Proceedings of the Human Factors and Ergonomics Society Annual Meeting
http://pro.sagepub.com/

Making eyes with robots: Readiness to engage in Human-Robot-Interaction depends on the attribution of intentionality

E. Wiese, A. Wykowska and H.J. Müller

Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2013 57: 1174
DOI: 10.1177/1541931213571261

The online version of this article can be found at:
http://pro.sagepub.com/content/57/1/1174

Published by:
SAGE
http://www.sagepublications.com

On behalf of:
Human Factors and Ergonomics Society

Additional services and information for *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* can be found at:

Email Alerts: http://pro.sagepub.com/cgi/alerts

Subscriptions: http://pro.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations: http://pro.sagepub.com/content/57/1/1174.refs.html

>> Version of Record - Sep 30, 2013

What is This?

Downloaded from pro.sagepub.com at TU Muenchen on December 15, 2013

Making eyes with robots: Readiness to engage in Human-Robot-Interaction depends on the attribution of intentionality

Wiese, E.[1,2], Wykowska, A.[1], & Müller, H.J.[1,3]
[1] Department of Experimental Psychology, Ludwig Maximilian University Munich,
Leopoldstr. 13, 80802 Munich, Germany
[2] Department of Psychology and Human Factors, George Mason University,
2055 David King Hall, 4400 University Drive, Fairfax, VA 22030, USA
[3] Birkbeck College, Department of Psychological Sciences, University of London,
Malet Street, London WC1E 7HX, UK

ABSTRACT. One of the most important aims of social robotics is to improve Human-Robot Interaction by providing robots with means to understand observed behavior and to predict upcoming actions of their interaction partners. The most reliable source for inferring the action goals of interaction partners is their gaze direction. Hence, to anticipate upcoming actions, it is necessary to identify where others are currently looking at and to shift the attentional focus to the same location. Interestingly, it has been shown that observing robot gaze direction also induces attentional shifts to the location that is gazed-at by the robot. Given this, gaze direction can be actively used by the robot to direct the attentional focus of interaction partners to important events in the world. In this paper, we review findings from two studies indicating that the readiness to engage attentional resources in interactions with robots is modulated by the degree to which intentionality can be attributed to the robot: Robots believed to behave similar to humans cause stronger gaze-cueing effects than robots perceived as machines, independently of their physical appearance. Based on these findings, and on results from a pilot study with a sample of patients diagnosed with autism spectrum disorder (ASD), we derive guidelines for improving human-robot interaction by emulating social gaze behavior.

Copyright 2013 by Human Factors and Ergonomics Society, Inc. All rights reserved. DOI 10.1177/1541931213571261

INTRODUCTION

Knowing where others are looking is an important social skill and a prerequisite for performing actions together with others. Attending to gaze direction helps us building up shared representations of events in the world and enables us to anticipate the others' action goals - as gaze indicates the location of objects of interest and determines what might be the target of subsequent actions. Thus, in order to anticipate upcoming actions of other people, it is helpful to identify where they are looking at and to shift our own attentional focus to the same location – a mechanism that is referred to as shared attention.

The ability to share attention with others develops early in life (Striano & Reid, 2006) and is thought to be the foundation for developing a **theory of mind** (Baron-Cohen, 1995), that is: the ability to reason about others' mental states, intentions and motivations. This process of **mentalizing** (Frith & Frith, 2006) enables us to make predictions about the behavior of others by combining information we deduce from the current state of the world with assumptions about their internal states. Thus, the interpretation of social content is thought to involve a bottom-up mechanism triggered by perceptual information in the environment, and a top-down mechanism that is activated based on background knowledge we have about others or deduce from perceived information (Frith & Frith, 2006). Thanks to this top-down mechanism, the human brain accounts for the social relevance of objects/events in the environment and modulates bottom-up responses that are reflexively evoked based on the perceptual features of the scene.

When we try to understand others, a network consisting of medial prefrontal cortex (mPFC), amygdala, superior temporal sulcus (STS), and intraparietal sulcus (IPS) is activated (Frith & Frith, 2006, for a review). The STS is coding for biological movements, such as head and gaze direction or pointing gestures and is more responsive to meaningful compared to random changes in gaze direction (Pelphrey, Singerman, Allison & McCarthy, 2003). The emotional content of perceived information is processed via reciprocal projections between STS and the amygdala, a structure of the limbic system involved in the processing of facial expression and emotions (Aggleton, Burton & Passingham, 1980). The STS is also involved in orienting of attentional resources via its connections with the IPS (Corbetta, Miezin, Shulman & Petersen, 1993). That way, perception of others' gaze direction can be linked to the spatial attention system, in order to initiate orienting of attentional resources to gazed-at locations.

