

Accepted Manuscript

Current knowledge on the role of the Inferior Frontal Gyrus in Theory of Mind – a
Commentary on Schurz and Tholen (2016)

Dr Charlotte E. Hartwright, Peter C. Hansen, Ian A. Apperly

PII: S0010-9452(16)30279-9

DOI: [10.1016/j.cortex.2016.10.005](https://doi.org/10.1016/j.cortex.2016.10.005)

Reference: CORTEX 1855

To appear in: *Cortex*

Received Date: 9 May 2016

Revised Date: 4 September 2016

Accepted Date: 14 October 2016

Please cite this article as: Hartwright CE, Hansen PC, Apperly IA, Current knowledge on the role of the Inferior Frontal Gyrus in Theory of Mind – a Commentary on Schurz and Tholen (2016), *CORTEX* (2016), doi: 10.1016/j.cortex.2016.10.005.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**Current knowledge on the role of the Inferior Frontal Gyrus in Theory of Mind – a
Commentary on Schurz and Tholen (2016)**

Charlotte E Hartwright^{a*}, Peter C Hansen^b, Ian A Apperly^b

^a Aston Brain Centre, School of Life and Health Sciences, Aston University,
Birmingham, B4 7ET, UK

^b Centre for Behavioural Brain Sciences, School of Psychology, University of
Birmingham, Birmingham, UK

*Corresponding author: Dr Charlotte E Hartwright

Email: c.hartwright@aston.ac.uk

Tel: 0121 204 3000

Current knowledge on the role of the Inferior Frontal Gyrus in Theory of Mind – a Commentary on Schurz and Tholen (2016)

Schurz and Tholen (2016) argue that common approaches to studying the neural basis of “theory of mind” (ToM) obscure a potentially important role for inferior frontal gyrus (IFG) in managing conflict between perspectives, and urge new work to address this question: “to gain a full understanding of the IFG's role in ToM, we encourage future imaging studies to use a wider range of control conditions.” (p332). We wholeheartedly agree, but note that this observation has been made before, and has already led to a programme of work that provides evidence from fMRI, EEG, and TMS on the role of IFG in managing conflict between self and other perspectives in ToM. We highlight these works, and in particular we demonstrate how careful manipulation within ToM tasks has been used to act as an internal control condition, wherein conflict has been manipulated within-subject. We further add to the discussion by framing key questions that remain regarding IFG in the context of these. Using limitations in the existing research, we outline how best researchers can proceed with the challenge set by Schurz and Tholen (2016).

Theory of Mind and cognitive control

Children’s and adults’ success at reasoning about the beliefs, desires and intentions of others – an ability termed “theory of mind” or “mindreading” – is associated with performance on tests of cognitive control (Carlson & Moses, 2001; Carlson, Moses, & Breton, 2002; Devine & Hughes, 2014; German & Hehman, 2006). It is commonly proposed that one reason for this is that cognitive control is necessary to overcome interference from one’s own “egocentric” or “self” perspective, when judging the perspectives of others. In light of such behavioural results, it is striking that neuroimaging work on ToM often describes the core “social brain” network as comprising the temporoparietal junction and medial prefrontal cortex, but not brain regions associated with cognitive control, such as the inferior frontal cortex. Indeed, the review presented in Schurz and Tholen (2016), and two recent quantitative reviews, are inconsistent regarding the presence of activations in IFG in ToM tasks

(Molenberghs, Johnson, Henry, & Mattingley, 2016; Schurz, Radua, Aichhorn, Richlan, & Perner, 2014). Schurz and Tholen (2016) rightly suggest that this may be because frequently-used paradigms aim to identify the neural correlates of ToM (in particular reasoning about false beliefs) by subtracting out baseline conditions that require closely-matched reasoning about non-social situations (perhaps most notably, reasoning about false photographs (for a typical example, see Saxe & Kanwisher, 2003)). While this approach is vital for identifying brain regions that might be domain-specific to ToM, it risks subtracting out functionally critical processes that are shared with other tasks (e.g., Apperly, 2011, 2013), including cognitive control processes. While questions about domain-specificity for ToM have had most prominence in the literature in the past 15 years, a growing number of more recent studies suggest that this approach misses critical information about the functional and neural basis of ToM. Consequently, though quantitative brain-based meta-analyses are informative, they are necessarily constrained by the prevalent methods in the prior literature, and the brain regions that are modulated as a result of those paradigmatic choices. Thus, where current meta analyses appear to indicate that IFG is only recruited for certain ToM tasks, this cannot be determined with any certainty due to the relative lack of published work that seeks to specifically manipulate cognitive control in ToM.

