

Do you see what I'm thinking about? How adults' eye-movements are influenced by representations of mental states.

Paula Rubio-Fernandez

Department of Psychology, Princeton University

Department of Literature, Area Studies and European Languages, Oslo University

prubio@princeton.edu

Stephen Butterfill

Department of Philosophy, Warwick University

s.butterfill@warwick.ac.uk

Daniel C. Richardson

Department of Psychology, University College London

dcr@eyethink.org

Following recent eye-tracking research on perspective-taking and false-belief reasoning in adults (Keysar, Lin & Barr, 2003; Ferguson, Scheepers & Sanford, 2010; Rubio-Fernandez & Glucksberg, *under review*), we propose to investigate the processes and representations that underlie eye-movements in Theory of Mind tasks. Focusing on those patterns of eye-movements that have been interpreted as evidence of an 'egocentric bias' in adult false-belief reasoning, our primary aim is to test hypotheses about which representations control eye-movements during comprehension of narratives involving false beliefs. Our collaborative research should thereby contribute to controversies concerning the nature and development of Theory of Mind cognition.

Background of the topic

Interest in eye-movements as a window on social cognition has been sparked by recent eye-tracking studies using non-verbal false-belief tasks showing that infants as young as 13 months are able to keep track of other people's beliefs (Onishi & Baillargeon, 2005; Surian, Caldi & Sperber, 2007; Southgate, Senju & Csibra, 2007). As ground-breaking as the use of eye-monitoring techniques has been in recent Theory of Mind research, earlier developmental studies using traditional false-belief tasks had already revealed children's improved performance using eye-movement measures (Garnham & Perner, 2001; Ruffman, Garnham, Import & Collony, 2001).

In a pioneering study monitoring children's eye-gaze, Clements and Perner (1994) observed within-child discrepancies on explicit versus implicit measures of false-belief understanding. For the implicit measure, they monitored where children looked in anticipation of a protagonist looking for a desired object when he was mistaken about the object's location. From 2 years 11 months to 4 years 5 months, about 90% of the children looked at the empty location where the protagonist thought the object was. This result is in sharp contrast with the actual responses that children in that age range gave when they were asked to predict where the character would look for the object: in the explicit measure, only about 45% of these children gave the empty location as their answer. As in other, non-social domains (e.g. Hood, Cole-Davies & Dias, 2003), it appears that there is a discrepancy between eye-movements and other responses concerning what subjects know.

How might the discrepancy be resolved? Although there are several proposals (e.g. Baillargeon, Scott & He, 2010; Perner & Ruffman, 2005), we focus here on just one. Apperly and Butterfill (2009) argue that the contrasting results of the above studies are consistent with the possibility that an early-developing system for tracking belief-like states guides children's eye movements, while a later-developing system guides children's explicit judgments about beliefs. According to Apperly and Butterfill, infants and perhaps some non-human animals are able to solve some Theory of Mind tasks by virtue of having one or more systems that are cognitively efficient but limited and inflexible. As children's mastery of language and executive function develops, they become able to deploy a new system for Theory of Mind reasoning that is highly flexible but cognitively demanding. In adults, both systems operate in parallel, with the cognitively efficient system playing a central role in online social interaction and communication while the cognitively flexible system enables adults to engage in top-down guidance of social interaction and explicit reasoning about mental states. According to Apperly and Butterfill (2009), efficient but limited systems for Theory of Mind cognition may also play a role in explaining how adults are able to communicate in ways that apparently depend on ascribing mental states to others.

While there are grounds for thinking that Apperly and Butterfill's conjecture about mental state ascriptions in communication will at least need refining (Rubio-Fernandez & Glucksberg, *under review*), they are surely right on the more general point that the interpretation of research on infants' and children's theory of mind cognition could be strengthened by consideration of adults' online performance, particularly their performance in communicative contexts.

In fact recent eye-tracking studies on adult Theory of Mind have also found different results using implicit and explicit false-belief tasks. Ferguson, Scheepers and Sanford (2010) asked adult participants to listen to a series of short false-belief stories while looking at various static scenes. Like the children in Clements and Perner (1994), adults were able to anticipate the behavior of a protagonist who was

mistaken about a certain state of affairs without showing interference from their own knowledge of the situation. In contrast, in another recent eye-tracking study using a standard false-belief task with adults, Rubio-Fernandez and Glucksberg (*under review*) observed that when participants were asked to predict where a character would look for a desired object that, unbeknownst to her, had been moved to different container, they would often look first towards the current location of the object before they would correct this initial tendency and respond that she would look for the object where she had left it (i.e., in the empty container).

The above studies suggest that both children and adults suffer more interference from their own knowledge of a situation in 'explicit' false-belief tasks where they are asked to make an overt prediction of a mistaken character's behavior, than in those 'implicit' tasks where they only have to watch a false-belief story while their eye-gaze is monitored as predictive of their expectations about the character's behavior. However, the fact that adults have shown a tendency to first fixate on the 'true-belief response' (i.e., the current location of the object rather than the empty container representing the character's false belief) in an explicit false-belief task might cast doubt on the hypothesis that participants' eye-movements are guided by a fast and efficient system to track other people's beliefs while their responses to the false-belief question are guided by a flexible yet cognitively demanding system (cf. Apperly & Butterfill, 2009). Thus, the first question that we would like to investigate in our collaborative research project is whether the first fixations on the true-belief response that have been observed in explicit false-belief tasks with adults might be driven by the same flexible yet cognitively demanding system to track beliefs that would guide the actual responses to the false-belief question, or whether eye-movements in false-belief tasks are always guided by an efficient but limited Theory of Mind system that would play a crucial role in online social interaction.

A second, related question that we propose to investigate is whether adult participants' first fixations on the true-belief location are *predictive* of the character's behaviour (i.e., whether participants are applying some sort of 'egocentric heuristic' to the effect that their true belief about the location of the toy works as an initial default) or whether these first eye-fixations might reflect a lower-level process having to do with memory retrieval. Recent eye-tracking studies by Richardson and collaborators have shown that people often look to uninformative locations in a visual display when trying to recall spoken information that has previously been associated with those locations (Richardson & Spivey, 2000; Hoover & Richardson, 2008). Research on memory and scene perception therefore offers an alternative interpretation of participants' first fixations on the true-belief container in explicit false-belief tasks: rather than predicting the character's behaviour on the basis of their own true belief about the location of the object, participants' first fixations might respond to spatial indexing behaviour. That is, as a result of an object-based attentional mechanism that uses the world as an 'external memory store', when participants in a false-belief task are asked about the desired object, they might initially look towards its current location as a way of retrieving associated cues that will help them recall relevant information about the object.