file'), where the latter is activated, on the current trial, through repetition of one of the features. Similar effects have been reported in inter-trial analyses of compound-search tasks in which the search-critical and response-critical stimulus attributes varied randomly, and independently, across consecutive trials (e.g., Töllner, Gramann, Kiss, Müller, & Eimer, 2008, Zehetleitner, Rangelov, & Müller, 2012). Töllner et al. (2008) and Zehetleitner et al. (2012) explained their results in terms of 'linked expectancies' with regard to the search- and response-critical attributes.

The most important implication of these results is that gaze cueing appeared to be modulated by participants' expectations with regard to upcoming action steps in an observed action sequence. In other words, relative to the baseline, the gaze-cueing effect was actually enhanced when the actor's gaze 'complied' with observers' expectations regarding the object that would be gazed at in order to achieve the ultimate action goal. That is, based on their background knowledge of the social/action context, humans activate a certain action schema (Manthley et al., 2003) and implicitly expect the gaze of the observed agent to precede (and, thus, provide a pointer to) successive action steps (Flanagan & Johansson, 2003). Hence, when the observed gaze behavior conflicts with participants' expectations concerning the action sequence, they gaze following seems to be overridden by top-down control. Since gaze cueing reflects mechanisms underlying joint attention (e.g., Friesen & Kingstone, 1998; Sebanz et al., 2006) – which in turn facilitates establishing common social context and/or is related to action understanding (Sebanz et al., 2006) – the present results suggest that humans might be reluctant to engage in a common social context with those who violate their expectations regarding gaze behavior.

Our results can also be interpreted along the lines of Wurm and Schubotz (2012), who performed a study using fMRI and behavioral measures to examine how action understanding is affected by the compatibility of an action with a context in which it is embedded. Wurm and Schubotz (2012) assumed that action sequences always unfold in a certain context, with given objects and in a particular order. Participants watched video clips of context-specific everyday actions performed in domestic settings and were to stop the video as soon as they had identified the action. The actions were performed in a compatible, an incompatible, or a neutral context (i.e. cutting bread in the kitchen, in the bathroom, or, in an empty white background, respectively). The results revealed RTs to be slower (by 100 ms, on average) for the incompatible versus the compatible condition; no differences were reported between neutral and compatible contexts. Likewise, BOLD responses showed an increased activation in the left ventrolateral prefrontal cortex when the context was incompatible, relative to the other conditions. Wurm and Schubotz (2012) concluded that the effects were driven by the conflict induced by incompatible contexts, which might violate expectations –similar to the current study.

In sum, the present results confirm the idea that gaze-cueing effects can be influenced by whether or not the gazed-at object fits with an overarching action scheme –which is in line with previous findings showing top-down influences on the bottom-up gaze-cueing effect (e.g., Teufel et al., 2010; Wiese et al., 2013; Wiese et al., 2012; Wykowska et al., 2014). However, going beyond previous findings, the present results show that the 'referring' function of the gaze cue critically depends on whether the referred-to object does or does not fit with a background action context, yielding an interactive influence on the deployment of attention: when the agent's gaze is action-congruent, the referring function of the gaze cue and thus the cueing effect is strengthened, whereas it is weakened when the gaze is action-incongruent.

4.3 Electrophysiological correlates of gaze cueing modulation

The present project examined also which stages of processing are influenced by expectations about others' action steps –using the event-related potentials (ERPs) of the EEG signal. Importantly we replicated the behavioral effect showing that gaze-cueing effects are modulated by whether the observed agent gazed at an object which was expected to be gazed-at, according to the action sequence. In addition to the behavioural effects, we found that validly cued locations elicited an increased amplitude in the occipital-temporal N1 component, relatively to invalidly cued locations; but only in action-congruent gaze conditions. Extensive evidence suggest that modulatory effects on N1 play a significant role in perceptual discrimination processes for stimuli at attended locations (Ritter, Simson, & Vaughan, 1983; Ritter, Simson, Vaughan, & Macht 1982; Senkowski & Hermann, 2002; Vogel & Luck, 2000; Fort, Besle, Giard, & Pemier, 2005; Luck & Hillyard, 1995; Mangun, 1995; Mangun & Hillyard, 1991; Parasuraman, 1980). Studies have shown that not only physical properties of the stimuli (e.g. color or shape, Vogel & Luck, 2000), but also task difficulty (defined in terms of the similarity of the target to the distractors: Fedota, McDonald, Roberts, & Parasuraman, 2012) can modulate N1 amplitudes. Hence, in our paradigm the top-down influence –generated by expectations about the forthcoming action sequence– corresponds to N1 effects previously reported. Importantly, to the best of our knowledge, these results reveal for the first time how such modulation unfolds in more naturalistic, social contexts.

Moreover, we found that the N300 component, locked to gaze-shift onset recorded in the temporal occipital areas, was more negative in the incongruent condition relative to the congruent condition. This is consistent with Senju et al. (2006, 2008) and Tipples et al. (2012) who found that the same component yielded a more negative amplitude in response to breaches in expectation of direction of eye/gaze cues. These studies suggest that incongruent conditions elicit a more negative amplitude relative to congruent conditions. However, in contrast to Senju and Tipples, our paradigm also included a baseline condition (i.e., an action-neutral condition).

Our comparisons showed that the incongruent condition did not elicit an ERP amplitude significantly different from the neutral condition. Conversely, our analyses revealed that it was the congruent condition that yielded a (significantly) more positive ERP amplitude relative to neutral baseline. Similarly, rather than an increased negativity related to incongruent gaze, the difference between the incongruent and congruent conditions represents a more positive amplitude related to congruent gaze. Hence, we suggest that the modulation of the N300 component might be better understood as a congruency-related positivity, rather than incongruency-related negativity. It is plausible to assume that larger amplitudes on N300 reflect update of expectations regarding the unfolding action sequence. While, in the gaze congruent condition expectations were matched, in the other two conditions unmatched (incongruent) or not defined (neutral) expectations required updates. Previous findings (Wolpert et al., 2003) suggest that mere action observation elicits inferences regarding the intentions and goals of others. Consistently, we assume that all conditions generated expectations regarding the observed agents' behavior, even in the neutral condition when the action goal was not specified. Therefore, more negative amplitudes in N300 reveal not only detection of unexpected movements but as well might reflect update of expectations. This update regarding other's behavior might provide signals for the top-down modulation of early sensory processing. Additionally, observed N300 component activation paralleled behavioral data, with significantly larger gaze-cueing effects for the congruent condition, relative to both the incongruent and the neutral conditions –which showed no difference.

Collectively, findings in Experiment 3 showed that modulation of gaze following by expectations occurs in early stages of processing. Namely, occipital components closely associated with attention/discrimination can be affected by expectations regarding (observed) gaze behavior within the context of complex actions. Furthermore, our findings suggest that update of expectations might play a role in modulation of early attentional components via top-down regulation.