One of the approaches used to examine cognitive control in ToM has manipulated the salience of the participant's own perspective, while they were engaged in reasoning about the false beliefs or differing visual perspectives of others (e.g., Hartwright, Apperly, & Hansen, 2015; van der Meer, Groenewold, Nolen, Pijnenborg, & Aleman, 2011; Vogeley et al., 2001). For example, Vogeley et al. (2001) found greater activation in right IFG when participants reasoned about others' mental states in a story scenario in which they themselves also featured, compared with a story in which they did not. Other relevant studies have found varying activation in IFG when they manipulated whether the perspective of the social target is congruent with the participant's (e.g., because they have a true belief, or see the same number of objects) or incongruent with the participant's (e.g., because they have a false belief, or see a different number of objects; e.g., Abraham, Rakoczy, Werning, von

Cramon, & Schubotz, 2010; McCleery, Surtees, Graham, Richards, & Apperly, 2011; Ramsey, Hansen, Apperly, & Samson, 2013; Rothmayr et al., 2011).

While IFG is associated with a variety of functions (see Liakakis, Nickel, & Seitz, 2011 for a meta-analysis) support for interpreting the above results in terms of inhibitory control comes from studies that included an independent inhibitory control task, and found that activity in IFG during the inhibitory control task overlapped with activity during the relevant ToM contrast (e.g., Rothmayr et al., 2011; Saxe, Schulz, & Jiang, 2006; van der Meer et al., 2011). These findings converge with evidence from a patient with right frontal brain injury (including damage to IFG), who showed significant impairments to inhibitory control, and showed strong egocentric biases on a range of ToM tasks (Samson, Apperly, Kathirgamanathan, & Humphreys, 2005; Samson, Houthuys, & Humphreys, 2015).

Distinguishing self-perspective inhibition from generic cognitive control

While the abovementioned studies all provide evidence that IFG may be involved in resisting interference from participants' "self" perspective or knowledge of reality during ToM, they are also to varying degrees consistent with a less interesting possibility that cognitive control processes are simply recruited in task conditions that are more challenging for participants to perform. Several recent studies, however, provide stronger evidence that self-perspective inhibition is not merely a confound in existing tasks, but that it is an integral component of thinking about alternative perspectives.

In Hartwright, Apperly, and Hansen (2012), participants undertook a ToM task that orthogonally manipulated whether the social target's belief about an object's location was true or false, and whether their desire for the object was positive or negative (they liked or disliked it). Critically, both the belief and desire factors affected task difficulty, but only the belief factor resulted in systematic differences between the participants' and target's perspectives. Thus, while both factors might recruit cognitive control processes to meet generic task demands, only the belief factor should recruit control processes associated with self-perspective inhibition.

Consistent with this distinction, both belief and desire factors modulated neural activity in anterior cingulate cortex, while only the belief factor modulated activity in vIPFC, specifically bilateral IFG. This result suggests that IFG is specifically responsive to the need for self-perspective inhibition, and not generic task demands.

In later work, we sought to evaluate self-perspective inhibition in mental- versus non mental-representation (Hartwright et al., 2015). By adapting the ToM localizer vignettes from Saxe and Kanwisher (2003), we developed a series of high and low salience of self-perspective scenarios, for both mental and non-mental representation. This factorial design enabled us to identify the neural correlates associated with representational content (mental-representation, e.g., false belief, versus non mental-representation, e.g., false photograph), self-perspective inhibition (high versus low salience of self-perspective), and the interaction between these. These data indicated an interaction in left IFG, where this region was modulated by salience in belief reasoning only (with greater activation for high salience of self-perspective) suggesting additional processing specific to ToM in the case where own perspective was highly salient and, thus, required inhibition.

