



## BIROn - Birkbeck Institutional Research Online

Hillebrandt, H. and Dumontheil, Iroise and Blakemore, S.J. and Roiser, J.P. (2013) Dynamic causal modelling of effective connectivity during perspective taking in a communicative task. *NeuroImage* 76 , pp. 116-124. ISSN 1053-8119.

Downloaded from: <https://eprints.bbk.ac.uk/id/eprint/6562/>

*Usage Guidelines:*

Please refer to usage guidelines at <https://eprints.bbk.ac.uk/policies.html>  
contact [lib-eprints@bbk.ac.uk](mailto:lib-eprints@bbk.ac.uk).

or alternatively

Manuscript Number: NIMG-12-2331R2

Title: Dynamic Causal Modelling of effective connectivity during perspective taking in a communicative task

Article Type: Regular Article

Section/Category: Social Neuroscience

Corresponding Author: Mr. Hauke Frerk Hillebrandt,

Corresponding Author's Institution:

First Author: Hauke Hillebrandt

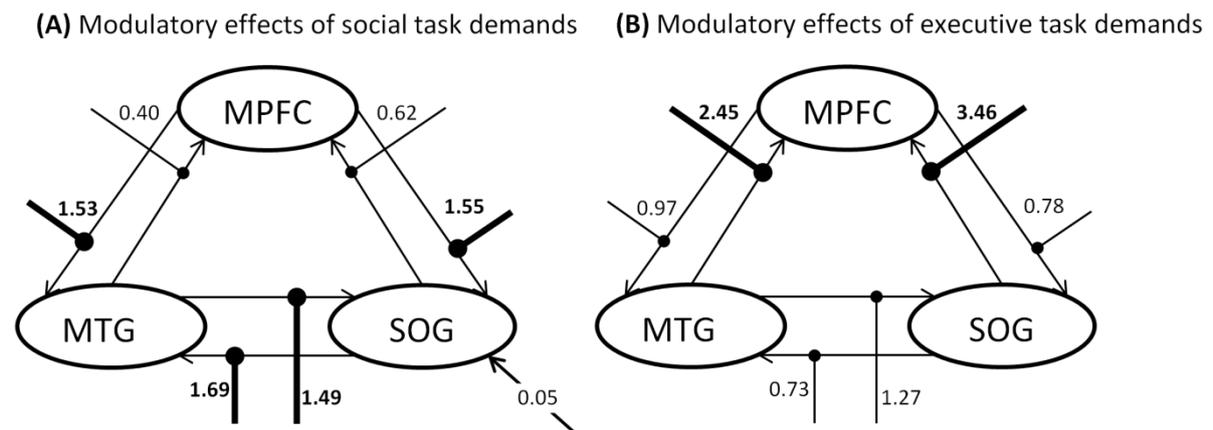
Order of Authors: Hauke Hillebrandt; Iroise Dumontheil; Sarah-Jayne Blakemore; Jonathan P Roiser

Abstract: Previous studies have shown that taking into account another person's perspective to guide decisions is more difficult when their perspective is incongruent from one's own compared to when it is congruent. Here we used dynamic causal modelling (DCM) for functional magnetic resonance imaging (fMRI) to investigate effective connectivity between prefrontal and posterior brain regions in a task that requires participants to take into account another person's perspective in order to guide the selection of an action. Using a new procedure to score model evidence without computationally costly estimation, we conducted an exhaustive search for the best of all possible models. The results elucidate how the activity in the areas from our previously reported analysis (Dumontheil et al., 2010) are causally linked and how the connections are modulated by both the social as well as executive task demands of the task. We find that the social demands modulate the backward connections from the medial prefrontal cortex (MPFC) more strongly than the forward connections from the superior occipital gyrus (SOG) and the medial temporal gyrus (MTG) to the MPFC. This was also the case for the backward connection from the MTG to the SOG. Conversely, the executive task demands modulated the forward connections of the SOG and the MTG to the MPFC more strongly than the backward connections. We interpret the results in terms of hierarchical predictive coding.

##### Highlights

- We investigated effective connectivity with dynamic causal modelling for fMRI
- We used a new exhaustive search procedure to find the best of all possible models
- Social demands modulate the backward and lateral connections most strongly
- Conversely, executive task demands modulated the forward connections most strongly
- We interpret the results in terms of Bayesian hierarchical predictive coding

## Graphical abstract



Effective connectivity analysis shows that backward and lateral connections are modulated more strongly by social task demands (A); whereas forward connections are modulated more strongly by executive task demands (B).

**Running head:** CONNECTIVITY OF PERSPECTIVE TAKING

**Dynamic Causal Modelling of effective connectivity during perspective taking in a communicative task**

Hauke Hillebrandt<sup>1</sup>, Iroise Dumontheil<sup>1</sup>, Sarah-Jayne Blakemore<sup>1</sup>, Jonathan P. Roiser<sup>1</sup>

1. Institute of Cognitive Neuroscience, University College London

**Corresponding author:**

Hauke Hillebrandt

UCL Institute of Cognitive Neuroscience, 17 Queen Square, London, WC1N 3AR, UK.

**E-mail:** [hauke.hillebrandt@gmail.com](mailto:hauke.hillebrandt@gmail.com)

**Tel:** +00 44 20 7679 1126

**Fax:** +00 44 20 7813 2835

Words in abstract: 226

Words in text: 6232

**Abstract**

Previous studies have shown that taking into account another person's perspective to guide decisions is more difficult when their perspective is incongruent from one's own compared to when it is congruent. Here we used dynamic causal modelling (DCM) for functional magnetic resonance imaging (fMRI) to investigate effective connectivity between prefrontal and posterior brain regions in a task that requires participants to take into account another person's perspective in order to guide the selection of an action. Using a new procedure to score model evidence without computationally costly estimation, we conducted an exhaustive search for the best of all possible models. The results elucidate how the activity in the areas from our previously reported analysis (Dumontheil et al., 2010) are causally linked and how the connections are modulated by both the social as well as executive task demands of the task. We find that the social demands modulate the backward connections from the medial prefrontal cortex (MPFC) more strongly than the forward connections from the superior occipital gyrus (SOG) and the medial temporal gyrus (MTG) to the MPFC. This was also the case for the backward connection from the MTG to the SOG. Conversely, the executive task demands modulated the forward connections of the SOG and the MTG to the MPFC more strongly than the backward connections. We interpret the results in terms of hierarchical predictive coding.

**Keywords:** Dynamic causal modelling, effective connectivity, Theory of mind, perspective taking, social brain, exhaustive modelling, hierarchical predictive coding.

## Introduction

Verbal and non-verbal social interactions both rely on an understanding of other people's mental states, also called Theory of Mind (ToM) or mentalising (Frith and Frith, 2007, 2012; for an excellent review of the extensive research in theory of mind literature we refer the reader to Apperly, 2011). During social interactions, in a complex real-world environment, ToM enables individuals to take decisions and choose actions that are appropriate to the present situation and the inferred mental states of the other people involved. Recent research suggests that it is important to investigate not only ToM development but also how individuals are able to efficiently use ToM information during decision making and reasoning (Samson and Apperly, 2010), and the distinction between ToM-specific processes and executive control (e.g. Saxe et al., 2006; Scholz et al., 2009; Van Overwalle, 2009, 2011; Dumontheil et al., 2012; Meyer et al., 2012).

### ***Studying the neural mechanisms of social cognition with the "Director" task***

Keysar and colleagues designed a paradigm to investigate real-world social decision-making, in which participants are faced with a real set of shelves containing objects that are either visible or not visible from the viewpoint of a "director" (a confederate; Keysar et al., 2000, 2003; Lin et al., 2010). The director asks participants to move objects in the shelves and critical instructions require the participant to use information about the director's viewpoint to interpret his instructions correctly. In this Director task, around 50% of the time adult participants fail to use information about the director's perspective and instead erroneously use their own (egocentric) viewpoint when trying to follow instructions (Keysar et al., 2003, 2000). These results were replicated using a computerised version of the task (Apperly et al., 2010; Dumontheil et al., 2010). The Director task differs from other ToM tasks in that it requires participants both to have a functioning ToM to compute the perspective and intentions of another person (the director), and to use this ToM information in

Connectivity of perspective taking  
concert with other cognitive processes such as attentional and inhibitory control to overcome their  
egocentric bias and select the appropriate response quickly and accurately (Apperly et al., 2010).