Our findings are consistent with previous studies that showed the impact of top-down control on gaze-cueing effects (Hoffman & Haxby, 2000; Kanwisher, 2000; Ristic & Kingstone, 2005; Bayliss, Schuch, & Tipper, 2010; Teufel et al., 2010; Wiese, Wykowska et al., 2012; Wykowska, Wiese et al., 2014). Moreover, our findings reveal links between higher-order cognition and lower-level social attention mechanisms, unveiling the functional role of the gaze following in naturalistic scenarios. Hence, prior knowledge helps to understand the underlying action goals of others' behavior online, actively –rather than reactively. Importantly, these predictions seem to impact earlier stages of information processing to the point of modulating deployment of attentional resources.

The observed gaze congruency effects on N300 component might be related to activation of STS (in line with Sebanz et al., (2006); Tipples et al. (2012). It presumably reflects a general mechanism for detecting unexpected events and processing of expectations. Previous studies also suggested that STS –together with the IPS (Pelphrey et al., 2003, 2005) and the FFG (Pelphrey et al., 2003)– is part of a network involved in understanding others' intentions, deployment of anticipatory mechanisms in action observation, and prediction of the successive action steps produced by (observed) others (Frith & Frith, 2006). Particularly, STS has been related with the encoding of: biological motion (i.e. point-light displays of a walking person) (Hirai et al., 2005; Jokisch et al., 2005), relations between perceived biological motion and context (Pelphrey, Morris, & McCarthy, 2004; Saxe, Xiao, Kovacs, Perrett, & Kanwisher, 2004), and observed gaze shifts (Pelphrey et al., 2003, 2004, 2005; Senju et al., 2006, Tipples et al., 2013). Therefore, it is plausible that the present effects are linked to the activity of the STS/IPS and FFG network. However, these are only speculations regarding the sources of the observed EEG signal, as one cannot draw firm based on scalp-recorded EEG without performing source reconstruction analyses.

4.4 Theoretical implications

The present findings have several important theoretical implications. First, they add to the mounting evidence that predictive processing is a ubiquitous mechanism of the human brain, operating in various cognitive domains (including social cognition) at different levels – from low-level perception (Friston, 2005) and motor control (Wolpert & Flanagan, 2001), through intermediate-level processes (Enns & Lleras, 2008) to higher-order cognition. Importantly, our results provide clear evidence that even the higher-order prediction mechanisms affect low-level fundamental processes (of social cognition, in the present case), thus supporting the notion of highly interactive nature of processing in the brain, allowing for bi-directional influences across domains (Friston, 2005; Hommel et al., 2001).

Second, the present results confirm the idea that attentional orienting in response to gaze cues, rather than being purely reflexive in nature (e.g., Driver et al., 1999; Friesen & Kingstone, 1998), is modifiable by top-down control processes (e.g., Teufel et al., 2010; Wiese et al., 2012; Wykowska et al., 2014). Accordingly, our findings support the dual-component account of gaze cueing recently proposed by Wiese et al. (2012), on which gaze cueing involves a (bottom-up) mechanism of automatic attentional orienting to the gazed-at side, which may however be modulated when additional, visual or social/action context information is available. Over and above the previous evidence, the present findings show that top-down modulation of gaze-induced attentional orienting can not only enhance gaze-cueing effects (as in the case of gaze behavior congruent with participants' expectations) relative to baseline (default gaze following), but also completely suppress the default mechanism (as in the case of gaze behavior incongruent with expectations).

Third, the results of this series of studies suggest that action prediction and gaze following are closely connected. During social interaction, each counterpart needs to know what the others are going to do next (Frith & Frith, 2006). Using gaze direction as informative about the identity (*What?*) and the spatial location of attended objects (*Where?*); collecting prior knowledge about the others' preferences (*Who?*) –acquired by observation, direct interaction, or through communication–, people are able to make inferences about others' internal states (*Why?*) and therefore, be able to predict the most likely upcoming actions under the given circumstances (Sebanz, Bekkering, & Knoblich, 2006). Thus, in our paradigm, knowing that the observed agent was asked to bring a glass of orange juice to a guest, one would predict that she would be looking for an empty glass and a bottle of orange juice (in order to pour some juice and bring the filled glass to the guest). Therefore, through linking processing of gaze direction with action prediction within a single paradigm, the current study clearly demonstrates that spatial information derived from gaze direction and context information about the action goal are combined in order to predict consecutive steps in an action sequence. This demonstration highlights the actual function of the gaze-following mechanism and the role it plays in natural daily-life scenarios: gaze following has developed not only to pick up signals that others convey regarding potentially relevant events in the environment, but also, and perhaps foremost, in order to enable us to infer what others are going to do next (Baron-Cohen, 1995; Frith & Frith, 2006).

Finally, our findings have another important implication: they show that humans might engage a mentalizing process (e.g., Baron-Cohen, 1995; Frith & Frith, 2006) even if this is not necessitated by the task. In our study, it was not necessary to infer the mental states of the observed actor to complete the task. All that participants were required to do was to discriminate the level of the liquid in the cup and keep in mind the first, action scenario image (for recall at the end of the trial). Technically, it was a simple target discrimination task combined with a short-term memory task. Yet, the data pattern shows that not only did participants (i) follow the actor's gaze, even though this was not relevant to any of the tasks (the gaze validity effect); but they also (ii) engaged in a (perhaps implicit) mentalizing process involving predictions and expectations regarding where the actor would gaze in order to realize her action goal. The mentalizing process was evidenced indirectly by the effect of gaze congruency on gaze following. If participants had not mentalized about Ella's action goals and intentions, they would not have had any expectations regarding her gaze behavior; hence, violations of these expectations would not have influenced the degree to which they followed Ella's gaze. The effect of gaze congruency on gaze validity cannot be explained by other factors, such as simple integration of background information about the context into the processing of the task-related information, as no effect of target congruency was observed on gaze-neutral trials in Experiment 1. Had the effect been due to the action context itself (rather than a process of mentalizing), we should have observed effects of context on target discrimination, and not the modulation of gaze cueing by gaze congruency. Therefore, by eliciting an 'automatic' mentalizing process – be it theorizing about the observed agent's mental states, in accordance with the Theory-Theory of social cognition (see Apperly, 2008), or by simulating her mental states through one's own cognitive processes, in line with the Simulation Theory (see Apperly, 2008; Sebanz & Knoblich, 2009) –our paradigm can serve the purpose of probing mentalizing in an implicit (i.e., not necessarily explicit) manner. This may be of particular relevance for research on autism spectrum disorder, as the affected individuals might actually be impaired in more implicit forms of social cognition, rather than explicit processes that are probed in typical theory-of-mind tasks (Schilbach et al., 2013).