In this network, the attribution of intentionality to observed gaze behavior is exerted by the mPFC through top-down modulation of the signals sent by the STS and IPS (Grezes, Frith & Passingham, 2004). In line with that, activity in the mPFC has been shown to be positively correlated with the amount of background knowledge one has about others (Saxe & Wechsler, 2005, Frith & Frith, 2006). Given this, the mPFC region seems to be crucial for aligning visual information in the environment with the observer's background knowledge and is thought to modulate the perceived information according to its social relevance.

Downloaded from pro.sagepub.com at TU Muenchen on December 15, 2013

Importantly, the operation of the top-down component has been shown to critically depend on whether or not changes in gaze direction are presented in a social context (Frith & Frith, 2003; Teufel, Fletcher & Davis, 2010). Context information can either be provided *visually* in the environment or *verbally* by inducing certain beliefs in participants about the observed person. Regarding visual context information, Wiese, Zwickel and Müller (accepted) showed that attentional allocation was modulated depending on the availability of peripheral objects to which changes in gaze direction could refer: spatial attention was deployed to the whole gaze-cued hemifield when no context was provided, whereas shifts of attention were specific to the gazed-at position in conditions when reference objects were presented. Consistent with this, Kawai (2011) found that gaze shifted attention only when the target seemed to be visible to the gazer, but not when it was occluded by a bar that was placed in between the gazer and the target.

With regard to verbal information, preexisting assumptions concerning the observed person have been shown to influence the readiness to attend to gaze direction (Ristic & Kingstone, 2005; Teufel et al., 2010). Teufel et al. (2010) reported in that context that adaptation to gaze direction can only be observed when participants believed that the gazer could see through a pair of goggles compared to when they believed that he could not. Consistent with this, Wiese, Wykowska and Müller (submitted) showed that the readiness to shift attention to gazed-at locations strongly depended on the trustworthiness of the gazer and that stronger attentional facilitation was induced when changes in gaze direction were believed to be predictive (i.e., indicated the target position with a high likelihood) compared to when believed to be nonpredictive. Analogously, Ristic and Kingstone (2005) presented an ambiguous stimulus that participants could perceive either as a car or a face: Their attention was cued when they believed that they were looking at a face, but not when they believed that the stimulus represented a car.

The importance of emulating social gaze behavior in HRI

Given that gaze direction reveals important information about the motivational and emotional states of others in a natural and straight-forward way, it is important to provide robots with an internal model that they can use to interpret others' gaze direction and to communicate their own action goals.

In order to implement gaze as means for communication in human-robot interaction, it is essential to investigate how humans react to robot gaze and whether responses to changes in gaze direction are comparable to those induced by human interaction partners. That is, in order to establish robots as social interaction partners, it is necessary to identify a set of parameters that determine under which conditions a robot is perceived as intentional agent, given that humans seem to be more willing to engage their attentional resources when they are able to interpret observed behavior in an intentional manner.

Therefore, the paper reviews findings of two previous experiments and reports results of a pilot study with ASD patients on how assumptions about robots' gaze behavior modulate the allocation of social attention. Furthermore, implications for the use of gaze control in HRI are discussed.

Attributing intentionality modulates engagement in HRI

For humans, there exists an essential difference between being engaged in social interactions with other human individuals vs. interacting with robots: When interacting with humans, we treat them as intentional systems and adopt an *intentional stance*: that is, we explain their behavior with reference to their internal states, such as desires, intentions or motivations (Dennett, 2003). A robot, by contrast, is not characterized by having mental states and its behavior is explained/predicted as mechanistic rather than intentional. As a result, the *design stance* is adopted to robots and their behavior is explained on the basis of how they have been designed to behave (i.e., being pre-programmed).

In two studies, we investigated whether the belief that an observed stimulus is representing an intentional agent influences the readiness to engage in social interactions with that particular agent and whether this effect depends on the physical appearance of the agent. We assumed that adopting the intentional stance would lead to an increased attentional engagement compared to when the design stance is adopted. Importantly, we expected this effect to be independent of the physical appearance of the gazer, but to depend only on the beliefs participants have about the gazer's internal states. The likelihood of adopting an intentional stance was manipulated by instruction.