In a previous fMRI study, we employed an adapted version of this Director task (Dumontheil et al., 2010), which in contrast to previous studies that were designed to assess error rates, included extensive task instructions such that participants performed at high levels of accuracy. As in the behavioural version of the task (Keysar et al., 2000), participants followed auditory instructions to move objects in a set of shelves. Using this modified paradigm for fMRI, we found that: (1) selection of an appropriate action when faced with alternatives (Object factor) was associated with domain-general bilateral brain activations located primarily in the frontal and parietal cortices, with additional activations in the inferior temporal cortex; (2) the processing of social information vs. symbolic cues (Director factor) was associated with specific activations in the dorsal medial prefrontal cortex (MPFC) and superior temporal sulcus; (3) the use of social cues as opposed to symbolic cues for the selection of the appropriate action from the alternative options (interaction) was associated with further recruitment of dorsal MPFC and middle temporal gyri, extending into the left temporal pole (Dumontheil et al., 2012, 2010).

Thus, part of the network of brain regions implicated in social cognition, specifically the MPFC and temporal cortex (Brothers, 1990; Frith and Frith, 2007; Van Overwalle, 2009), was recruited when the guiding information was of a social nature compared to more arbitrary symbolic stimuli.

Research using visual search paradigms suggests that the prefrontal cortex (PFC) supports the integration of information from the current environment and internal representations, thereby providing a “top-down” influence (i.e. intentionally driven by knowledge, expectations and goals) on attentional orientation and action selection appropriate with current goals (Burgess et al., 2007; Fuster, 2008, 2000; Koechlin and Summerfield, 2007), in contrast with stimulus-driven “bottom-up”

Connectivity of perspective taking mechanisms (Beck and Kastner, 2009; Hahn et al., 2006). Therefore one interpretation of these findings is that the dorsal MPFC, similarly to lateral parts of the PFC, may play a role in providing a top-down influence for the selection of the correct target among distractors when the relevant guiding information is in the social domain. To test this hypothesis we examined the top-down and bottom-up influences of social and executive manipulations on network coupling during the Director task, using Dynamic Causal Modelling (DCM: Friston et al., 2003).

***The current study: dynamic causal modelling***

DCM estimates the experimental modulation of forwards and backwards connections between regions that are active during a particular task in a directional manner, and thus makes it possible to infer whether experimental manipulations affect top-down or bottom-up influences. We refer to forward and backward connections in the framework of hierarchical predictive coding, in which sensory input is passed forward and processed in the brain hierarchically, from primary sensory to secondary sensory areas, then on to association areas and finally to higher (frontal) areas (Clark, 2012; Friston, 2010, 2005). We used DCM to investigate coupling between frontal, temporal and occipital brain regions (which represented the aforementioned hierarchy in descending order) involved in the Director task, and its modulation by social cues, using fMRI data from a group of adults (Dumontheil et al., 2010).

An important methodological advance in our analysis is the use of a new post hoc model selection procedure (Rosa et al., 2012) to find (1) the best model out of all possible connection architectures with Bayesian model selection (BMS), (2) posterior probabilities resulting from family level inferences testing whether a parameter exists or not, and (3) Bayesian parameter averages (BPA) over all possible models showing how strong fixed connections were and how much they were modulated. Until recently, DCM required very specific hypotheses about the structure of the model

Connectivity of perspective taking (e.g. which connections are modulated by the experimental manipulations). This is because the estimation of each different model takes a few seconds and with increasing number of nodes in each model the combinatorial explosion of possible models that makes it prohibitively expensive in computational terms to estimate all possible models in model space. Instead, we used a new method to find the model evidence for all possible models without estimating them (Friston and Penny, 2011; Friston et al., 2011; Rosa et al., 2012). This approach permits the selection of the winning model as well as family level inferences (Penny et al., 2010) over all possible models to find (1) the probability of certain connections existing and (2) whether these connections are modulated by the experimental manipulations.

We hypothesized that, while occipital and temporal cortex regions process the social aspects of the stimuli in a bottom-up manner (faces and bodies of the directors), the MPFC is involved in the computation, maintenance, and use of perspective information to guide the selection of an appropriate action. These processes are recruited in the Director present vs. Director absent conditions, where the role of the MPFC may be particularly important in the 3-object condition, which requires, on half of the trials, the inhibition of the prepotent bottom-up responses related to one's own perspective.

## **Material and methods**

### ***Participants***

Fourteen adult (mean age 24.9 years, standard deviation (SD) 3.0, range 21.3–30.6) right-handed female volunteers included in Dumontheil et al. (2010) were considered for DCM analysis, of which 11 were included in the final analysis (see *Volume of Interest Extraction section below*). All participants spoke English fluently and had no history of psychiatric or neurological disorders.

Participants gave written informed consent and the study was approved by the University College London ethics committee.

### ***Experimental design***

Our paradigm includes two manipulations embedded in a  $2 \times 2$  factorial design with the factors Director (“Director present” vs. “Director absent”) and Object (“3-object” vs. “1-object”). In the Director present conditions, two directors are shown, one female and one male. This enabled the participant to identify easily which director was speaking by the sound of their voice. One director stands behind the shelves, facing the participant, while the other stands on the same side of the shelves as the participant. The position of the male and female directors changed within blocks and was counterbalanced between conditions and within and between participants. Therefore the gender of the directors was not confounded with the different experimental conditions. In the 3-object conditions, the instructions refer to an object that is one of three exemplars in the shelves; the correct object to move depends on which director is speaking and whose perspective to take (see **Figure 1A**). Thus in the Director present 3-object trials, participants need to use the social cues, i.e. the position of the speaking director, to select and move the appropriate object. On half of the Director present 3-object trials the perspective of the director issuing the instruction is different from that of the participant; on the other half the director’s and participant’s perspectives are the same. This is varied on a trial-by-trial basis, and thus participants need to consider the director’s perspective on every trial. Note that this is not an experimental factor (our analyses collapsed across these trial types) but a manipulation that ensures participants integrate trial-specific cues. In Director present 1-object trials, there is no need to take into account the director’s perspective to identify the correct object (e.g. “Move the turtle left”), as there are no distractors or other referents; this resembles a bottom-up, visual pop-out as opposed to an effortful top-down visual search

Connectivity of perspective taking (Buschman and Miller, 2007). The Director absent conditions were logically equivalent to the Director present conditions, but the directors were replaced by symbolic cues (see **Figure 1B**).

Stimuli consisted of sets of  $4 \times 4$  shelves with objects located in half of the shelves. Five of the shelves had a grey background (**Figure 1**; see Dumontheil et al. (2010) for details). On each trial, participants were given instructions via headphones, by either a male or a female voice, to move one of the eight objects in the shelves to a different slot, either up, down, left or right (note that this was the participant's left or right). A  $2 \times 2$  factorial within-subject design was used with the factors Director (Present vs. Absent) and Object (1-object vs. 3-object) varying between blocks.

#### *Director factor*

In the Director present conditions the display included two directors, one female and one male. In the Director absent conditions, there were no directors in the display (**Figure 1**). Instead, the letter “F” for female and “M” for male were shown beside the shelves. Below each of the letters there was either one transparent box, which indicated to participants that only objects in open shelves should be moved; or two boxes, one grey and one transparent, which indicated that there was no restriction on the participant's choice and all objects (both in open shelves and occluded shelves) could be moved (see **Figure 1B** for an example). These rules had precisely the same consequences as the position of the director in the Director present conditions. In Director present conditions the physical position of the director issuing the instruction varied on a trial-by-trial basis; similarly, in Director absent conditions the M/F rules changed on a trial-by-trial basis.