4.5 Conclusions

To conclude, the present series of studies is, to our knowledge, the first to provide the missing link between the prediction mechanism involved in action understanding and (covert) gaze following. With the use of a novel, naturalistic paradigm, we show that higher-order predictive processes modulate (either enhance or even entirely suppress) low-level mechanisms of social cognition –gaze following– dependent on whether gaze direction of an observed agent is in line or at variance with the expectations that participants hold with regard to the unfolding action sequence. These findings support the idea that predictive processing is a ubiquitous mechanism of the human brain operating in various cognitive domains (including social cognition) at several levels of the information processing hierarchy (from lower-lever predictive coding to higher-order cognitive predictions) and is potent enough to influence rudimentary and fundamental low-level mechanisms, such as gaze following.

5. REFERENCES (Abstract, introduction and general discussion)

Adolphs, R. (1999). Social cognition and the human brain. *Trends in Cognitive Sciences*, *3*(12), 469–479. http://doi.org/10.1016/S1364-6613(99)01399-6

Aggleton, J. P., Burton, M. J., & Passingham, R. E. (1980). Cortical and subcortical afferents to the amygdala of the rhesus monkey (Macaca mulatta). *Brain Research*, *190*(2), 347–368. http://doi.org/10.1016/0006-8993(80)90279-6

Aggleton, J. P. (1993). The contribution of the amygdala to normal and abnormal emotional states. *Trends in Neurosciences*. http://doi.org/10.1016/0166-2236(93)90110-8

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: Role of the STS region. *Trends in Cognitive Sciences*. http://doi.org/10.1016/S1364-6613(00)01501-1

Apperly, I. A. (2008). Beyond Simulation-Theory and Theory-Theory: Why social cognitive neuroscience should use its own concepts to study "theory of mind." *Cognition*, *107*(1), 266–283. http://doi.org/10.1016/j.cognition.2007.07.019

Baron Cohen, S. (1995). *Mindblindness*. *Learning, development, and conceptual change.*

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with autism or Asperger syndrome. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *38*(7), 813–822. http://doi.org/10.1111/j.1469-7610.1997.tb01599.x

Bayliss, A. P., Schuch, S., & Tipper, S. P. (2010). Gaze cueing elicited by emotional faces is influenced by affective context. *Visual Cognition*, *18*(8), 1214–1232. http://doi.org/10.1080/13506285.2010.484657

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological Studies of Face Perception in Humans. *Journal of Cognitive Neuroscience*, *8*(6), 551–565. http://doi.org/10.1162/jocn.1996.8.6.551

Bickart, K. C., Dickerson, B. C., & Feldman Barrett, L. (2014). The amygdala as a hub in brain networks that support social life. *Neuropsychologia*, *63*, 235–248. http://doi.org/10.1016/j.neuropsychologia.2014.08.013

Brothers, L. (1990). The neural basis of primate social communication. *Motivation and Emotion*, *14*(2), 81–91. http://doi.org/10.1007/BF00991637

Bubic, A., von Cramon, D. Y., & Schubotz, R. I. (2010). Prediction, cognition and the brain. *Frontiers in Human Neuroscience*, *4*(March), 25. http://doi.org/10.3389/fnhum.2010.00025

Bukowski, H., Hietanen, J. K., & Samson, D. (2016). From gaze cueing to perspective taking: Revisiting the claim that we automatically compute where or what other people are looking at. *Visual Cognition*, *23*(8), 1020–1042. http://doi.org/10.1080/13506285.2015.1132804

Campbell, R., Heywood, C. A., Cowey, A., Regard, M., & Landis, T. (1990). Sensitivity to eye gaze in prosopagnosic patients and monkeys with superior temporal sulcus ablation. *Neuropsychologia*, *28*(11), 1123–1142. http://doi.org/10.1016/0028-3932(90)90050-X

Castelli, F., Happé, F., Frith, U., & Frith, C. (2000). Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. *NeuroImage*, *12*(3), 314–25. http://doi.org/10.1006/nimg.2000.0612

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *The Behavioral and Brain Sciences*, *36*(3), 181–204. http://doi.org/10.1017/S0140525X12000477

Corbetta, M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Ollinger, J. M., Drury, H. A., ... Shulman, G. L. (1998). A common network of functional areas for attention and eye movements. *Neuron*, *21*(4), 761–773. http://doi.org/10.1016/S0896-6273(00)80593-0

Coull, J. T. (1998). Neural correlates of attention and arousal: Insights from electrophysiology, functional neuroimaging and psychopharmacology. *Progress in Neurobiology*. http://doi.org/10.1016/S0301-0082(98)00011-2

Coull, J. T., & Nobre, A. C. (2008). Dissociating explicit timing from temporal expectation with fMRI. *Current Opinion in Neurobiology*. http://doi.org/10.1016/j.conb.2008.07.011

de C. Hamilton, A. F., & Grafton, S. T. (2007). The motor hierarchy: from kinematics to goals and intentions. *Sensorimotor Foundations of Higher Cognition*, pp. 381. http://doi.org/DOI:10.1093/acprof:oso/9780199231447.003.0018

DeLong, K. a, Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117–1121. http://doi.org/10.1038/nn1504

Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze perception triggers reflexive visuospatial orienting. *Visual Cognition*, *6*(5), 509–540. http://doi.org/10.1080/135062899394920

Eacott, M. J., Heywood, C. A., Gross, C. G., & Cowey, A. (1993). Visual discrimination impairments following lesions of the superior temporal sulcus are not specific for facial stimuli. *Neuropsychologia*, *31*(6), 609–619. http://doi.org/10.1016/0028-3932(93)90055-5

Emery, N. J., Lorincz, E. N., Perrett, D. I., Oram, M. W., & Baker, C. I. (1997). Gaze Following and Joint Attention in Rhesus Monkeys (Macaca mulatto). *Journal of Comparative Psychology*, *111*(3), 286–293.

Enns, J. T., & Lleras, A. (2008). What's next? New evidence for prediction in human vision. *Trends in Cognitive Sciences*, *12*(9), 327–333. http://doi.org/10.1016/j.tics.2008.06.001

Farroni, T., Johnson, M. H., Brockbank, M., & Simion, F. (2000). Infants' use of gaze direction to cue attention: The importance of perceived motion. *Visual Cognition*, *7*(6), 705–718. http://doi.org/10.1080/135062800501443999

Farroni, T., Johnson, M. H., & Csibra, G. (2004). Mechanisms of eye gaze perception during infancy. *Journal of Cognitive Neuroscience*, *16*(8), 1320–1326. http://doi.org/10.1162/0898929042304787

Farroni, T., Mansfield, E. M., Lai, C., & Johnson, M. H. (2003). Infants perceiving and acting on the eyes: Tests of an evolutionary hypothesis. *Journal of Experimental Child Psychology*, *85*(3), 199–212. http://doi.org/10.1016/S0022-0965(03)00022-5