Study 1: Behavioral findings

In the first study (see Wiese, Wykowska, Zwickel & Müller, 2012), two different stimuli were used, namely a human and a robot face (see **Figure 1**). The likelihood of adopting an intentional stance was manipulated by using three different instructions in which participants were told that they are observing 1) a human and a robot, 2) a human and a robot, whose eyes are controlled by a human or 3) a mannequin and a robot. In conditions in which participants were made to believe to observe intentional behavior, attentional engagement was expected to be significantly larger as compared to conditions when adopting the intentional stance was less likely.

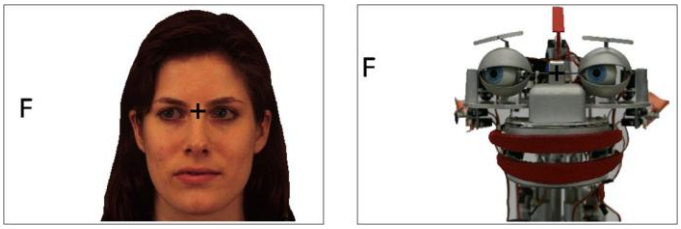

Figure 1. Human and robot used as stimuli (F represents the target).

The stimuli were presented in a classical gaze-cueing paradigm (Friesen & Kingstone, 1998), in which gaze direction is used to shift the attentional focus of the observer to the gazed-at location, resulting in faster responses to targets appearing at the gaze-cued position compared to targets at the uncued position. Attentional engagement was determined by measuring the magnitude of the cueing effects that have been induced by the human and the robot gazer depending on which instruction was given.

Downloaded from pro.sagepub.com at TU Muenchen on December 15, 2013

The results clearly supported the intentional stance hypothesis: In conditions in which participants were made to believe they were observing intentional behavior (i.e., changes in gaze direction were caused by a human), cueing effects were significantly larger compared to conditions when adopting the intentional stance was unlikely (i.e., changes in gaze direction were pre-programmed). Most importantly, the effect of adopting the intentional stance to the gazer was independent of whether a human or a robot face was presented (**Figure 2**). That is, the same stimulus elicited cueing effects to varying degrees depending on which belief has been induced by instruction: the human face yielded reduced cueing effects when it was believed to represent a mannequin (comparable to the effects of the robot face in the baseline condition: instruction 1), while the robot face elicited enhanced cueing effects when it was believed to be controlled by a human (comparable to the effects of the human face in the baseline condition: instruction 1). This pattern has been shown very consistently and generalized from more complex discrimination tasks (decide whether capital letter F or T is shown) to other tasks with different attentional demands, such as target localization tasks (decide whether a target is shown on the left or the right side of the screen).

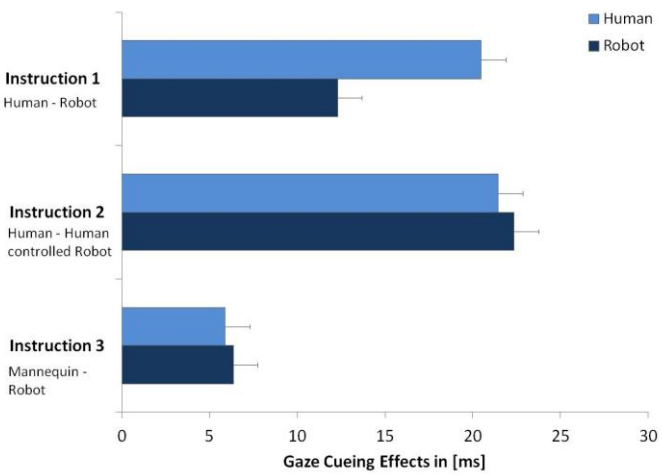

Figure 2. Size of the gaze-cueing effects as a function of cue identity (human, robot) and Instruction (1,2,3). **Instruction 1**: Human, Robot, **Instruction 2:** Human, Human-controlled Robot, and **Instruction 3:** Mannequin, Robot. Figure adapted from Wiese et al. (2012).

Study 2: EEG findings

In Study 2 (Wykowska, Wiese, Prosser & Müller, under review), the EEG/ERP-method was used to identify the neural correlates underlying modulatory effects on gaze cueing exerted by adopting the intentional stance towards the gazer.