#### *Object factor*

Instructions in *1-object* conditions (e.g., in **Figure 1**, “Move the turtle left”) referred to a unique target object (there was only one turtle), which was in an open shelf. Instructions in *3-*

Connectivity of perspective taking *object* conditions (e.g., “Move the large ball up”) could refer to an object in a closed shelf (with a grey background) or an object in an open shelf, which could both be described with the same instruction (e.g. “large ball”). Which of the possible referents was in fact correct was determined by whether the director giving the instruction (identified as male or female by his/her voice) was at the back or front of the shelves (in Director present), or whether the cues indicated that only objects in open shelves could be moved (in Director absent). This manipulation ensured that in Director present 3-object blocks participants had to consider the director's perspective (which was different from their own perspective on 50% of the trials) in order to know which was the correct object to move. In Director present 1-object blocks the director's perspective made no difference to the correct interpretation of his or her instructions, and thus participants could use their own perspective to select the appropriate object on all trials. In the Director absent condition, perspective taking was not involved. On each trial the visual stimulus and the auditory instruction were presented over a period of 2.2 s, after which the display remained on the screen for another 3.8 s. Four such trials formed one block in our block design and each scanning session consisted of 16 task blocks lasting 24.8 s and four fixation blocks lasting 20 s. Participants responded with their right hand using a trackball mouse (see Dumontheil et al. (2012, 2010); for more detailed methods information).

### ***fMRI data acquisition***

3D T<sub>1</sub>-weighted fast-field echo structural images and multi-slice T<sub>2</sub>-weighted echo-planar volumes with blood-oxygen level dependent (BOLD) contrast (35 axial slices with a voxel resolution of 3 × 3 × 3 mm covering the whole brain; TR = 3 s; TE = 50 ms; TA = 2.9143 s) were obtained using a 1.5 T MRI scanner (Siemens TIM Avanto, Erlangen, Germany). Functional imaging data were acquired in three scanning sessions (one session of one participant was discarded due to low accuracy) lasting approximately 8 min 40 s each in which 174 volumes were obtained. The first 2 volumes of each

Connectivity of perspective taking session were discarded to allow for  $T_1$  equilibrium effects. A  $T_1$ -weighted anatomical image lasting 5 min 30 s was acquired after the first two functional sessions for each participant.

## **Data analysis**

### *fMRI data*

fMRI data were analysed using Statistical Parametric Mapping (SPM8 for the GLM and SPM12a for DCM, [www.fil.ion.ucl.ac.uk/spm8](http://www.fil.ion.ucl.ac.uk/spm8)). Detailed methods of the preprocessing and General Linear Model (GLM) analysis of the fMRI data can be found in Dumontheil et al. (2012). Here we report results of data reanalyzed with concatenated sessions for the purpose of DCM. Thus, the results of the GLM differ slightly due to the concatenation of sessions. Briefly, images were realigned, slice timing corrected, and normalised to a standard EPI template based on the Montreal Neurological Institute (MNI) reference brain. The resulting  $3 \times 3 \times 3$  mm images were spatially smoothed with an 8-mm Gaussian kernel. The time series were modelled with boxcar regressors of the instructions, fixation, four types of task blocks, and with one event-related regressor representing error trials. In addition, we included constant session effects. Appropriate stimulus functions were convolved with the canonical hemodynamic response function to form regressors. Together with regressors representing residual movement-related artifacts and the mean over scans, these regressors comprised the full model for each session. A flexible factorial 2<sup>nd</sup> level analysis was performed to identify significant regional effects for the Director present – Director absent contrast and the Director  $\times$  Object interaction contrast ([Director present 3-object – Director present 1-object] – [Director absent 3-object – Director absent 1-object], masked inclusively by [Director present 3-object – Director present 1-object]) to ensure that any interactions observed were driven by effects in the Director present condition (see **Table 1**).

### *Volume of Interest Extraction*

Two volumes of interest (VOIs) were extracted based on the peaks of the Director present > Director absent contrast (see Results section for GLM results and Dumontheil et al. 2010). The first VOI was located in the left superior occipital gyrus (SOG) (-12 -94 4, note that this peak coordinate was labelled as left cuneus in Dumontheil et al., 2010, because the activation was more diffuse due to a less conservative statistical threshold). The domain general function of the SOG is visual perception relevant for motor control (Lui et al., 2006). Activation in this region is observed in more specific social contexts such as during the perception of bodies (Kret et al., 2011). The second VOI was the left middle temporal gyrus (MTG) (-45 -70 13). The MTG domain general function is thought to be motion processing and multisensory integration (Hein and Knight, 2011; Onitsuka et al., 2004), and is recruited in more specific social cognitive tasks during face and speech processing (for a review see, Hein and Knight, 2011). A third VOI was extracted based on the most active peak from the (masked) interaction contrast. This VOI was located in the left dorsal MPFC (-9 38 34). The MPFC is thought to support the processing of one's own and others' mental states (Amodio and Frith, 2006; Van Overwalle 2009).

Timeseries from VOIs associated with different contrasts were summarised using the SPM12 Eigenvariate toolbox: we extracted each participant's principal eigenvariate around the participant-specific local maxima activation nearest to the peak voxel of the group (2<sup>nd</sup> level) GLM analysis (see **Table 1**). All VOIs were taken from the left hemisphere, which showed higher levels of activity, possibly due to the language demands of the task, the fact that participants responded using a trackball mouse held in their right hand and/or the fact that the directors were presented in the right visual field. For the purpose of VOI extraction the normalised images were smoothed with a 4 mm Gaussian kernel to improve anatomical accuracy. The radius of the VOI spheres was 8 mm and the search radius for local maxima from the group analysis was restricted to 16 mm for the main effect and 20 mm for the interaction, which had a more diffuse activation. All voxels contributing to the

Connectivity of perspective taking eigenvariates were significant at  $p < 0.05$  uncorrected and adjusted at  $p < 0.05$ , for the effects of interest (i.e. only for those regressors that were used in the DCMs for input or modulation). We were unable to create VOIs for 3 participants from the interaction contrast, as they did not show any activation above threshold within the search radius. The interaction contrast had less power (due to a lower number of trials in each level contrast) and showed weaker effects overall than the main effect. Hence, we conducted all analyses on 11 of the initial 14 participants.

### *Specification of dynamic causal models*

We created and estimated DCMs (Friston et al., 2003) with DCM12 (version 4750) as implemented in SPM12a. The DCM estimation routines in SPM12a differ slightly from those in SPM8 and DCM10, in that the hyperpriors have been adjusted to reflect a more realistic signal-to-noise ratio in regional (VOI) timeseries. The DCMs were based on the VOIs reported above and used the main effects of director and object to modulate the connections between regions. All our DCMs were deterministic (as opposed to stochastic, see Daunizeau et al., 2012), bilinear (as opposed to nonlinear, see Stephan et al., 2008), two-state models (Marreiros et al., 2008), with mean-centred inputs. Two-state DCMs differ from one-state models in that activity in one brain region is modelled so that it has both excitatory and inhibitory neuronal populations, and introduces positivity constraints that allow extrinsic (between regional) populations influences of one region on another to only be excitatory (Marreiros et al., 2008). To simplify the models and ease interpretation, we disallowed the modulation of self-connections within each region. In the present study the diagonals in DCM.B matrix, which describes the change in coupling strength (Friston et al., 2003), were thus set to zero, while all other connections were set to one. Self-connections were only present in the DCM.A matrix, which represents the *fixed* connection strength between areas (also referred to as endogenous, direct, context-independent, or average connectivity; see Friston et al., 2003). The driving input into the model – represented by the DCM.C matrix (Friston et al., 2003) – was the main effect of Director

Connectivity of perspective taking present vs. absent. This driving input entered the most posterior region, the SOG. Our hypothesis was that this region would be the first region showing sensitivity to the presence of cues (faces and bodies), indicating the Director present vs. absent, and that it would subsequently influence activity in more anterior regions.

### *Post-hoc Bayesian model selection*

Until recently, one aspect of DCM was that one had to have specific hypotheses about the structure of the model (e.g. which connections are modulated by the experimental manipulation, Stephan et al., 2010). A model space with  $n$  nodes has  $2^{n \times n}$  permutations of connections that can be turned on or off, which can be modulated by different experimental manipulations (Friston and Penny, 2011). This combinatorial explosion makes it prohibitively expensive to estimate a large number models (Friston et al., 2007). Instead, we used a new method to find the model evidence for all possible models by only inverting (estimating) the full model (Friston & Penny, 2011; Rosa et al., 2012) to select the winning model. This was achieved by searching over all possible models to find the probability of certain connections existing and estimating whether certain connections are modulated by certain experimental manipulations or not (Friston & Penny, 2011; Rosa et al., 2012).