Fedota, J. R., Mcdonald, C. G., Roberts, D. M., & Parasuraman, R. (2012). Contextual task difficulty modulates stimulus discrimination: Electrophysiological evidence for interaction between sensory and executive processes. *Psychophysiology*, *49*(10), 1384–1393. http://doi.org/10.1111/j.1469-8986.2012.01455.x

Flanagan, J. R., & Johansson, R. S. (2003). Action plans used in action observation. *Nature*, *424*(6950), 769–771. http://doi.org/10.1038/nature01861

Fleischer, J. G. (2007). Neural Correlates of Anticipation in Cerebellum , Basal Ganglia , and Hippocampus. *Behavioral Neuroscience*, 19–34. http://doi.org/10.1007/978-3-540-74262-3

Fort, A., Besle, J., Giard, M. H., & Pernier, J. (2005). Task-dependent activation latency in human visual extrastriate cortex. *Neuroscience Letters*, *379*(2), 144–148. http://doi.org/10.1016/j.neulet.2004.12.076

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review*, *5*(3), 490–495. http://doi.org/10.3758/BF03208827

Friesen, C. K., Moore, C., & Kingstone, A. (2005). Does gaze direction really trigger a reflexive shift of spatial attention? *Brain and Cognition*, *57*(1), 66–9. http://doi.org/10.1016/j.bandc.2004.08.025

Friesen, C. K., Ristic, J., & Kingstone, A. (2004). Attentional effects of counterpredictive gaze and arrow cues. *Journal of Experimental Psychology. Human Perception and Performance*, *30*(2), 319–29. http://doi.org/10.1037/0096-1523.30.2.319

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *360*(1456), 815–36. http://doi.org/10.1098/rstb.2005.1622

Friston, K. J., & Stephan, K. E. (2007). Free-energy and the brain. *Synthese*, *159*(3), 417–458. http://doi.org/10.1007/s11229-007-9237-y

Frith, C. D., & Frith, U. (2006). How we predict what other people are going to do. *Brain Research*, *1079*(1), 36–46. http://doi.org/10.1016/j.brainres.2005.12.126

Fuster, J. M. (2001). The Prefrontal Cortex—An Update. *Neuron*, *30*(2), 319–333. http://doi.org/10.1016/S0896-6273(01)00285-9

Gallagher, H. L., Jack, A. I., Roepstorff, A., & Frith, C. D. (2002). Imaging the intentional stance in a competitive game. *NeuroImage*, *16*, 814–821. http://doi.org/10.1006/nimg.2002.1117

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain☐: A Journal of Neurology*, *119 (Pt 2*, 593–609. http://doi.org/10.1093/brain/119.2.593

George, N., Driver, J., & Dolan, R. J. (2001). Seen gaze-direction modulates fusiform activity and its coupling with other brain areas during face processing. *NeuroImage*, *13*(6 Pt 1), 1102–1112. http://doi.org/10.1006/nimg.2001.0769

Gómez, C. M., Vaquero, E., & Vázquez-Marrufo, M. (2004). A neurocognitive model for short-term sensory and motor preparatory activity in humans. *Psicologica*. http://doi.org/p217-229 2004

Greene, D. J., & Zaidel, E. (2012). Spatial orienting of attention simultaneously cued by automatic social and nonsocial cues. *Experimental Brain Research. Experimentelle Hirnforschung. Expérimentation Cérébrale*, *221*(1), 115–22. http://doi.org/10.1007/s00221-012-3152-2

Greene, D. J., & Zaidel, E. (2011). Hemispheric differences in attentional orienting by social cues. *Neuropsychologia*, *49*(1), 61–8. http://doi.org/10.1016/j.neuropsychologia.2010.11.007

Grèzes, J., Frith, C., & Passingham, R. E. (2004). Brain mechanisms for inferring deceit in the actions of others. *The Journal of Neuroscience☐: The Official Journal of the Society for Neuroscience*, *24*(24), 5500–5505. http://doi.org/10.1523/JNEUROSCI.0219-04.2004

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2002). Human neural systems for face recognition and social communication. *Biological Psychiatry*, *51*(1), 59–67. http://doi.org/10.1016/S0006-3223(01)01330-0

Haxby, J., Hoffman, E., & Gobbini, M. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*(6), 223–233. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10827445

Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision*, *3*(1), 49–63. http://doi.org/10.1167/3.1.6

Heider, F., & Simmel, M. (1944). An Experimental Study of Apparent Behavior. *The American Journal of Psychology*, *57*(2), 243–259. http://doi.org/10.1017/CBO9781107415324.004

Heywood, C. A., & Cowey, A. (1992). The role of the “face-cell” area in the discrimination and recognition of faces by monkeys. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences*, *335*(1273), 31–38. http://doi.org/10.1098/rstb.1992.0004

Hietanen, J. K. (1999). *Does your gaze direction and head orientation shift my visual attention? Neuroreport* (Vol. 10). http://doi.org/10.1097/00001756-199911080-00033

Hirai, M., Senju, A., Fukushima, H., & Hiraki, K. (2005). Active processing of biological motion perception: An ERP study. *Cognitive Brain Research*, *23*(2-3), 387–396. http://doi.org/10.1016/j.cogbrainres.2004.11.005

Hoffman, E. A., & Haxby, J. V. (2000). *Distinct representations of eye gaze and identity in the distributed human neural system for face perception. Nature neuroscience* (Vol. 3). http://doi.org/10.1038/71152

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The Theory of Event Coding (TEC): a framework for perception and action planning. *The Behavioral and Brain Sciences*, *24*(5), 849–878; discussion 878–937. http://doi.org/10.1017/S0140525X01000103

Hommel, B. (2004). Event les: feature binding in and across perception and action. *Trends in Cognitive Sciences*, *8*(11). http://doi.org/10.1016/j.tics.2004.08.007

Hommel, B. (1998). Event Files: Evidence for Automatic Integration of Stimulus-Response Episodes. *Visual Cognition*, *5*(1-2), 183–216. http://doi.org/10.1080/713756773

Hood, B. M., Willen, J. D., & Driver, J. (1998). Adult ’ S Eyes Trigger Shifts of Visual Attention in Human Infants. *Psychological Science*, *9*(2), 131–134. http://doi.org/10.1111/1467-9280.00024

Hooker, C. I., Paller, K. A., Gitelman, D. R., Parrish, T. B., Mesulam, M. M., & Reber, P. J. (2003). Brain networks for analyzing eye gaze. *Cognitive Brain Research*, *17*(2), 406–418. http://doi.org/10.1016/S0926-6410(03)00143-5

Itier, R. J., & Batty, M. (2009). Neural bases of eye and gaze processing: The core of social cognition. *Neuroscience and Biobehavioral Reviews*. http://doi.org/10.1016/j.neubiorev.2009.02.004

