Journal: NeuroImage

Format: 'Comments and Controversies' article

Word count: 2150

Date submitted: 19th June 2015; reviews received 9th December 2015; revised version submitted 16th December 2015; accepted 9th February 2016

Avatars and Arrows in the Brain

Caroline Catmur[a,b]*, Idalmis Santiesteban[c], Jane R. Conway[d], Cecilia Heyes[e], & Geoffrey Bird[d,f]

[a] School of Psychology, University of Surrey, Guildford, Surrey, GU2 7XH, UK.

[b] Department of Psychology, Institute of Psychiatry, Psychology & Neuroscience, King's College London, SE1 1UL, UK. caroline.catmur@kcl.ac.uk.

[c] Department of Psychology, University of Cambridge, Downing Street, Cambridge, CB2 3EB, UK. is405@cam.ac.uk.

[d] MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, SE5 8AF, UK. jane_rebecca.conway@kcl.ac.uk; geoff.bird@kcl.ac.uk.

[e] All Souls College, University of Oxford, Oxford, OX1 4AL, UK. cecilia.heyes@all-souls.ox.ac.uk.

[f] Institute of Cognitive Neuroscience, University College London, London, WC1N 3AR, UK.

* Corresponding author: Caroline Catmur, caroline.catmur@kcl.ac.uk

Abstract

In this *Commentary* article we critically assess the claims made by Schurz, Kronbichler, Weissengrubler, Surtees, Samson and Perner (2015) relating to the neural processes underlying theory of mind and visual perspective taking. They attempt to integrate research findings in these two areas of social neuroscience using a perspective taking task contrasting mentalistic agents ('avatars'), with non-mentalistic control stimuli ('arrows'), during functional Magnetic Resonance Imaging. We support this endeavour whole-heartedly, agreeing that the integration of findings in these areas has been neglected in research on the social brain. However, we cannot find among the behavioural or neuroimaging data presented by Schurz et al. evidence supporting their claim of 'implicit mentalizing' - the automatic ascription of mental states to another representing what they can see. Indeed, we suggest that neuroimaging methods may be ill-suited to address the existence of implicit mentalizing, and suggest that approaches utilizing neurostimulation methods are likely to be more successful.

Keywords: Visual perspective taking; implicit mentalizing; theory of mind; domain-general; attentional orienting

Avatars and Arrows in the Brain

In a recent paper, Schurz et al. (2015) seek to integrate research on the neural mechanisms underlying theory of mind and visual perspective taking. They do this using a perspective taking task devised by Samson et al. (2010), involving 'avatars', and a control condition that we added recently to this task (Santiesteban et al., 2014), involving arrows. We applaud their clear-sighted identification of a "conceptual gap" (Schurz et al., p. 386) in research on the social brain, and attempt to fill it using an avatar-arrow comparison in Samson et al.'s perspective taking task. However, as we shall explain in this commentary, we cannot find among the behavioural or neuroimaging data presented by Schurz et al. evidence of implicit mentalizing, rather than domain-general processing.

In the dot perspective task, participants see an image of a room with discs or dots on the walls. In the centre of the room is a human figure, an avatar, who, because s/he is facing to the left or the right, can see either all, or only some, of the dots. Participants are asked to judge either how many dots they themselves can see: an *explicit self-perspective* judgement; or how many dots the avatar can see: an *explicit other-perspective* judgement. Importantly, because the avatar cannot always see the same number of dots as the participant themselves, this task allows researchers to address the question of whether the avatar's visual perspective is spontaneously processed even when judging self-perspective. On trials when the avatar can see all of the dots, the visual perspectives of the avatar and of the participant are *consistent*; whereas on trials when the avatar can see only some of the dots, the two visual perspectives are *inconsistent*. A number of behavioural experiments have confirmed that response times to judge self-perspective are longer when the two visual perspectives are inconsistent, than when they are consistent (Samson et al., 2010; Qureshi et al., 2010; McCleery et al., 2011; Santiesteban et al., 2014). This result has been taken to support the claim that the avatar's visual perspective is spontaneously processed, under the assumption

that the inconsistency between the two visual perspectives leads to an interference effect which underlies the increase in response time. Thus the contrast between inconsistent and consistent trials when judging self-perspective, the *self-consistency effect*, is considered a measure of *implicit* visual perspective taking.