We expected to find a modulation of the *sensory gain* mechanism, dependent on whether participants could attribute the mind to the gazer. The sensory gain mechanism is typically reflected in the P1/N1 ERP components and is said to improve perceptual coding by increasing the signal-to-noise ratio either through i) enhanced processing of stimuli at the attended locations (N1) and/or ii) suppression at other, interfering locations (P1) (Luck, Vogel & Woodman, 2000). If humans are more likely to attend where others look when they adopt the intentional stance, larger cueing effects (i.e., better performance for targets presented at gazed-at locations) should be observed for intentional compared to the mechanistic agents. This should be paralleled in modulatory effects on the sensory gain mechanism.

In one experiment, we presented the same human and robot faces as in Study 1. While participants were performing a target discrimination task, EEG was measured. We expected larger cueing effects for the human, relative to the robot condition, based on the assumption that participants would more likely attribute the mind to the human than to the robot. Hence we hypothesized that they would follow the gaze of the robot to a lesser degree than of the human. We observed a modulation of the early P1 component in the time window of 100-130 ms post-target onset. In particular, the amplitude of the EEG-signal was more enhanced for valid relative to invalid trials, but only for the human condition (**Figure 3**). This clearly indicated that the sensory gain mechanism is modulated by higher-level processes, such as adopting the intentional stance.

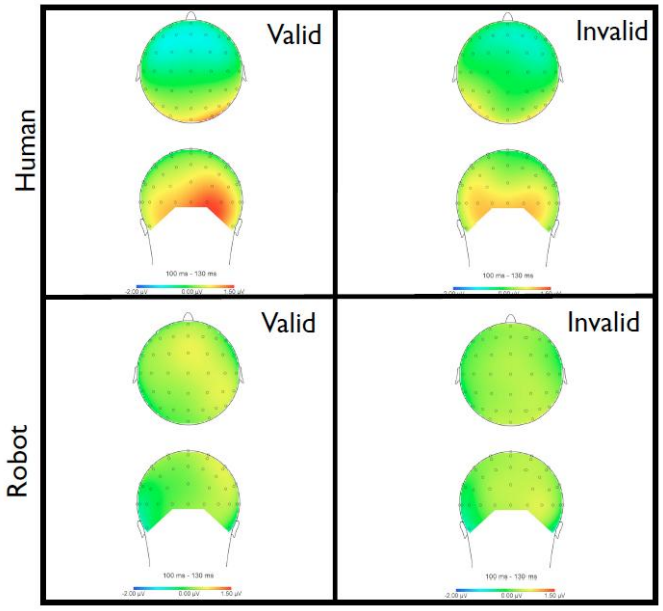

Figure 3. Voltage distribution of the ERPs in the time window of 100-130 ms, relative to target onset, view from the top (upper graphs in each panel) and from the back (lower graphs in each panel). Human condition is represented in the upper two panels while the robot condition is depicted in the lower two panels. Warm colors represent positive amplitudes, colder colors mark negative amplitudes. As can be seen from the upper left panel, the most positive amplitude was evoked by the human valid condition, relative to other three conditions.

In a second experiment, we tested whether or not the same effect would be observed if the identity of the gazer remained constant across trials (i.e., the robot), and mind attribution was manipulated through instruction (i.e., adopting the intentional stance to identical stimuli was manipulated within-partici-

Downloaded from pro.sagepub.com at TU Muenchen on December 15, 2013

pants): Instruction 1 stated that the robot's eyes were controlled by a human; while Instruction 2 informed that the gaze behavior of the robot was pre-programmed. Interestingly, the results of this experiment replicated the findings of the earlier EEG experiment: that is, larger cueing effects were observed for the human-controlled robot than for the pre-programmed robot and P1 was modulated by instruction: enhanced amplitudes were observed for valid relative to invalid trials, but only when the intentional stance was adopted towards the robot. These findings are very remarkable, given that differential effects on early sensory processing were obtained even for physically identical stimuli.

Further support for the idea that adopting the intentional stance modulates social attention mechanisms comes from a recent pilot study with ASD patients. The crucial aspect of this study is that autistic people have been shown to possess reduced empathizing skills while having increased systemizing skills (Baron-Cohen, 2002): *Empathizing* is the capability of attributing mental states to others, to predict other people's behavior, to identify other's emotional states, and to respond to these with an appropriate reaction. *Systemizing*, on the other hand, is the capability to analyze particular variables of a system, to derive rules underlying the system's behavior and to predict or control the behavior of a system. Hence, given that people with ASD have increased systemizing and decreased empathizing skills, it should be easier for them to interpret the behavior of a machine (i.e., robot face) in a meaningful way compared to the behavior of an intentional system (i.e., human face), resulting in larger cueing effects for the former than for the latter.