Jeannerod, M. (2001). Neural Simulation of Action: A Unifying Mechanism for Motor Cognition. *NeuroImage*, *14*(1), S103–S109. http://doi.org/10.1006/nimg.2001.0832

Jeannerod, M. (2006). *Motor Cognition*. *Motor Cognition: What Actions Tell the Self*. http://doi.org/10.1093/acprof:oso/9780198569657.001.0001

Jokisch, D., Daum, I., Suchan, B., & Troje, N. F. (2005). Structural encoding and recognition of biological motion: Evidence from event-related potentials and source analysis. *Behavioural Brain Research*, *157*(2), 195–204. http://doi.org/10.1016/j.bbr.2004.06.025

Jonides, J. (1998). Voluntary versus automatic control over the mind’s eye's movement. *Psychonomic Bulletin & Review*. http://doi.org/10.1037/0096-1523.29.5.835

Kanwisher, N. G. (2000). Domain specificity in face perception. *Nature Neuroscience*, *3*(8), 759–763. http://doi.org/10.1038/77664

Keller, P. E., & Koch, I. (2008). Action planning in sequential skills: relations to music performance. *Quarterly Journal of Experimental Psychology*, *61*(2), 275–291. http://doi.org/10.1080/17470210601160864

Kilner, J. M., Vargas, C., Duval, S., Blakemore, S.-J., & Sirigu, A. (2004). Motor activation prior to observation of a predicted movement. *Nature Neuroscience*, *7*(12), 1299–1301. http://doi.org/10.1038/nn1355

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: An account of the mirror neuron system. *Cognitive Processing*. http://doi.org/10.1007/s10339-007-0170-2

Kingstone, A., Friesen, C. K., & Gazzaniga, M. S. (2000). Reflexive Joint Attention Depends on Lateralized Cortical Connections. *Psychological Science*, *11*(2), 159–166. http://doi.org/10.1111/1467-9280.00232

Kingstone, A. (1992). Combining Expectancies. *The Quarterly Journal of Experimental Psychology*, *44*(September), 69–104. http://doi.org/10.1080/14640749208401284

Kingstone, A., Tipper, C., Ristic, J., & Ngan, E. (2004). The eyes have it!: An fMRI investigation. *Brain and Cognition*, *55*(2), 269–271. http://doi.org/10.1016/j.bandc.2004.02.037

Klein, R., & Hansen, E. (1990). Chronometric analysis of apparent spotlight failure in endogenous visual orienting. *Journal of Experimental Psychology. Human Perception and Performance*, *16*(4), 790–801. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2148592

Klein, R. M., Kingstone, A., & Pontefract, A. (1992). Orienting of visual attention. *Eye Movements and Visual Cognition*, (1967), 46–65. http://doi.org/10.1007/978-1-4612-2852-3_4

Klein, R. M., & Shore, D. . (2000). Relations among modesof visual orienting. *Attention & Performance XVIII: Control of Cognivite Processes*, 195–208.

Klingstone, a, & Klein, R. (1991). Combining shape and position expectancies: hierarchical processing and selective inhibition. *Journal of Experimental Psychology. Human Perception and Performance*, *17*(2), 512–519. http://doi.org/10.1037/0096-1523.17.2.512

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*. http://doi.org/10.1016/j.tins.2004.10.007

Knutson, B., & Cooper, J. C. (2005). Functional magnetic resonance imaging of reward prediction. *Current Opinion in Neurology*, *18*(4), 411–7. http://doi.org/10.1097/01.wco.0000173463.24758.f6

Kotz, S. A., Schwartze, M., & Schmidt-Kassow, M. (2009). Non-motor basal ganglia functions: A review and proposal for a model of sensory predictability in auditory language perception. *Cortex*, *45*(8), 982–990. http://doi.org/10.1016/j.cortex.2009.02.010

Kveraga, K., Boshyan, J., & Bar, M. (2007). Magnocellular projections as the trigger of top-down facilitation in recognition. *The Journal of Neuroscience□: The Official Journal of the Society for Neuroscience*, *27*(48), 13232–40. http://doi.org/10.1523/JNEUROSCI.3481-07.2007

Langton, L., & Bruce, B. (1999). Reflexive Visual Orienting in Response to the Social Attention of Others. *Visual Cognition*, *6*(5), 541–567. http://doi.org/10.1080/135062899394939

Luck, S. J., & Hillyard, S. A. (1995). The role of attention in feature detection and conjunction discrimination: an electrophysiological analysis. *The International Journal of Neuroscience*, *80*(1-4), 281–297. http://doi.org/10.3109/00207459508986105

Mangun, G. R. (1995). Neural mechanisms of visual selective attention. *Psychophysiology*, *32*(1), 4–18. http://doi.org/10.1111/j.1469-8986.1995.tb03400.x

Mangun, G. R., & Hillyard, S. A. (1991). Modulations of sensory-evoked brain potentials indicate changes in perceptual processing during visual-spatial priming. *Journal of Experimental*

Psychology. Human Perception and Performance, *17*(4), 1057–1074. http://doi.org/10.1037/0096-1523.17.4.1057

Manthey, S., Schubotz, R. I., & von Cramon, D. Y. (2003). Premotor cortex in observing erroneous action: an fMRI study. *Brain Research. Cognitive Brain Research*, *15*(3), 296–307. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12527103

Martin, A., & Weisberg, J. (2003). Neural foundations for understanding social and mechanical concepts. *Cognitive Neuropsychology*, *20*(3-6), 575–87. http://doi.org/10.1080/02643290342000005

Mattler, U. (2003). Combined perceptual or motor-related expectancies modulated by type of cue. *Perception & Psychophysics*, *65*(4), 649–666. http://doi.org/10.3758/BF03194589

Mattler, U. (2005). Combined expectancy effects: An accumulator model. *Cognitive Psychology*, *51*(3), 214–255. http://doi.org/10.1016/j.cogpsych.2005.05.002

Mattler, U., van der Lugt, A., & Münte, T. F. (2006). Combined expectancies: electrophysiological evidence for the adjustment of expectancy effects. *BMC Neuroscience*, *7*, 37. http://doi.org/10.1186/1471-2202-7-37

McCabe, K., Houser, D., Ryan, L., Smith, V., & Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(20), 11832–11835. http://doi.org/10.1073/pnas.211415698

Mehta, B., & Schaal, S. (2002). Forward models in visuomotor control. *Journal of Neurophysiology*, *88*(2), 942–953. http://doi.org/10.1152/jn.00804.2001

Miall, R. C., & Wolpert, D. M. (1996). Forward models for physiological motor control. *Neural Networks*. http://doi.org/10.1016/S0893-6080(96)00035-4

Müller, H. J., & Rabbitt, P. M. A. (1989). Reflexive and Voluntary Orienting of Visual Attention□: Time Course of Activation and Resistance to Interruption. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(2), 315–330. http://doi.org/10.1037/0096-1523.15.2.315