However, Santiesteban et al. (2014) showed that a similar pattern of increased response times on inconsistent, versus consistent, trials can occur when the avatar is replaced with an arrow which points to some or all of the dots. This finding casts doubt on the claim that the self-consistency effect reflects spontaneous processing of the avatar's visual perspective, i.e. that it is due to representation by the participant of the avatar's mental state, specifically, of what the avatar can see. It raises the possibility that, both when the central stimulus is an avatar and when it is an arrow, the self-consistency effect is due to a domain-general process, such as automatic attentional orienting. On this domain-general account, the directional, rather than agentive, features of the avatar speed responding in consistent trials and/or slow responding in inconsistent trials simply by directing attention to the dots in front of the avatar.

Following Santiesteban et al. (2014), Schurz et al. (2015) also included in their experiment trials in which the avatar was replaced with an arrow, allowing measurement of the effect of *animacy* on behavioural and neural responses. They proposed that the inclusion of these trials would allow exploration of whether the (implicit) self-consistency effect found for the avatar is the result of different neural mechanisms than those underlying the self-consistency effect for the arrow. (Schurz et al. also included two other non-mental-state stimuli - a lamp and a brick wall. These control conditions are problematic, primarily due to their spatial features, but since the data for these conditions are not fully reported they will not be considered further here.)

Response Time (RT) Data

When reporting their behavioural data, Schurz et al. (2015) stated that their arrow stimuli had "failed to produce the same reaction time effect as that of Santiesteban et al. (2014)" (p. 393). This is potentially misleading for two reasons. First, like Santiesteban et al., Schurz et al. found a main effect of consistency (longer RTs in inconsistent than in consistent trials), and no interaction between consistency and animacy. Thus, their self-consistency effect was not significantly greater when the central stimulus was an avatar rather than an arrow. Second, even if Schurz et al. had found a smaller consistency effect in their arrow condition than in their avatar condition, or in the arrow condition used by Santiesteban et al., this would not have been a failure of replication, or had any bearing on whether consistency effects are due to implicit visual perspective taking or to domain-general processing. This is because, in contrast with Santiesteban et al., Schurz et al. opted to use arrow stimuli that were not matched in terms of their low-level visual properties with their avatar stimuli. For example, their avatar stimuli were vertically aligned while their arrow stimuli were horizontally aligned. Therefore, it is possible that the arrows differed in salience from the avatar stimuli, or that the directional properties of the arrow stimuli were more or less discriminable than those of the avatar stimuli. Either of these differences could have led to a smaller consistency effect in the arrow than the avatar condition – a fact acknowledged by Schurz et al., "we could have produced a quantitatively stronger cueing effect with an arrow similar to that of Santiesteban et al. (2014)" (p. 393) – and in neither case would it have amounted to a failure to replicate Santiesteban et al., or to evidence that the two effects are mediated by different psychological processes. The purpose of the arrow condition is to show that other directional stimuli, besides avatars, can produce a self-consistency effect, and thus to highlight that it cannot be assumed that the self-consistency effect observed with the avatar is due to ascription of mental states to the avatar. Schurz et al. acknowledge this

possibility, but suggest that their neuroimaging data support the interpretation that the self-consistency effects in the avatar and arrow conditions are due to different processes: mental state ascription in the avatar condition and attentional orienting in the arrow condition. We disagree with this interpretation, for the reasons given below.

Neuroimaging Data

Schurz et al. (2015) reported both a whole-brain analysis of their data and region of interest analyses centred on putative theory of mind areas (right posterior temporoparietal junction (rTPJp), ventral medial prefrontal cortex (vmPFC), and ventral precuneus). In our view, claims are made about the response in these areas that are not supported by the data presented in the paper.