Below, we first present the results of Bayesian model selection comparing all models at once to find the winning model. Then we present the model posterior probability over parameters (with and without a given free parameter) of whether a fixed connection or a particular modulation exists at all using family-level inferences (Penny et al., 2010). The posterior probability is the probability that a model (or family of models) provides the best explanation for the measured data across participants (Penny et al., 2004). The log-evidences for all subsequent analysis were pooled in a fixed effects fashion, because we assumed that the underlying model structure did not vary across the participants for whom VOI timeseries could be extracted. Then we present the average BPA

Connectivity of perspective taking parameter estimates for the model with the highest evidence (the winning model) to elucidate the quantitative nature of the connection e.g. how much a connection is modulated or how much fixed connectivity there is (Friston et al., 2003). BPA computes a joint posterior probability density over parameter estimates for a group of participants, by using the posterior from one participant as the prior for the next participant, whose posterior then serves as the prior for the next participant etc. (Kasess et al., 2010; Stephan et al., 2010). Note that the fixed and modulatory parameters were always scale parameters (exponentiated) to ensure positivity as per convention for two-state DCMs, so that the extrinsic connections are always excitatory (Marreiros et al., 2008). Scale parameters of two-state DCMs are thus higher than parameter estimates from one-state DCMs. Our unexponentiated parameter estimates ranged from -1.3 to 1.8 Hz, similar to one-state DCM parameter estimates reported in other studies (Goulden et al., 2012; Rosa et al., 2012).

In addition to the average BPA estimates, the post-hoc optimisation also provides BPA parameter estimates for individual participants that can be compared with conventional frequentist statistics (Stephan et al., 2010). We supplemented our Bayesian inference with analysis of variance of individual changes in connection strengths to demonstrate the consistency of these effects across subjects. Critically, post-hoc routines and the conventional variational free energy approach have been shown to yield very similar results (Rosa et al., 2012).

## **Results**

### ***General Linear Model results***

As reported previously (Dumontheil et al., 2010), the Director present > Director absent contrast, performed with a whole-brain voxel-level FWE-corrected threshold of  $p < .05$ , showed increased BOLD signal in the left middle temporal gyrus (MTG VOI), left superior occipital gyrus (SOG VOI), and right middle temporal gyrus (**Table 1**). We performed three small volume corrections (FWE  $p < .05$ ) in

Connectivity of perspective taking the left posterior superior temporal sulcus (STS)/temporo-parietal junction (TPJ), left temporal pole and left MPFC on the basis of coordinates from previous studies (Dumontheil et al., 2010, for details). This revealed significant interactions in the left middle temporal gyrus and left dorsal MPFC (**Table 1**); the latter was the most significant after small volume correction. Hence, this region was selected as a VOI for the DCM analyses.

### ***Dynamic Causal Modelling results***

DCMs were created using the three VOIs described above: SOG; MTG; and MPFC in the left hemisphere (**Table 1**). The post-hoc analysis (see Methods for details) finds the best model and also furnishes the posterior probability of whether individual parameters exist or not. The latter is equivalent to family comparison inferences, which test whether a family of models without a certain parameter (e.g. a connection between two areas) has a higher probability than the family of models with this parameter (Penny et al., 2010). Finally, we show the results of the parameter estimates using the BPA of the winning models parameter estimates.

### ***Bayesian model selection and family-level inferences***

We first assessed the model with the best evidence (a metric in which model fit is traded off against model complexity). A comparison of the evidence for all possible models showed that the winning (optimal) model with the highest probability had a probability of 0.30 (**Figure 2A**). The winning model was the full model that had all connections and all modulations (**Figure 2B**). The next most probable model's probability was 0.27. When comparing this model to the winning model, the Bayes factor is only 1.11. In Bayesian statistics, this would not be considered strong evidence for the full model over the next probable model. In frequentist, classical hypothesis testing this would correspond to a non-significant difference. However, traditional Bayesian as well as frequentist approaches are used for smaller model spaces. In large model spaces like ours it is to be expected

Connectivity of perspective taking that the posterior mass is diluted over a high number of models and no single model has a very high probability (Rosa, Friston, & Penny, 2012), since all models that share some characteristics with the full winning model will have a non-zero probability in the presence of any noise. For instance, using the post-hoc approach with synthetic data, Rosa et al. (2012) were able to obtain the model they knew a priori to be the true model in a model space that was even smaller than ours (half the size) and a Signal to Noise Ratio (SNR) of 2.6. However, when they compared this winning true model (that the data was generated from) to the next probable (false) model the resulting Bayes factor was just 1.94 (which is comparable to ours). **Figure 2A** shows the (log-)posterior probability of all models examined (256 models overall). Crucially, the “best model” approach used in Bayesian model selection is very useful but can become brittle (as in software brittleness; see Penny et al. 2010), when one compares a large number of models (Penny et al. 2010) so, in addition, we performed a family level inference.

When assessing the probabilities for each parameter existing or not using family-level comparison (Penny et al., 2010), we found evidence for reciprocal fixed connectivity between all three regions and modulation of all connections. The results of the family-level inference of all models showed that the posterior probability (over parameters) for reciprocal fixed connectivity between all the VOIs was (almost) 1. Moreover, the evidence that these connections were all modulated by both contrasts (main effects of director and object) was also (almost) 1.

Since the full model had the highest evidence, this indicates that none of the alternative, reduced models outperformed the full model. Moreover, because the family-level inferences also showed that all fixed connections were modulated by both modulators, the full model was selected for further analyses of the parameter estimates in order to examine the strengths of the fixed connectivity and the strengths of the modulation by experimental manipulation.

*Comparison of connection strength*

We tested whether there were quantitative differences between the selected model's Bayesian parameter estimates to find out whether connections between different areas were subject to different modulation. To achieve this, we conducted repeated measures ANOVAs with SPSS 20.0 for both modulators (Director present vs. Director absent and 3-object vs. 1-object) on the individual participants' (BPA) parameter estimate values of the optimal model with connection as a factor with 6 levels. While the two state-DCMs use exponentiated scale parameters that introduce positivity constraints and are so more plausible to interpret, these values are likely to not be normally distributed and heteroscedastic, because the exponential function is the inverse function of the natural logarithm (which is commonly used to transform data to meet the assumption of a normal distribution, see (Bland and Altman, 1996a, 1996b)). Thus, we used the original unexponentiated non-scale parameter estimates for these statistics (see **Table 2**), but plot the exponentiated values in **Figure 3 and 4a-b**. When significant, the main effect of connection was followed-up by paired post-hoc comparisons between the connections.

The 6 (connections)  $\times$  2 (modulator) repeated measures ANOVA showed no significant main effects of connection or modulator, but a significant connection  $\times$  modulator interaction ( $F(5,50) = 8.26, p < 0.001$ ). When considering the modulators separately, there was a main effect of connection for both the Director present vs. absent modulation ( $F(5,50)=4.26, p=0.003$ ), and the 3-object vs. 1-object modulation ( $F(2.4,24.4)=8.93, p=0.001$ ; Greenhouse-Geisser corrected). Post-hoc tests showed that connections from the MPFC to the SOG and MTG, and the lateral reciprocal connections between them, were more strongly modulated by the Director present vs. absent modulation (**Table 2**; only post-hoc comparisons of connections with connections going to the MPFC are displayed in **Table 2**) than were the connections to the MPFC. By contrast, the 3-object vs. 1-object manipulation

Connectivity of perspective taking modulated connections leading to the MPFC more strongly than those projecting from this region or the reciprocal connections between the SOG and MTG (**Table 2**).

## **Discussion**

We used DCM to explore effective connectivity between SOG, the MTG, and the MPFC, in a task that required participants to take into account another person's perspective in order to guide action selection. We used a novel post-hoc model selection routine (Friston and Penny, 2011; Rosa et al., 2012) to look at all possible dynamic causal models. Family-level inference provided strong evidence for reciprocal fixed connectivity between all three areas and modulations by both the Director and the Object factors, which respectively manipulate the nature of the stimuli (social vs. non-social) and the need for top-down influences on action selection.