Nobre, A. C., Sebestyen, G. N., Gitelman, D. R., Mesulam, M. M., Frackowiak, R. S. J., & Frith, C. D. (1997). Functional localization of the system for visuospatial attention using positron emission tomography. *Brain*, *120*(3), 515–533. http://doi.org/10.1093/brain/120.3.515

Parasuraman, R. (1980). Effects of information processing demands on slow negative shift latencies and N100 amplitude in selective and divided attention. *Biological Psychology*, *11*(3-4), 217–233. http://doi.org/10.1016/0301-0511(80)90057-5

Pelphrey, K., Singerman, J., Allison, T., & McCarthy, G. (2003). Brain activation evoked by perception of gaze shifts: the influence of context. *Neuropsychologia*, *41*, 156–170. Retrieved from http://www.sciencedirect.com/science/article/pii/S002839320200146X

Pelphrey, K. A., Morris, J. P., Michelich, C. R., Allison, T., & McCarthy, G. (2005). Functional anatomy of biological motion perception in posterior temporal cortex: An fMRI study of eye, mouth and hand movements. *Cerebral Cortex*, *15*(12), 1866–1876. http://doi.org/10.1093/cercor/bhi064

Pelphrey, K. A., Singerman, J. D., Allison, T., & McCarthy, G. (2003). Brain activation evoked by perception of gaze shifts: The influence of context. *Neuropsychologia*, *41*(2), 156–170. http://doi.org/10.1016/S0028-3932(02)00146-X

Perrett, D. I., Hietanen, J. K., Oram, M. W., & Benson, P. J. (1992). Organization and functions of cells responsive to faces in the temporal cortex. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *335*(1273), 23–30. http://doi.org/10.1098/rstb.1992.0003

Perrett, D. I., Smith, P. a, Potter, D. D., Mistlin, a J., Head, a S., Milner, a D., & Jeeves, M. a. (1985). Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character. Royal Society (Great Britain)*, *223*(1232), 293–317. http://doi.org/10.1098/rspb.1985.0003

Perrett, D. I., Harries, M. H., Mistlin, A. J., Hietanen, J. K., Benson, P. J., Bevan, R., ... Brierley, K. (1990). Social signals analyzed at the single cell level: someone is looking at me, something touched me, something moved! *International Journal of Comparative Psychology*, *4*(1), 25–55. http://doi.org/10.5811/westjem.2011.5.6700

Perrett, D. I., & Mistlin, A. J. (1990). Perception of facial characteristics by monkeys. In *Comparative Perception, Vol. 2: Complex Signals* (pp. 187–215).

Perrett, D. I. I., & Emery, N. J. J. (1994). Understanding the intentions of others from visual signals: Neurophysiological evidence. *Current Psychology of Cognition*, *13*(5), 683–694. Retrieved from http://doi.apa.org/psycinfo/1995-24608-001

Ploghaus, a, Tracey, I., Gati, J. S., Clare, S., Menon, R. S., Matthews, P. M., & Rawlins, J. N. (1999). Dissociating pain from its anticipation in the human brain. *Science (New York, N.Y.)*, *284*(5422), 1979–1981. http://doi.org/10.1126/science.284.5422.1979

Porro, C. A., Cettolo, V., Francescato, M. P., & Baraldi, P. (2003). Functional activity mapping of the mesial hemispheric wall during anticipation of pain. *NeuroImage*, *19*(4), 1738–1747. http://doi.org/10.1016/S1053-8119(03)00184-8

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*, 3–25. http://doi.org/10.1080/00335558008248231

Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*(2), 160–174. http://doi.org/10.1037/0096-3445.109.2.160

Premack, D., & Woodruff, G. (1978). Premack and Woodruff□: Chimpanzee theory of mind. *Behavioral and Brain Sciences*, *4*, 515–526.

Puce, a, Allison, T., Bentin, S., Gore, J. C., & McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. *The Journal of Neuroscience□: The Official Journal of the Society for Neuroscience*, *18*(6), 2188–2199.

Quadflieg, S., Mason, M. F., & Macrae, C. N. (2004). The owl and the pussycat: Gaze cues and visuospatial orienting. *Psychonomic Bulletin & Review*, *11*(5), 826–831. http://doi.org/10.3758/BF03196708

Rangelov, D., Müller, H. J., & Zehetleitner, M. (2012). The multiple-weighting-systems hypothesis: Theory and empirical support. *Attention, Perception, & Psychophysics*, *74*(3), 540–552. http://doi.org/10.3758/s13414-011-0251-2

Rilling, J. K., Sanfey, A. G., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2004). The neural correlates of theory of mind within interpersonal interactions. *NeuroImage*, *22*(4), 1694–1703. http://doi.org/10.1016/j.neuroimage.2004.04.015

Ristic, J., & Kingstone, A. (2005). Taking control of reflexive social attention. *Cognition*, *94*(3), B55–65. http://doi.org/10.1016/j.cognition.2004.04.005

Ritter, W., Simson, R., & Vaughan, H. G. (1983). Event-related potential correlates of two stages of information processing in physical and semantic discrimination tasks. *Psychophysiology*, *20*(2), 168–179. http://doi.org/10.1111/j.1469-8986.1983.tb03283.x

Ritter, W., Simson, R., Vaughan, H. G., & Macht, M. (1982). Manipulation of event-related potential manifestations of information processing stages. *Science (New York, N.Y.)*, *218*(4575), 909–911. http://doi.org/10.1126/science.7134983

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*, 169–192. http://doi.org/10.1146/annurev.neuro.27.070203.144230

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews. Neuroscience*, *2*(9), 661–670. http://doi.org/10.1038/35090060

Rolfs, M., Jonikaitis, D., Deubel, H., & Cavanagh, P. (2010). Predictive remapping of attention across eye movements. *Nature Neuroscience*, *14*(2), 252–256. http://doi.org/10.1038/nn.2711

Saxe, R., Xiao, D. K., Kovacs, G., Perrett, D. I., & Kanwisher, N. (2004). A region of right posterior superior temporal sulcus responds to observed intentional actions. *Neuropsychologia*, *42*(11), 1435–1446. http://doi.org/10.1016/j.neuropsychologia.2004.04.015

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: the prospective brain. *Nature Reviews Neuroscience*, *8*(9), 657–661. http://doi.org/10.1080/08995600802554748

Schacter, D. L., & Addis, D. R. (2011). On the Nature of Medial Temporal Lobe Contributions to the Constructive Simulation of Future Events. In *Predictions in the Brain* (pp. 58–69). Oxford University Press. http://doi.org/10.1093/acprof:oso/9780195395518.003.0024

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, *36*(4), 393–414. http://doi.org/10.1017/S0140525X12000660