In order to test this hypothesis, we used the same paradigm as in the first experiment of Study 2 (human vs. robot face) with the patient sample. Fourteen patients with ASD (age range: 18-23 years) participated in the experiment and performed a discrimination task (F vs. T) while RTs were measured. Similar to the experiments described above, cueing effects were calculated as the difference in RTs between *valid* trials (face is looking at the side where the target is presented) and *invalid* trials (face is looking to the side opposite of the target).

The results of the patient sample show the inversed pattern compared to the healthy controls (Study 1): there was a significant gaze-cueing effect for the robot face (Δ_{RT}= 16 ms, $t(13)$= 2.672, p= .019*), but no significant cueing effect for the human face (Δ_{RT}= 2 ms, $t(13)$= .553, p= .590), see **Figure 4.** The results provide evidence that the degree to which attentional resources are deployed in social interactions seems to depend on whether or not observed gaze behavior can be interpreted in a way that is meaningful to the observer: while meaningful means *intentional* for healthy participants, interpreting eye movements in a *systematic* or mechanistic way seems to be more meaningful for patients with ASD.

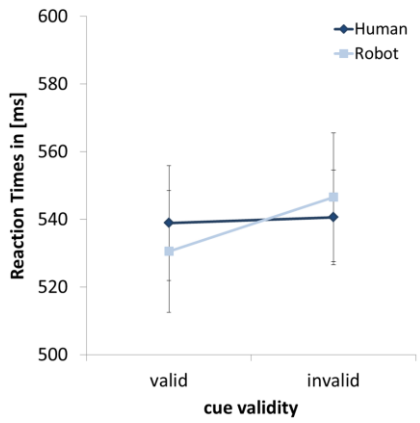

Figure 4. Reaction times for valid and invalid trials in the ASD sample, both for the human and robot face: no cueing effect was observed in the human condition, but a large cueing effect in the robot condition.

DISCUSSION

The findings reported in the series of studies reviewed here provide evidence that the allocation of attention following changes in gaze direction is influenced by the degree to which intentionality can be attributed to observed gaze behavior: that is, gazers that were believed to behave like humans induced stronger gaze-cueing than gazers that were perceived as machines. The EEG-results further show that the modulation of cueing effects by adopting the intentional stance to the observed behavior is independent of the gazer's physical appearance and originates from modulatory effects on early sensory gain processes (reflected in P1).

A modulation of attentional engagement in social interactions depending on whether or not intentionality is attributed to observed behavior is highly plausible, as it allows for attentional mechanisms to be sensitive to the social relevance of the objects/events in the environment: the bottom-up mechanism ensures a general preparedness to social signals transmitted by others, while the top-down mechanism warrants flexible allocation of attention to objects/events that are socially meaningful as they constitute the targets of another's intentions**.**

Two component-model of social attention

Based on the findings of the three experiments outlined above, we postulate a two component-model of social attention that integrates the influences of the bottom-up and top-down processes (**Figure 5**): while the bottom-up mechanism induces a default mode of orienting of attention to the gaze-cued hemifield, the top-down mechanism assigns social relevance to the observed scene and modulates the bottom-up response of the brain. Due to this weighting, attentional resources are allocated in a way that objects or events that are relevant for the social interaction are prioritized.

Downloaded from pro.sagepub.com at TU Muenchen on December 15, 2013

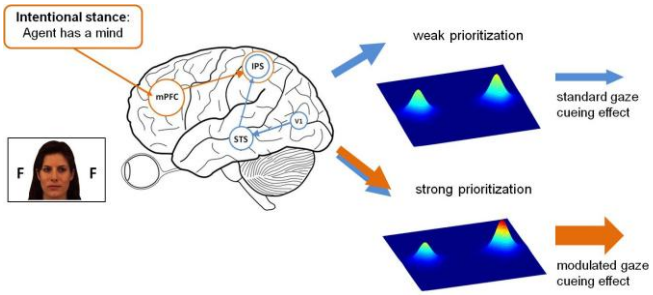

Figure 5. Two-component model of social attention. Bottom-up orienting to gaze cues is carried out by a network consisting of visual areas (V1), STS and IPS (in blue). Top-down modulation of the bottom-up signal is exerted via connections with the mPFC region (in orange) and in case of adopting the intentional stance in particular via the anterior cingulate (Gallagher et al., 2002).