Here we highlight three specific claims for which we could not identify supporting data. The first claim is that there was a main effect of consistency in theory of mind areas: "activation was higher when self- and other-perspectives were inconsistent" (p. 394). The second suggests that there was a consistency x animacy interaction: "stronger activation for inconsistent > consistent perspectives for avatar" (p. 391) and "theory of mind areas engage when the scene shows a perspective difference" (i.e., a difference in content between two perspectives) (p. 395). The third suggests a three-way interaction between perspective, consistency, and animacy, or at least a simple interaction effect between consistency and animacy on self-judgement trials: "these areas are spontaneously processing information linked to the other's perspective during self-perspective judgements" (abstract, p. 386); "activation was also sensitive to the consistency between perspectives — again only for the avatar and not for the arrow" (p. 391); and "vmPFC and ventral precuneus were engaged in spontaneous other-perspective taking during self-judgements" (p. 394).

We have not been able to find anywhere in the published report (Schurz et al., 2015) data to support these conclusions, i.e. a significant main effect of consistency, a consistency x animacy interaction or a perspective x consistency x animacy interaction. The authors explicitly stated that non-significant results were not reported. Therefore, we can only infer that these effects were not present either in the ROI data or the whole-brain analysis.

Although the absence of these effects is problematic for how Schurz et al. (2015) interpret their data, in our view there is a more fundamental problem with the logic of the experiment. Schurz et al. take their results to support the suggestion that, even when judging one's own perspective, the presence of another agent (the avatar) prompts an automatic process that causes the participant to represent the avatar's mental state, specifically what the avatar can see. In the behavioural data this effect can be seen when comparing consistent and inconsistent trials, where it is reflected in a significant difference in RT between these trial types, but this is not the case for the imaging data. If the ascription of mental states elicits reliable patterns of activation then this activation should be seen on both consistent and inconsistent trials. Comparison of the imaging data for consistent and inconsistent trials (or any interaction involving consistency) will therefore reveal activation related to the resolution of response conflict associated with inconsistent trials, but will not reveal activation related to the ascription of mental states.

Under the implicit mentalizing position, then, both the consistent and inconsistent avatar conditions should prompt implicit mentalizing, while the arrow conditions should not. Implicit mentalizing would therefore be revealed by a main effect of animacy (avatar vs arrow). But there is a major problem: the contrast of avatar vs arrow confounds the eliciting stimulus (the avatar) and the process putatively elicited by that stimulus (mentalizing). Any activation revealed by the avatar vs arrow contrast could, therefore, reflect either mentalizing or a basic perceptual response to faces or bodies. With respect to this point it is interesting

that the peak activation for the main effect of animacy in the region of rTPJp is within smoothing distance of an area known preferentially to process images of the human body (the extrastriate body area; Downing, Jiang, Shuman, & Kanwisher, 2001); and previous fMRI studies have found that the mere presence of a human-like figure activates areas including mPFC (Dumontheil et al., 2010) and TPJ (Abraham et al., 2008)[1]. This problem is usually solved within fMRI designs by holding the stimulus constant and requiring different tasks to be performed in response to the same stimulus. For example, if one is interested in facial emotional recognition, one contrasts activation during a task in which emotional faces are presented and the participant is asked to recognise emotion, with activation during a task in which emotional faces are presented and the participant is asked to judge gender. As the stimulus (emotional faces) stays constant in this comparison, any differential activation can be assumed to be due to the process of recognising facial emotion. If it is the case that spontaneous implicit mentalizing occurs whenever one sees an agent, it becomes impossible to separate non-mentalistic processes related to processing of the stimulus from the mentalizing which hypothetically accompanies perception of the stimulus. Thus, one cannot interpret the main effects of animacy reported by Schurz et al. (2015), and any interactions with (or simple effects of) consistency may therefore reflect purely response conflict-related processes (or such response conflict-related processes along with their downstream effects on relevant representations), rather than implicit mentalizing.