Two studies provide critical evidence on the causal role of IFG in self-perspective inhibition. First, the recent neuropsychological study by Samson et al. (2015) (referred to by Schurz and Tholen (2016)) demonstrated a double-dissociation between impaired self-perspective inhibition and impaired reasoning about participants' own negative versus positive desires, with right IFG being one of the unique areas of overlap in the lesions of the two patients with impaired self-perspective inhibition. Second, based on our paradigm in Hartwright et al. (2012), we used neuro-navigated transcranial magnetic stimulation (TMS) to demonstrate that stimulation of the right vIPFC affected participants' predictions according to others' false beliefs (where self-perspective inhibition is required), whilst no such influence was seen for desires (where there was no perspective conflict). Notably, the morphology of right vIPFC and TPJ were relevant to the direction of the influence of TMS (see Hartwright, Hardwick, Apperly, & Hansen, 2016).

Further questions regarding the Inferior Frontal Gyrus in Theory of Mind

The evidence reviewed supports the view that IFG is relevant for ToM in managing conflict between perspectives. Of course, in the wider literature right IFG associated with the suppression of irrelevant information – acting as a ‘brake’ – whereas the left IFG is suggested to support the controlled retrieval of information (Aron, Robbins, & Poldrack, 2004, 2014; Badre, Poldrack, Pare-Blagoev, Insler, & Wagner, 2005; Badre & Wagner, 2007). It is likely that both the left and right IFG are relevant for ToM, but that they serve a different role in managing perspective difference. It remains unclear whether they show any domain-specific specialisation for these purposes.

Further investigation is warranted to test whether IFG is involved in inhibiting self-perspective, or instead managing conflict between perspectives, and that the same process is recruited when it is the other person’s perspective that must be inhibited (McCleery et al., 2011; Ramsey et al., 2013). Still, it has not been established whether the effects discussed here are particular to conflicting self- and other-perspectives – the participant’s own knowledge state and the target agent’s [false] belief (thus requiring ‘self-perspective inhibition’) – or whether conflicting other-other or self-self perspectives would be sufficient. For example, it is possible for a participant to know that two people have opposing beliefs – e.g., about the contents of a box – but not to know whether either belief is true. Likewise, some circumstances require the participant to hold in mind both their own current perspective and their own conflicting past, future or counterfactual perspective – e.g., “Had I known it would rain I would have brought an umbrella”. Will IFG be equally important in resolving such conflicting perspectives?

More broadly, what also remains unclear is how IFG – which is a large anatomical region – might be divided functionally in supporting ToM. Schurz and Tholen (2016) allude to a possible division where more dorsal parts of IFG support mirroring, due to activations in this region being identified in affective ToM tasks, whereas more ventral IFG might support cognitive ToM, in terms of self-perspective inhibition. Nonetheless, more fine-grained work needs to be conducted in order to fully

disambiguate the circumstances under which IFG is engaged, and more precisely how its functions are organised in ToM.

There are two notable challenges to determining whether IFG is inhibiting self-perspective, or managing conflict between perspectives. First, little is understood about the time-course of ToM. McCleery et al. (2011) provide EEG evidence that inconsistency between visual perspectives is managed by a late component in the right lateral prefrontal cortex, and that this process occurs after perspectives have been calculated. It is not yet clear, however, whether visual perspective taking and mental-state representation follow identical neural computations, and limited agreement in paradigmatic focus across the two areas of research makes this difficult to infer on the basis of existing studies (Schurz et al., 2015). A better understanding of the temporal course of ToM would facilitate more powerful neuroimaging paradigms, wherein timing information can be accommodated within modelling of brain responses. A second difficulty is that parts of the IFG are known to support language. Many of the prior ToM tasks involve written language or would likely engage subvocalizing. Criticisms have been raised suggesting that some prior ToM experimental manipulations do not balance language differences between conditions (Callejas, Shulman, & Corbetta, 2011; see also Hartwright et al., 2015 for further discussion). It is therefore important to avoid systematic biases that would also modulate language areas. Still, this is not straightforward. Language relating to unobservable mental phenomena is likely to reflect more complex syntax, where the interpretation of semantic information may be more varied. On this basis, the widely used false-belief and false-photograph comparison has been suggested to be suboptimal. Only the representational content of one of these scenarios – the misinformed belief – is technically false: a photograph, even if the scene changes immediately after its acquisition, is an accurate representation of the moment in time that it depicts (Perner & Leekam, 2008). This is a matter of individual construal, but one that may introduce bias nonetheless.