The bottom-up mechanism ensures a general preparedness to the social signals transmitted by others and is executed by a brain network that is linking sensory information from *visual cortex* (dark iris is moving relative to white sclera), to the *STS* (interpretation of observed behavior as eye movement) and the *IPS* (shifting attention to gaze cued objects). In order to weight context information according to its social relevance, the top-down mechanism (executed via *mPFC*) assures that attentional resources are deployed according to the intentionality underlying the observed changes in gaze direction.

Guidelines for implementing gaze control in HRI

The results provide evidence that social attention mechanisms are influenced by the likelihood with which intentionality can be attributed to the observer: that is, robots believed to behave like humans induce stronger engagement than robots believed to behave like machines. Most importantly, the attribution of intentionality seems not to depend on the appearance of the robot, but only on participants' beliefs about the robot's internal states. Furthermore, the findings imply that gaze direction can be actively used by robots to convey information about their internal states and to direct the attentional focus of others to important events in the world.

Taken together, gaze direction seems to be a very powerful tool that can be used in human-robot interaction to communicate intentions and upcoming action goals in a natural way. To effectively use gaze control in social robotics, robots must not only be able to interpret the gaze direction of humans, but also be capable of showing meaningful (e.g., action congruent) eye movements themselves. It will be important for future research to determine behavioral manifestations of intentional systems and to equip social robots with these behaviors in order to trigger an adoption of the intentional stance to the robot.

References

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(1980)*, 347-368.

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

Baron-Cohen, S. (2002). The extreme male brain theory of autism. *TICS, 6(6)*, 248-254.

Corbetta, M., Miezin, F.M., Shulman, G.L., & Petersen, S.E. (1993). A PET study of visuospatial attention. *Journal of Neuroscience, 13*, 1202-1226.

Dennett, D.C. (2003). True believers: the intentional strategy and why it works. In O'Connor, T., & Robb, D. (eds.). *Philosophy of Mind: Contemporary Readings*, pp. 370-390. London: Routledge.

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

Frith, C.D., & Frith, U. (2006). How we predict what other people are going to do. *Brain Research, 1079*, 36-46.

Gallagher, H.L., Jack, A.I., Roepstorff, A., & Frith, C.D. (2002). Imaging the intentional stance in a competitive game. *Neuroimage, 16(3*), 814-821.

Grezes, J., Frith, C.D., & Passingham, R.E. (2004). Brain mechanisms for inferring deceit in the actions of others. *Journal of Neurosciences, 24(24)*, 5500-5505.

Kawai, N. (2011). Attentional shift by eye gaze requires joint attention: Eye gaze cues are unique to shift attention. *Japanese Psychological Research, 53*, 292–301.

Luck, S.J., Woodman, G.F., & Vogel E.K. (2000). Event-related potential studies of attention. *TICS, 4*, 432-440.

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*, 156-170.

Ristic, J., & Kingstone, A. (2005) Taking control of reflexive social attention, *Cognition 94(3):* B55-B65.

Saxe, R., & Wexler, A. (2005). Making sense of another mind: the role of the right temporo-parietal junction. *Neuropsychologia, 43(10)*, 1391-1399.

Striano, T. & Reid, V.M. (2006). Social cognition in the first year. *TICS, 10(10)*, 471-476.

Teufel, C., Fletcher, P. C., & Davis, G. (2010). Seeing other minds: attributed mental states influence perception. *TICS*, 14, 376–382.

Wiese, E., Zwickel, J. & Müller, H.J. (accepted). The importance of context information on the spatial allocation of social attention. *Attention, Perception & Psychophysics*.

Wiese, E., Wykowska, A. & Müller, H.J. (submitted). What we observe is biased by what other people tell us: Beliefs about the reliability of gaze behavior modulate attentional orienting to gaze cues. *Cognition.*

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): e45391*.

Wykowska, Wiese, Prosser & Müller (submitted). Readiness to engage in social interactions depends on beliefs about the minds of others, an EEG study. *JoCN*

Downloaded from pro.sagepub.com at TU Muenchen on December 15, 2013