This point can be generalized to other fMRI studies on implicit theory of mind. In Kovacs et al. (2014) and Schneider et al. (2014) the relevant contrast is between conditions where an agent holds false versus true beliefs. We suggest a similar problem occurs in these

[1] We thank a reviewer for pointing out that the activation in the contrast avatar>arrow (Figure 4, top) overlaps with that in the contrast lamp>arrow (Figure 5, bottom), casting further doubt on the claim that the former contrast delineates brain areas that are specific to the process of mentalizing.

studies: the agent's beliefs should be represented in both conditions, and therefore any response in the contrast false belief>true belief cannot reflect the representation of beliefs *per se*, (i.e. "belief tracking"; Kovacs et al., 2014) but instead reflects conflict-related processes.

While fMRI may be ill-suited to investigate predictions related to implicit mentalizing, future studies could use causal techniques to test whether theory of mind areas such as rTPJ are indeed necessary for spontaneous representation of another's visual perspective during self-perspective trials. If, as Schurz et al. (2015) suggest, rTPJ is involved in producing the consistency effect, then stimulation of rTPJ should influence the effect of consistency on RTs; if, in contrast, rTPJ is involved in explicit visual perspective taking then rTPJ stimulation should influence the effect of explicit perspective taking on RTs. Brain stimulation studies can therefore build on these neuroimaging results by investigating whether the brain areas putatively identified in the present study are involved in the underlying psychological processes.

References

Abraham, A., Werning, M., Rakoczy, H., von Cramon, D.Y., Schubotz, R.I., 2008. Minds, persons, and space: An fMRI investigation into the relational complexity of higher-order intentionality. Conscious. Cogn. 17, 438–450. doi:10.1016/j.concog.2008.03.011

Downing, P.E., Jiang, Y., Shuman, M., Kanwisher, N., 2001. A cortical area selective for visual processing of the human body. Science 293, 2470–2473. doi:10.1126/science.1063414

Dumontheil, I., Küster, O., Apperly, I.A., Blakemore, S.J., 2010. Taking perspective into account in a communicative task. Neuroimage 52, 1574–1583. doi:10.1016/j.neuroimage.2010.05.056

Kovács, A.M., Kühn, S., Gergely, G., Csibra, G., Brass, M., 2014. Are all beliefs equal? Implicit belief attributions recruiting core brain regions of theory of mind. PLoS One. 9(9), e106558. doi: 10.1371/journal.pone.0106558

McCleery, J.P., Surtees, A.D.R., Graham, K. A, Richards, J.E., Apperly, I. A., 2011. The neural and cognitive time course of theory of mind. J. Neurosci. 31, 12849–12854. doi:10.1523/JNEUROSCI.1392-11.2011

Qureshi, A.W., Apperly, I. A., Samson, D., 2010. Executive function is necessary for perspective selection, not Level-1 visual perspective calculation: Evidence from a dual-task study of adults. Cognition 117, 230–236. doi:10.1016/j.cognition.2010.08.003

Samson, D., Apperly, I. A., Braithwaite, J.J., Andrews, B.J., Bodley Scott, S.E., 2010. Seeing it their way: evidence for rapid and involuntary computation of what other people see. J. Exp. Psychol. Hum. Percept. Perform. 36, 1255–1266. doi:10.1037/a0018729

Santiesteban, I., Catmur, C., Coughlan Hopkins, S., Bird, G., Heyes, C., 2014. Avatars and arrows: implicit mentalizing or domain-general processing? J. Exp. Psychol. Hum. Percept. Perform. 40, 929–37. doi:10.1037/a0035175

Schneider, D., Slaughter, V.P., Becker, S.I., Dux, P.E., 2014. Implicit false-belief processing in the human brain. Neuroimage 101, 268-75. doi: 10.1016/j.neuroimage.2014.07.014

Schurz, M., Kronbichler, M., Weissengruber, S., Surtees, A., Samson, D., Perner, J., 2015. Clarifying the role of theory of mind areas during visual perspective taking: Issues of spontaneity and domain-specificity. Neuroimage. 117, 386-96. doi:10.1016/j.neuroimage.2015.04.031