In sum, there are a number of imaging studies that indicate that IFG is a reliable correlate of a key component of ToM – managing conflict between perspectives.

Lesion studies (Samson et al., 2005; Samson et al., 2015) and our recent TMS study (Hartwright et al., 2016) provide pivotal evidence for a causal role when mentalizing in the presence of competing perspectives. More work is required to elucidate the specificity and timing of IFG's function, as well as detailed functional architecture of IFG, in representational tasks.

Acknowledgements

This research was supported by the Economic and Social Research Council (grant number ES/G01258X/1).

References

- Abraham, A., Rakoczy, H., Werning, M., von Cramon, D. Y., & Schubotz, R. I. (2010). Matching mind to world and vice versa: Functional dissociations between belief and desire mental state processing. *Social Neuroscience*, 5(1), 1-18. doi: 10.1080/17470910903166853
- Apperly, I. (2011). *Mindreaders: The Cognitive Basis of Theory of Mind*. Hove: Psychology Press.
- Apperly, I. (Ed.). (2013). *Can theory of mind grow up? Mindreading in adults, and its implications for the neuroscience and development of mindreading* (3rd ed.): Oxford Scholarship Online.
- Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends in Cognitive Sciences*, 8(4), 170-177. doi: 10.1016/j.tics.2004.02.010
- Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2014). Inhibition and the right inferior frontal cortex: one decade on. *Trends in Cognitive Sciences*, 18(4), 177-185. doi: <http://dx.doi.org/10.1016/j.tics.2013.12.003>
- Badre, D., Poldrack, R. A., Pare-Blagoev, E. J., Insler, R. Z., & Wagner, A. D. (2005). Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron*, 47(6), 907-918. doi: 10.1016/j.neuron.2005.07.023
- Badre, D., & Wagner, A. D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. [Review]. *Neuropsychologia*, 45(13), 2883-2901. doi: 10.1016/j.neuropsychologia.2007.06.015
- Callejas, A., Shulman, G. L., & Corbetta, M. (2011). False Belief vs. False Photographs: A Test of Theory of Mind or Working Memory? *Frontiers in Psychology*, 2. doi: 10.3389/fpsyg.2011.00316
- Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development*, 72(4), 1032-1053.
- Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development*, 11(2), 73-92. doi: 10.1002/icd.298
- Devine, R. T., & Hughes, C. (2014). Relations Between False Belief Understanding and Executive Function in Early Childhood: A Meta-Analysis. [Review]. *Child Development*, 85(5), 1777-1794. doi: 10.1111/cdev.12237
- German, T., & Hehman, J. (2006). Representational and executive selection resources in 'theory of mind': Evidence from compromised belief-desire reasoning in old age. *Cognition*, 101(1), 129-152. doi: 10.1016/j.cognition.2005.05.007
- Hartwright, C. E., Apperly, I. A., & Hansen, P. C. (2012). Multiple roles for executive control in belief-desire reasoning: Distinct neural networks are recruited for self perspective inhibition and complexity of reasoning. *Neuroimage*, 61(4), 921-930. doi: 10.1016/j.neuroimage.2012.03.012
- Hartwright, C. E., Apperly, I. A., & Hansen, P. C. (2015). The special case of self-perspective inhibition in mental, but not non-mental, representation. *Neuropsychologia*, 67, 183-192. doi: 10.1016/j.neuropsychologia.2014.12.015