### *Fixed and modulatory connectivity*

Our family-level inferences suggest that there is high probability of reciprocal fixed connectivity between all three areas that was context-independent and thus driven by the input of the Director present vs. absent manipulation. Our hypothesis was that, similar to the role played by lateral parts of the PFC during tasks that are novel or not automatic and require attentional control and action selection (Burgess et al., 2007; Fuster, 2008, 2000; Koechlin and Summerfield, 2007), the MPFC plays a role in the control of attention and action selection when the guiding and informative cues are of a social nature. We thus predicted that, in the condition that requires perspective-taking information to be used to identify the correct target object (Director present 3-object), we would observe increased effective connectivity from the MPFC towards the SOG and MTG. We found strong evidence that all connections were modulated by both the social and executive manipulations – however, we were also able to disambiguate the relative contribution of modulatory strength on different connections by the two different experimental manipulations.

*Social manipulation*

Consistent with our prediction, analysis of the modulatory effects revealed that the presence of a social cue (Director present vs. absent contrast) increased the strength of the backward connections from the MPFC more strongly than the forward connections from the SOG and the MTG.

Interestingly, this was also the case for the backward connection from the MTG to the SOG. The dorsal MPFC consistently shows increased activation in response to social stimuli and mentalising, in particular when participants are thinking about others' thoughts, intentions or personality traits, and has been suggested to hold the representation of other people's mental states (Tamir and Mitchell, 2010). The left MTG region was close to TPJ which is also consistently implicated in social cognition and perspective taking (Frith and Frith, 2006). Our interpretation of these findings is that the social context leads the dorsal MPFC and the MTG to increase activation in regions processing socially-relevant information, and also to increase the coupling between these regions.

*Attentional and executive control manipulation*

The 3-object vs. 1-object factor increased bottom-up connectivity by modulating the forward connections of the SOG and the MTG to the MPFC more strongly than the backward connections. This may reflect the transmission of the socially relevant information (location of the directors, presence or not of the grey background on the two possible target objects) to the MPFC for further computation of viewpoint and intention of the relevant director. In 3-object trials, the position of the directors, and the relative position or relative size of the three objects needed to be considered to infer the intention of the director correctly. This was associated with longer reaction times (see Dumontheil et al., 2010) and increased demands of processing both the social and non-social information. Note, that in previous perspective-taking tasks (Aichhorn et al., 2006; David et al., 2006; Vogeley et al., 2004) the participants only judge whether another person can see a particular object,

whereas in our task, participants have to infer and select which object another person refers to, based on their perspective (Dumontheil et al., 2010).

However, executive demands sometimes increase activity in MPFC (Levy and Wagner, 2011) areas and social demands increase activity in MTG and SOG (as evidenced by our fMRI data), which may at least partly account the corresponding modulating effects on the connectivity between these brain areas. Furthermore, the visual search, and the computation and selection of the correct action in the 3-object vs. 1-object condition, which we have summarized as increased executive demands like the inhibition of prepotent responses and working memory demands, is likely to have also required more attention and cognitive resources (Levy and Wagner, 2011). This is partly supported by the activation of regions associated with attention and control in the 3-object vs. 1-object main effect contrast (Dumontheil et al., 2012, 2010; Duncan, 2010). Finally, since SOG and MTG are essentially sensory areas, the top-down predictions to these areas can only serve to mediate the inhibition of prepotent responses by selectively reducing the attentional capture by incorrect but prepotent stimuli (Ji and Neugebauer, 2012; Levy and Wagner, 2011; Sharp et al., 2010).

#### *Bayesian hierarchical predictive coding during social behaviour*

Our results can be interpreted within the framework of recent theories that attempt to explain (social) behaviour in terms of Bayesian hierarchical predictive coding (Brown and Brüne, 2012; Clark, 2012; Friston, 2010). Predictive coding postulates that forward, bottom-up connections pass prediction error signals with (unexpected) sensory information about the stimulus from 'lower' (sensory) areas to areas 'higher' in the cortical hierarchy (Friston, 2010). In this framework, the top-down backward connections pass predictions based on an internal (generative) model about the stimulus to lower sensory areas to minimize sensory prediction error (by selectively sampling the stimulus array) and to induce behavioural responses (Clark, 2012; Friston, 2010). In our 3-object factor, the stimulus array has to be sampled more extensively than in the 1-object condition like in

Connectivity of perspective taking traditional top-down visual search paradigms (see for instance, Buschman and Miller, 2007) to find the correct target ('which ball is referred to?) amidst distractors, and so 'surprising' unpredicted sensory information about distractors is passed forward in the cortical hierarchy as prediction error, unlike in bottom-up searches (where a ball 'pops out' among trucks). Lastly, the main effect of director modulates particularly strongly not only the backward connections (potentially by passing down predictions about the director), but also modulates what is usually characterized as a forward connection from SOG to MTG (**Figure 4A**). Because the task requires multisensory integration of visual stimuli with auditory instructions, the connection from SOG to MTG might be a backward connection where the SOG predicts auditory responses in the MTG in a top-down fashion.

### *Limitations*

Our study has some limitations. First, by using the DCM post-hoc procedure, like Rosa et al. (2012), our posterior mass was diluted over a large number of models and so we were not able to reach a statistically significant Bayes Factor when comparing our winning model with the next probable model in our large model space (see results section for more details). This is important to consider for future research using the post-hoc estimation routines with large model spaces. Furthermore, we did not test for any potential nonlinear modulatory gating effects that the activity in brain regions exerts on the connections between them (Stephan et al., 2008) in order to simplify the model and ease interpretability. For the same reason, we chose to look only at brain connectivity in the left hemisphere and omitted extracting a VOI around the right MTG, which was active during Director present vs. Director absent contrast. Indeed, as we only included three left hemisphere regions recruited in the social (Director present) condition, and not right hemisphere regions or regions responsive to the executive demands, our more constrained model would have failed to capture mediating effects and connections between hemispheres and between social and executive regions enabling task performance.

*Summary*

Our results suggest that effortful search and selection of the correct mental state of other people for the purpose of action selection involves lower sensory areas and multisensory integration areas (SOG, MTG, Hein and Knight, 2011) as well as prefrontal areas (MPFC). We provide evidence for an account in which social predictions based on an internal model modulate backward connections, whereas when a perspective diverges from one's own perspective one has to search for the correct perspective, which modulates forward connections.

## **Acknowledgements**

We are grateful to K. Friston for valuable discussion and comments on the manuscript. This study was funded by grants from the Royal Society to S.J.B., a Fellowship from the European Commission to I.D., a University College London scholarship and a Charlotte and Yule Bogue Research Fellowship to H.H.

## References

- Aichhorn, M., Perner, J., Kronbichler, M., Staffen, W., Ladurner, G., 2006. Do visual perspective tasks need theory of mind? *NeuroImage* 30, 1059–1068.
- Apperly, I.A., 2011. *Mindreaders: The cognitive basis of “Theory of Mind”*. Psychology Press, Hove, UK.
- Apperly, I.A., Carroll, D.J., Samson, D., Humphreys, G.W., Qureshi, A., Moffitt, G., 2010. Why are there limits on theory of mind use? Evidence from adults’ ability to follow instructions from an ignorant speaker. *Q J Exp Psychol (Colchester)* 63, 1201–1217.
- Beck, D.M., Kastner, S., 2009. Top-down and bottom-up mechanisms in biasing competition in the human brain. *Vision research* 49, 1154–65.
- Bland, J.M., Altman, D.G., 1996a. Statistics Notes: Transforming data. *BMJ* 312, 770–770.
- Bland, J.M., Altman, D.G., 1996b. Statistics notes: Transformations, means, and confidence intervals. *BMJ* 312, 1079–1079.
- Brothers, L., 1990. The social brain: A project for integrating primate behavior and neurophysiology in a new domain. *Concepts in Neuroscience* 1, 27–61.
- Brown, E.C., Brüne, M., 2012. The role of prediction in social neuroscience. *Front Hum Neurosci* 6, 147.
- Burgess, P.W., Gilbert, S.J., Dumontheil, I., 2007. Function and localization within rostral prefrontal cortex (area 10). *Philos Trans R Soc Lond B Biol Sci* 362, 887–899.
- Buschman, T.J., Miller, E.K., 2007. Top-Down Versus Bottom-Up Control of Attention in the Prefrontal and Posterior Parietal Cortices. *Science* 315, 1860–1862.
- Clark, 2012. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*. Sec. 3, 1–86.
- Daunizeau, J., Stephan, K.E., Friston, K.J., 2012. Stochastic dynamic causal modelling of fMRI data: Should we care about neural noise? *Neuroimage* 62, 464–481.
- David, N., Bewernick, B.H., Cohen, M.X., Newen, A., Lux, S., Fink, G.R., Shah, N.J., Vogeley, K., 2006. Neural representations of self versus other: visual-spatial perspective taking and agency in a virtual ball-tossing game. *J Cogn Neurosci* 18, 898–910.
- Dumontheil, I., Hillebrandt, H., Apperly, I.A., Blakemore, S.-J., 2012. Developmental Differences in the Control of Action Selection by Social Information. *Journal of Cognitive Neuroscience* 24, 2080–2095.
- Dumontheil, I., Küster, O., Apperly, I.A., Blakemore, S.-J., 2010. Taking perspective into account in a communicative task. *NeuroImage* 52, 1574–1583.
- Duncan, J., 2010. The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends Cogn. Sci. (Regul. Ed.)* 14, 172–179.
- Friston, K., 2005. A theory of cortical responses. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 360, 815–836.
- Friston, K., 2010. The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138.
- Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., Penny, W., 2007. Variational free energy and the Laplace approximation. *Neuroimage* 34, 220–234.
- Friston, K., Penny, W., 2011. Post hoc Bayesian model selection. *Neuroimage* 56, 2089–2099.
- Friston, K.J., Harrison, L., Penny, W., 2003. Dynamic causal modelling. *Neuroimage* 19, 1273–1302.
- Friston, K.J., Li, B., Daunizeau, J., Stephan, K.E., 2011. Network discovery with DCM. *Neuroimage* 56, 1202–1221.
- Frith, C.D., Frith, U., 2006. The Neural Basis of Mentalizing. *Neuron* 50, 531–534.
- Frith, C.D., Frith, U., 2007. Social cognition in humans. *Curr. Biol* 17, R724–732.
- Frith, C.D., Frith, U., 2012. Mechanisms of Social Cognition. *Annual Review of Psychology* 63, 287–313.
- Fuster, J., 2008. *The Prefrontal Cortex, Fourth Edition*, 4th ed. Academic Press.
- Fuster, J.M., 2000. Executive frontal functions. *Exp. Brain Res.* 133, 66–70.