Schubotz, R. I., & von Cramon, D. Y. (2002). Predicting perceptual events activates corresponding motor schemes in lateral premotor cortex: an fMRI study. *NeuroImage*, *15*(4), 787–96. http://doi.org/10.1006/nimg.2001.1043

Schubotz, R. I. (2007). Prediction of external events with our motor system: towards a new framework. *Trends in Cognitive Sciences*, *11*(5), 211–218. http://doi.org/10.1016/j.tics.2007.02.006

Schubotz, R. I., & Von Cramon, D. Y. (2003). Functional-anatomical concepts of human premotor cortex: Evidence from fMRI and PET studies. In *NeuroImage* (Vol. 20). http://doi.org/10.1016/j.neuroimage.2003.09.014

Schuller, a M., & Rossion, B. (2001). Spatial attention triggered by eye gaze increases and speeds up early visual activity. *Neuroreport*, *12*(11), 2381–2386. http://doi.org/10.1097/00001756-200108080-00019

Schultz, W., & Dickinson, A. (2000). Neuronal coding of prediction errors. *Program*, *23*, 473–500. http://doi.org/10.1146/annurev.neuro.23.1.473

Schyns, P. G., Jentzsch, I., Johnson, M., Schweinberger, S. R., & Gosselin, F. (2003). A principled method for determining the functionality of brain responses. *Neuroreport*, *14*(13), 1665–9. http://doi.org/10.1097/01.wnr.0000088408.04452.e9

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in Cognitive Sciences*, *10*(2), 70–6. http://doi.org/10.1016/j.tics.2005.12.009

Sebanz, N., Knoblich, G., & Prinz, W. (2003). Representing others' actions: Just like one's own? *Cognition*, *88*(3). http://doi.org/10.1016/S0010-0277(03)00043-X

Sebanz, N., & Knoblich, G. (2009). Prediction in Joint Action: What, When, and Where. *Topics in Cognitive Science*, *1*(2), 353–367. http://doi.org/10.1111/j.1756-8765.2009.01024.x

Senju, A., Csibra, G., & Johnson, M. H. (2008). Understanding the referential nature of looking: infants' preference for object-directed gaze. *Cognition*, *108*(2), 303–19. http://doi.org/10.1016/j.cognition.2008.02.009

Senju, A., Johnson, M. H., & Csibra, G. (2006). The development and neural basis of referential gaze perception. *Social Neuroscience*, *1*(3-4), 220–34. http://doi.org/10.1080/17470910600989797

Senju, A., Johnson, M. H., & Csibra, G. (2006). The development and neural basis of referential gaze perception. *Social Neuroscience*, *1*(3-4), 220–34. http://doi.org/10.1080/17470910600989797

Senkowski, D., & Herrmann, C. S. (2002). *Effects of task difficulty on evoked gamma activity and ERPs in a visual discrimination task. Clinical neurophysiology□: official journal of the International Federation of Clinical Neurophysiology* (Vol. 113). http://doi.org/10.1016/S1388-2457(02)00266-3

Taylor, M. J., Itier, R. J., Allison, T., & Edmonds, G. E. (2001). Direction of gaze effects on early face processing: Eyes-only versus full faces. *Cognitive Brain Research*, *10*(3), 333–340. http://doi.org/10.1016/S0926-6410(00)00051-3

Teufel, C., Alexis, D. M., Clayton, N. S., & Davis, G. (2010). Mental-state attribution drives rapid, reflexive gaze following. *Attention, Perception & Psychophysics*, *72*(3), 695–705. http://doi.org/10.3758/APP.72.3.695

Teufel, C., Fletcher, P. C., & Davis, G. (2010). Seeing other minds: attributed mental states influence perception. *Trends in Cognitive Sciences*, *14*(8), 376–82. http://doi.org/10.1016/j.tics.2010.05.005

Thomas, K. M., Drevets, W. C., Whalen, P. J., Eccard, C. H., Dahl, R. E., Ryan, N. D., & Casey, B. J. (2001). Amygdala response to facial expressions in children and adults. *Biological Psychiatry*, *49*(4), 309–316. http://doi.org/10.1016/S0006-3223(00)01066-0

Tipples, J., Johnston, P., & Mayes, A. (2012). Electrophysiological responses to violations of expectation from eye gaze and arrow cues. *Social Cognitive and Affective Neuroscience*, *8*(5), 509–14. http://doi.org/10.1093/scan/nss024

Tollner, T., Gramann, K., Muller, H. J., Kiss, M., & Eimer, M. (2008). Electrophysiological markers of visual dimension changes and response changes. *J Exp Psychol Hum Percept Perform*, *34*(3), 531–542. http://doi.org/10.1037/0096-1523.34.3.531

Tversky, B., & Hard, B. M. (2009). Embodied and disembodied cognition: Spatial perspective-taking. *Cognition*, *110*(1), 124–129. http://doi.org/10.1016/j.cognition.2008.10.008

Ueda, K., Okamoto, Y., Okada, G., Yamashita, H., Hori, T., & Yamawaki, S. (2003). Brain activity during expectancy of emotional stimuli: an fMRI study. *Neuroreport*, *14*(1), 51–55. http://doi.org/10.1097/00001756-200301200-00010

Umiltà, M. A., Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C., & Rizzolatti, G. (2001). I know what you are doing: A neurophysiological study. *Neuron*, *31*(1), 155–165. http://doi.org/10.1016/S0896-6273(01)00337-3

Vecera, S. P., & Johnson, M. H. (1995). Gaze detection and the cortical processing of faces: Evidence from infants and adults. *Visual Cognition*, *2*(1), 59–87. http://doi.org/10.1080/13506289508401722

Verfaillie, K., & Daems, A. (2002). Representing and anticipating human actions in vision. *Visual Cognition*, *9*(1-2), 217–232. http://doi.org/10.1080/13506280143000403

Vogel, E. K., & Luck, S. J. (2000). The visual N1 component as an index of a discrimination process. *Psychophysiology*, *37*(2), 190–203. http://doi.org/10.1111/1469-8986.3720190

Vuilleumier, P., George, N., Lister, V., Armony, J., & Driver, J. (2005). Effects of perceived mutual gaze and gender on face processing and recognition memory, *12*(1), 85–102. http://doi.org/10.1080/13506280444000120

Wicker, B., Michel, F., Henaff, M. a, & Decety, J. (1998). Brain regions involved in the perception of gaze: a PET study. *NeuroImage*, *8*(2), 221–7. http://doi.org/10.1006/nimg.1998.0357

Wiese, E., Wykowska, A., Zwickel, J., & Müller, H. J. (2012). I See What You Mean: How Attentional Selection Is Shaped by Ascribing Intentions to Others. *PLoS ONE*, *7*(9). http://doi.org/10.1371/journal.pone.0045391

Wiese, E., Zwickel, J., & Müller, H. J. (2013). The importance of context information for the spatial specificity of gaze cueing. *Attention, Perception & Psychophysics*, *75*(5), 967–82. http://doi.org/10.3758/s13414-013-0444-y

Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society of London.*

Series B, Biological Sciences, *358*(1431), 593–602. http://doi.org/10.1098/rstb.2002.1238

Wolpert, D. M., & Flanagan, J. R. (2001). Motor prediction. *Current Biology□: CB*, *11*(18), R729–R732. http://doi.org/10.1016/S0960-9822(01)00432-8

Wurm, M. F., & Schubotz, R. I. (2012). Squeezing lemons in the bathroom: contextual information modulates action recognition. *NeuroImage*, *59*(2), 1551–9. http://doi.org/10.1016/j.neuroimage.2011.08.038

Wykowska, A., Wiese, E., Prosser, A., & Müller, H. J. (2014). Beliefs about the minds of others influence how we process sensory information. *PloS One*, *9*(4), e94339. http://doi.org/10.1371/journal.pone.0094339

Wylie, G. R., Javitt, D. C., & Foxe, J. J. (2006). Jumping the gun: Is effective preparation contingent upon anticipatory activation in task-relevant neural circuitry? *Cerebral Cortex*, *16*(3), 394–404. http://doi.org/10.1093/cercor/bhi118

APPENDIX

Jairo Perez-Osorio

Education

2012 – present	**Doctoral Student** Ph.D. in Systemic Neurosciences Graduate School of Systemic Neurosciences, Ludwig Maximilian University, Munich Project name: Visual attention in social contexts. Supervised by Prof. Dr. Agnieszka Wykowska, Prof. Dr. Hermann Müller.
2010 – 2012	**M.Sc., Neuro-Cognitive Psychology** Department of Psychology Ludwig Maximilian University, Munich Overall qualification: 1.77 (1 is the highest)
2000 – 2005	**Psychologist** National University of Colombia, Bogotá, Colombia ECAES: 123.4/140 (State quality exam for superior education)

Professional Experience

08.2014 – present	Studentische/Wissenschaftliche Hilfskraft. Parmenides Fundation • Experiment design, behavioral and fMRI data collection and analysis.
12.2011 – 11.2012	Studentische/Wissenschaftliche Hilfskraft. Ludwig Maximilians University, Munich. • Coordination EEG lab . Design, program, collect and analyze data for behavioral and EEG/ERPs experiments.
03.2009 – 11.2010	Clinical Psychologist, Emotional Support Lifeline, Health Department, Colombian National Police, Bogotá. • Emotional crisis telephone attention to police men and their families. Active listening and focus on costumer were necessary. As well participated actively in design of the National Mental Health Policy, and the National Police's II Study of Mental Health in Colombia, as a co-researcher.

10.2008 – 06.2009 Psychosocial support - Construction of a clothing plant in Bogotá with demobilized people, Ap&p Colombia.

- The aim was to provide educational and psychosocial support for ex-combatants from Colombia's internal conflict, during the development of a productive project.

10 – 12.2008 Raising psychosocial and group profiles of beneficiaries and receptor community of demobilized people in Bogotá – Pilot Study, Ap&p Colombia.

- The objective was to determine psychosocial profiles to locate participants (ex- combatants from Colombia's internal conflict) in adequate productive projects. Financed by Japanese International Cooperation Agency (JICA).

03.2006 – 06.2008 Psychological Promotion and Prevention (P&P) Leader
Colombian National Police, Arauca Police Department, Arauca, Colombia.

- This work implied the development and execution of P&P subprograms (Social Skills, Family strengthening and Attention to sociopolitical violence victims), as well as Clinical Psychology consulting.

01.2004 – 12.2004 Neuropsychology Assessment Psychology (Internship)
Bogotá, Colombia.

- The job involved the application of test batteries to a wide spectrum of patients for a whole year. From this experience I gathered the information and knowledge required to write my thesis.

Key skills

Software

- Competent user of text, spreadsheet, presentation, photo editing processing programs
- Proficient user of SPSS, E-prime, Matlab (SPM), Brain Vision (Recorder and Analyzer)

Languages

- Spanish (mother tongue), advanced English, and basic German.

Academic references

Prof. Dr. Agnieszka Wykowska
LuleP University of Technology
Department of Business Administration, Technology and Social Sciences
Human Work Science, Engineering Psychology,
971 87 LuleP
Phone: +46 (0)920 493534
Email: agnieszka.wykowska(at)ltu.se

Dipl.-Psych. Dr. Michael Öllinger
Parmenides Center for the Study of Thinking
Kirchplatz 1, 82049, Pullach im Isartal
Phone.: +49 (0) 89 4520935 23
Email: michael.oellinger(at)parmenides-foundation.org

Prof. Dr. Hermann J. Müller
Department of General and Experimental psychology, LMU Munich
Leopoldstraße 13, 80802 München
Phone: +49 (0) 89 / 2180 5327
Email: hmueller(at)psy.lmu.de

Author contributions

- **Perez-Osorio, J.**, Müller, H.J., Wiese, E., & Wykowska, A. (2015) Gaze Following Is Modulated by Expectations Regarding Others' Action Goals. *PLoS ONE 10(11):* e0143614. doi:10.1371/journal. Pone.0143614
    - AW JP HJM, Conceived and designed the experiments. JP, Performed the experiments. JP AW, Analyzed the data. JP AW HJM EW, Wrote the paper.

- **Perez-Osorio, J.**, Müller, H.J., & Wykowska, A. (under review) Expectations regarding action sequences modulate electrophysiological correlates of the gaze-cueing effect. *Psychophysiology.*
    - AW JP HJM, Conceived and designed the experiments. JP, Performed the experiments. JP AW, Analyzed the data. JP AW HJM, Wrote the paper.

Publications

Published:

- Leszczyński, M., Wykowska, A., **Perez-Osorio, J.**, & Müller, H.J. (2013) Deployment of Spatial Attention towards Locations in Memory Representations. An EEG Study. *PLoS ONE 8(12):* e83856. doi:10.1371/journal.pone.0083856
- **Perez-Osorio, J.**, Müller, H.J., Wiese, E., & Wykowska, A. (2015) Gaze Following Is Modulated by Expectations Regarding Others' Action Goals. *PLoS ONE 10(11):* e0143614. doi:10.1371/journal. pone.0143614

Under review:

- **Perez-Osorio, J.**, Müller, H.J., & Wykowska, A. (under review) Expectations regarding action sequences modulate electrophysiological correlates of the gaze-cueing effect. *Psychophysiology*.

EIDESSTATTLICHE VERSICHERUNG/AFFIDAVIT

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation **Visual attention in social contexts** selbstständig angefertigt habe, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

I hereby confirm that the dissertation **Visual attention in social contexts** is the result of my own work and that I have only used sources or materials listed and specified in the dissertation.

München, den

Munich, date Jairo Perez-Osorio