- Hartwright, C. E., Hardwick, R. M., Apperly, I. A., & Hansen, P. C. (2016). Resting state morphology predicts the effect of theta burst stimulation in false belief reasoning. *Hum Brain Mapp*. doi: 10.1002/hbm.23255
- Liakakis, G., Nickel, J., & Seitz, R. J. (2011). Diversity of the inferior frontal gyrus-A meta-analysis of neuroimaging studies. *Behavioural Brain Research*, 225(1), 341-347. doi: 10.1016/j.bbr.2011.06.022
- McCleery, J. P., Surtees, A. D., Graham, K. A., Richards, J. E., & Apperly, I. A. (2011). The neural and cognitive time course of theory of mind. *J Neurosci*, 31(36), 12849-12854. doi: 10.1523/jneurosci.1392-11.2011
- Molenberghs, P., Johnson, H., Henry, J. D., & Mattingley, J. B. (2016). Understanding the minds of others: A neuroimaging meta-analysis. *Neuroscience & Biobehavioral Reviews*, 65, 276-291. doi: <http://dx.doi.org/10.1016/j.neubiorev.2016.03.020>
- Perner, J., & Leekam, S. (2008). The curious incident of the photo that was accused of being false: Issues of domain specificity in development, autism, and brain imaging. *The Quarterly Journal of Experimental Psychology*, 61(1), 76-89. doi: 10.1080/17470210701508756
- Ramsey, R., Hansen, P., Apperly, I., & Samson, D. (2013). Seeing it my way or your way: frontoparietal brain areas sustain viewpoint-independent perspective selection processes. *Journal of Cognitive Neuroscience*, 25(5), 670-684. doi: 10.1162/jocn_a_00345
- Rothmayr, C., Sodian, B., Hajak, G., Döhnel, K., Meinhardt, J., & Sommer, M. (2011). Common and distinct neural networks for false-belief reasoning and inhibitory control. *NeuroImage*, 56(3), 1705-1713. doi: 10.1016/j.neuroimage.2010.12.052
- Samson, D., Apperly, I. A., Kathirgamanathan, U., & Humphreys, G. W. (2005). Seeing it my way: a case of a selective deficit in inhibiting self-perspective. *Brain*, 128(Pt 5), 1102-1111. doi: 10.1093/brain/awh464
- Samson, D., Houthuys, S., & Humphreys, G. W. (2015). Self-perspective inhibition deficits cannot be explained by general executive control difficulties. *Cortex*, 70, 189-201. doi: 10.1016/j.cortex.2014.12.021
- Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind". *NeuroImage*, 19(4), 1835-1842. doi: 10.1016/s1053-8119(03)00230-1
- Saxe, R., Schulz, L. E., & Jiang, Y. V. (2006). Reading minds versus following rules: Dissociating theory of mind and executive control in the brain. *Social Neuroscience*, 1(3-4), 284-298. doi: 10.1080/17470910601000446
- 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, 386-396. doi: <http://dx.doi.org/10.1016/j.neuroimage.2015.04.031>
- Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews*, 42, 9-34. doi: <http://dx.doi.org/10.1016/j.neubiorev.2014.01.009>
- Schurz, M., & Tholen, M. G. (2016). What brain imaging did (not) tell us about the Inferior Frontal Gyrus in theory of mind – A commentary on Samson et al.,

(2015). *Cortex*, 74, 329-333. doi:

<http://dx.doi.org/10.1016/j.cortex.2015.08.011>

van der Meer, L., Groenewold, N. A., Nolen, W. A., Pijnenborg, M., & Aleman, A.

(2011). Inhibit yourself and understand the other: Neural basis of distinct processes underlying Theory of Mind. *NeuroImage*, 56(4), 2364-2374. doi: 10.1016/j.neuroimage.2011.03.053

Vogeley, K., Bussfeld, P., Newen, A., Herrmann, S., Happé, F., Falkai, P., . . . Zilles, K.

(2001). Mind Reading: Neural Mechanisms of Theory of Mind and Self-Perspective. *NeuroImage*, 14(1), 170-181. doi: 10.1006/nimg.2001.0789

ACCEPTED MANUSCRIPT