- Goulden, N., Elliott, R., Suckling, J., Williams, S.R., Deakin, J.F.W., McKie, S., 2012. Sample size estimation for comparing parameters using dynamic causal modeling. *Brain Connect* 2, 80–90.
- Gusnard, D.A., Raichle, M.E., 2001. Searching for a baseline: Functional imaging and the resting human brain. *Nature Reviews Neuroscience* 2, 685–694.
- Hahn, B., Ross, T.J., Stein, E.A., 2006. Neuroanatomical dissociation between bottom-up and top-down processes of visuospatial selective attention. *Neuroimage* 32, 842–853.
- Hein, G., Knight, R.T., 2011. Superior Temporal Sulcus—It’s My Area: Or Is It? *Journal of Cognitive Neuroscience* 20, 2125–2136.
- Ji, G., Neugebauer, V., 2012. Modulation of medial prefrontal cortical activity using in vivo recordings and optogenetics. *Mol Brain* 5, 36.
- Kasess, C.H., Stephan, K.E., Weissenbacher, A., Pezawas, L., Moser, E., Windischberger, C., 2010. Multi-subject analyses with dynamic causal modeling. *Neuroimage* 49, 3065–3074.
- Keysar, B., Barr, D.J., Balin, J.A., Brauner, J.S., 2000. Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science* 11, 32–38.
- Keysar, B., Lin, S., Barr, D.J., 2003. Limits on theory of mind use in adults. *Cognition* 89, 25–41.
- Koechlin, E., Summerfield, C., 2007. An information theoretical approach to prefrontal executive function. *Trends in cognitive sciences* 11, 229–235.
- Kret, M.E., Pichon, S., Grèzes, J., De Gelder, B., 2011. Similarities and differences in perceiving threat from dynamic faces and bodies. An fMRI study. *NeuroImage* 54, 1755–1762.
- Levy, B.J., Wagner, A.D., 2011. Cognitive control and right ventrolateral prefrontal cortex: reflexive reorienting, motor inhibition, and action updating. *Annals of the New York Academy of Sciences* 1224, 40–62.
- Lin, S., Keysar, B., Epley, N., 2010. Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *J. Exp. Soc. Psychol.* 46, 551–556.
- Lui, L.L., Bourne, J.A., Rosa, M.G.P., 2006. Functional Response Properties of Neurons in the Dorsomedial Visual Area of New World Monkeys (*Callithrix jacchus*). *Cereb. Cortex* 16, 162–177.
- Marreiros, A.C., Kiebel, S.J., Friston, K.J., 2008. Dynamic causal modelling for fMRI: a two-state model. *Neuroimage* 39, 269–278.
- Meyer, M.L., Spunt, R.P., Berkman, E.T., Taylor, S.E., Lieberman, M.D., 2012. Evidence for social working memory from a parametric functional MRI study. *Proc. Natl. Acad. Sci. U.S.A.* 109, 1883–1888.
- Onitsuka, T., Shenton, M.E., Salisbury, D.F., Dickey, C.C., Kasai, K., Toner, S.K., Frumin, M., Kikinis, R., Jolesz, F.A., McCarley, R.W., 2004. Middle and Inferior Temporal Gyrus Gray Matter Volume Abnormalities in Chronic Schizophrenia: An MRI Study. *Am J Psychiatry* 161, 1603–1611.
- Penny, W.D., Stephan, K.E., Daunizeau, J., Rosa, M.J., Friston, K.J., Schofield, T.M., Leff, A.P., 2010. Comparing Families of Dynamic Causal Models. *PLoS Comput Biol* 6, e1000709.
- Penny, W.D., Stephan, K.E., Mechelli, A., Friston, K.J., 2004. Comparing dynamic causal models. *NeuroImage* 22, 1157–1172.
- Rosa, M.J., Friston, K., Penny, W., 2012. Post-hoc selection of dynamic causal models. *J. Neurosci. Methods* 208, 66–78.
- Samson, D., Apperly, I.A., 2010. There is more to mind reading than having theory of mind concepts: new directions in theory of mind research. *Infant and Child Development* 19, 443–454.
- Saxe, R., Schulz, L.E., Jiang, Y.V., 2006. Reading minds versus following rules: dissociating theory of mind and executive control in the brain. *Soc Neurosci* 1, 284–298.
- Scholz, J., Triantafyllou, C., Whitfield-Gabrieli, S., Brown, E.N., Saxe, R., 2009. Distinct regions of right temporo-parietal junction are selective for theory of mind and exogenous attention. *PLoS ONE* 4, e4869.
- Sharp, D.J., Bonnelle, V., Boissezon, X.D., Beckmann, C.F., James, S.G., Patel, M.C., Mehta, M.A., 2010. Distinct frontal systems for response inhibition, attentional capture, and error processing. *PNAS*.

- Stephan, K.E., Kasper, L., Harrison, L.M., Daunizeau, J., Den Ouden, H.E.M., Breakspear, M., Friston, K.J., 2008. Nonlinear dynamic causal models for fMRI. *Neuroimage* 42, 649–662.
- Stephan, K.E., Penny, W.D., Moran, R.J., Den Ouden, H.E.M., Daunizeau, J., Friston, K.J., 2010. Ten simple rules for dynamic causal modeling. *Neuroimage* 49, 3099–3109.
- Tamir, D.I., Mitchell, J.P., 2010. Neural correlates of anchoring-and-adjustment during mentalizing. *PNAS* 107, 10827–10832.
- Van Overwalle, F., 2009. Social cognition and the brain: a meta-analysis. *Hum Brain Mapp* 30, 829–858.
- Van Overwalle, F., 2011. A dissociation between social mentalizing and general reasoning. *Neuroimage* 54, 1589–1599.
- Vogeley, K., May, M., Ritzl, A., Falkai, P., Zilles, K., Fink, G.R., 2004. Neural correlates of first-person perspective as one constituent of human self-consciousness. *J Cogn Neurosci* 16, 817–827.

**Table 1: GLM results.**

The Director present > Director absent contrast was performed with FWE correction at  $p < 0.05$ , whole brain level. For the Director  $\times$  Object [(Director present 3-object > 1-object) – (Director absent 3-object > 1-object)], inclusively masked by (Director present 3-object > 1-object), only those regions are listed that survived small volume FWE correction at  $p < 0.05$  based on a 12mm sphere centred on coordinates from previous literature (see Dumontheil et al., 2010 for details). \* and # indicate FWE-corrected values for the whole brain or a small volume (12 mm sphere) respectively.

Area	MNI Coordinates	Cluster size	Z-value	p-value for peak coordinate (FWE-corrected)
<b><i>Director present vs. Director absent</i></b>				
Left middle temporal gyrus	-45 -70 13	16	5.07*	0.006*
Right middle temporal gyrus	60 -64 13	6	4.95*	0.01*
Left superior occipital gyrus	-12 -94 4	9	4.91*	0.012*
<b><i>Director <math>\times</math> Object interaction</i></b>				
Left superior dorsal medial prefrontal cortex	-9 41 34	5	3.96#	0.004#
Left middle temporal gyrus	-36 8 -26	4	3.33#	0.027#

**Table 2: Comparison of modulatory connection strengths.**

Post-hoc paired comparisons of connections strengths of the modulatory effects of the Director present > absent manipulation and the 3-object > 1-object manipulation. Connections leading to the SOG and MTG were significantly more strongly modulated by the Director present vs. absent manipulation than those leading to the MPFC. By contrast, the 3-object vs. 1-object manipulation modulated the connections leading to the MPFC significantly more strongly than those leading to the MTG and SOG. For clarity, we only report the values of regions going to the SOG and MTG for the Director present vs. absent contrast and the values going to the MPFC for the 3-object vs. 1-object here.

**Modulation by Director present vs. absent contrast**

Backward/lateral connection	Forward connection	Mean Difference between backward/lateral - forward connections	Std. Error	<i>p</i> value
MTG to SOG	SOG to MPFC	0.446*	0.121	0.004
	MTG to MPFC	0.530*	0.163	0.009
MPFC to SOG	SOG to MPFC	0.488	0.267	0.097
	MTG to MPFC	0.572*	0.179	0.009
SOG to MTG	SOG to MPFC	0.507*	0.122	0.002
	MTG to MPFC	0.591*	0.209	0.018
MPFC to MTG	SOG to MPFC	0.456	0.214	0.059
	MTG to MPFC	0.540*	0.151	0.005

**Modulation by 3-object vs. 1-object contrast**

Forward connection	Backward/lateral connection	Mean Difference between forward - backward/lateral connections	Std. Error	<i>p</i> value
SOG to MPFC	MTG to SOG	0.759*	0.271	0.019
	MPFC to SOG	1.091*	0.271	0.002
	SOG to MTG	0.896*	0.211	0.002
MTG to MPFC	MPFC to MTG	0.867*	0.203	0.002
	MTG to SOG	0.473*	0.18	0.025
	MPFC to SOG	0.804*	0.151	<0.001
	SOG to MTG	.609*	0.224	0.022
	MPFC to MTG	.581*	0.162	0.005

\*. The mean difference is significant at the .05 level.

**Figure legends**

**Figure 1:** Examples of a 3-object trial in the Director present (A) and the Director absent (B) conditions. In this example, participants hear the instruction: “Move the large ball up” in either a male or a female voice. Two of the objects could correspond to this instruction depending on the director’s viewpoint. The largest ball (or equivalent) was always located in a closed shelf (not visible from the back), while the second largest ball was located in an open shelf. Here, if the voice is female, the object to be moved would be the basketball, since in the Director present condition (A) the female Director is standing in front of the shelves and can see all the objects, while in the Director absent condition (B) the two boxes below the “F” (for “female”) indicate that all objects can be moved by the participant. If the voice is male, the only appropriate response would be to move the football, since in the Director present condition (A) the male Director is standing behind the shelves and therefore cannot see the larger basketball (which was thus a distractor), while in the Director absent condition (B) the single a transparent box below the “M” (for “male”) indicates that only objects in clear slots can be moved by the participant. The third object was another type of distractor that did not fit any of the perspectives (the tennis ball).

**Figure 2:** (A) The left graph shows the range of log-posterior probability of all possible models examined. The right graph shows the posterior probability of all models. Model 256 had the highest posterior probability of 0.3. (B) The winning model: The full model was the winning model, with the highest evidence: in this model all connections were modulated by both modulators. The driving input was fixed and went into the most posterior region, the SOG.

**Figure 3:** (A) VOIs used in the DCM analyses and illustration of the fixed connectivity between them. Two VOIs were based on the peaks of the Director present > Director absent main effect contrast and were in the superior occipital gyrus (SOG; -12 -94 4; circled in blue) and the left middle temporal

Connectivity of perspective taking gyrus (MTG; -45 -70 13; circled in green). One VOI was based on the PFC peak from Director  $\times$  Object contrast [(Director present 3-object > 1-object) – (Director absent 3-object > 1-object), inclusively masked by (Director present 3-object > 1 –object)] and was located in the left dorsal medial prefrontal cortex (MPFC; -9 38 34; circled in red). The colour of the line represents the source of the strongest bidirectional fixed connection, while its width represents its absolute strength (see web version for a colour figure of this graph). (B) Bayesian Parameter estimates under selected (winning) model of fixed connections strength in absence of task specific modulation. Values indicate scale (exponentiated) parameter estimates for each particular connection.

**Figure 4:** Modulatory connectivity. (A) Modulatory effects of the Director present vs. absent contrast: Backward connections coming from the MPFC (highlighted in green) and lateral connections between SOG and MTG are subject to stronger modulation. (B) Modulatory effects of the 3-object vs. 1-object contrast: connections leading to the MPFC (highlighted in red) are subject to stronger modulation. Values indicate scale (exponentiated) parameter estimates for each particular connection. On (A), the input strength to the SOG is shown. The modulatory parameters estimates can be interpreted as percentage change in connection strength (e.g. the model with the highest evidence has a parameter estimate of 1.55 for the modulation of the MPFC-SOG connection, which corresponds to an increase in MPFC-SOG coupling of 55% during the Director present condition).

Figure 1 (color for the Web)

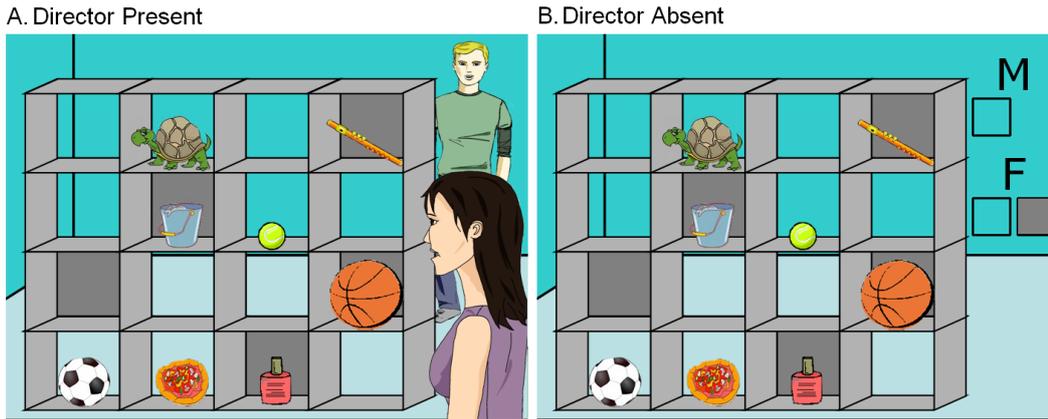


Figure 2

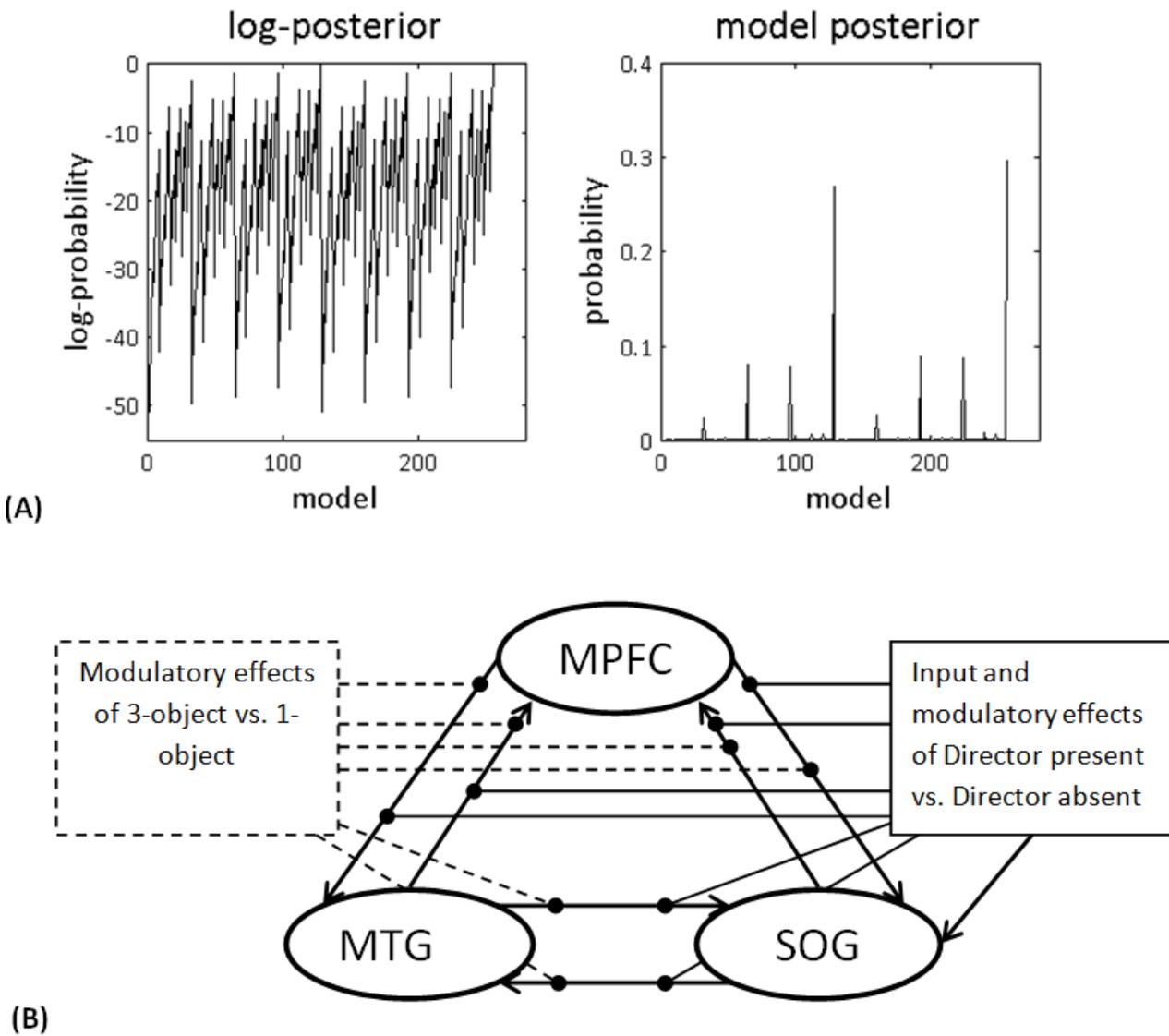
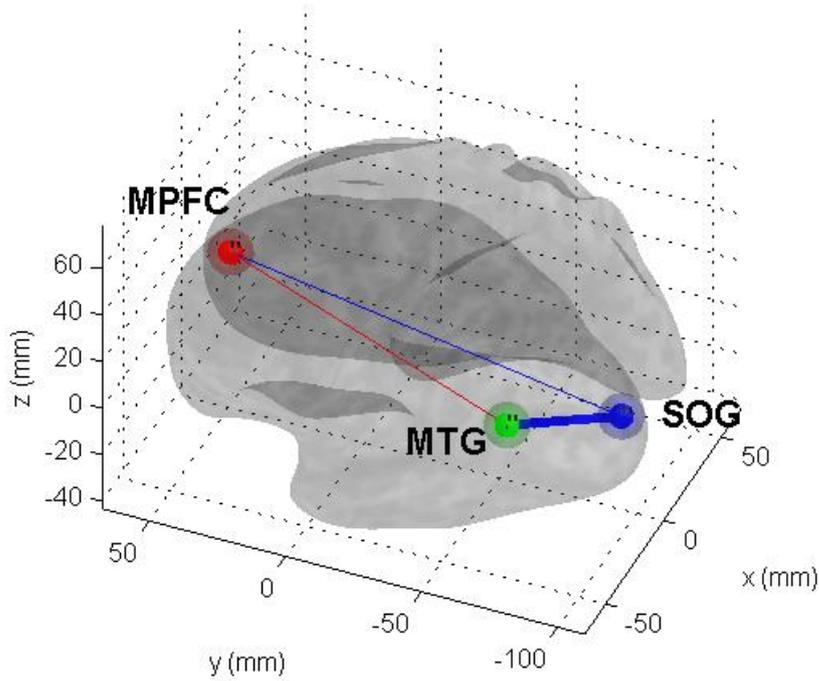
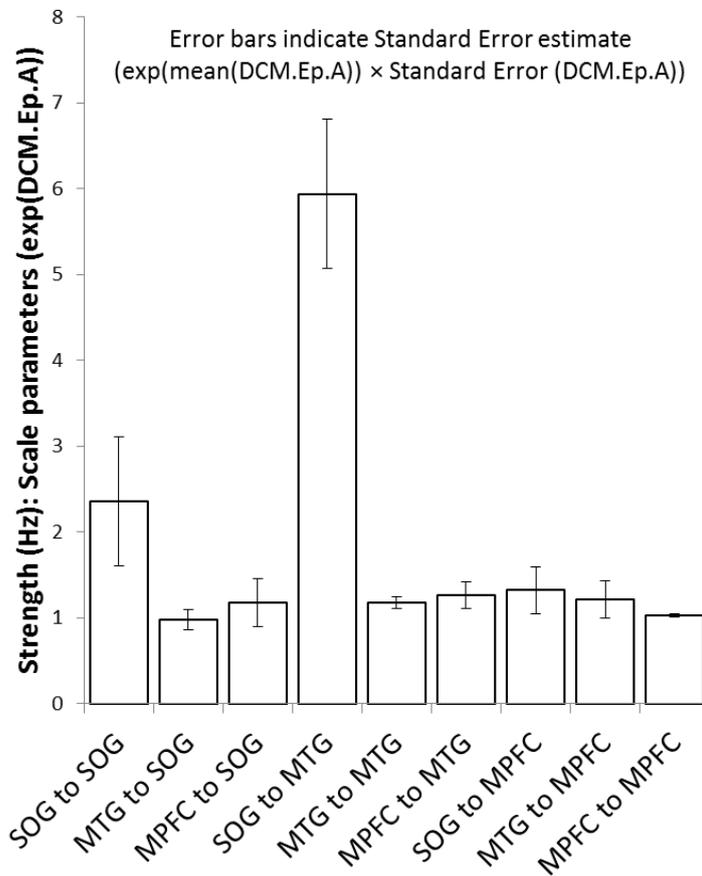


Figure 3 (color for the Web)



(A)

**Fixed connectivity parameter estimates**



(B)

Figure 4

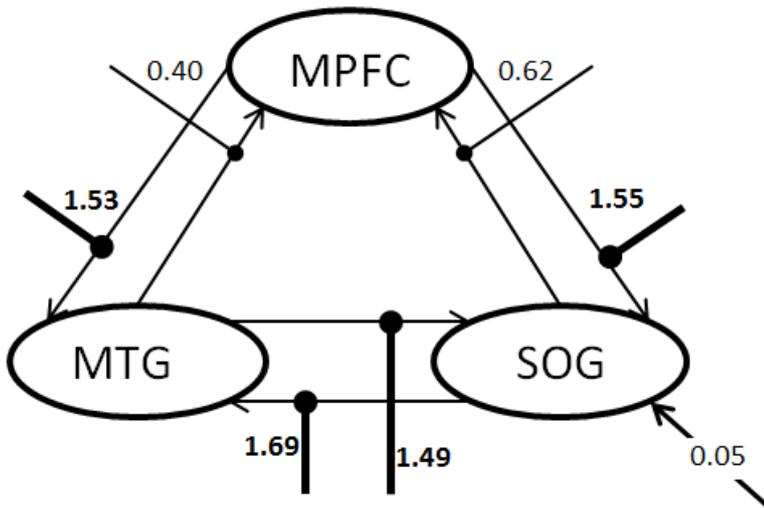
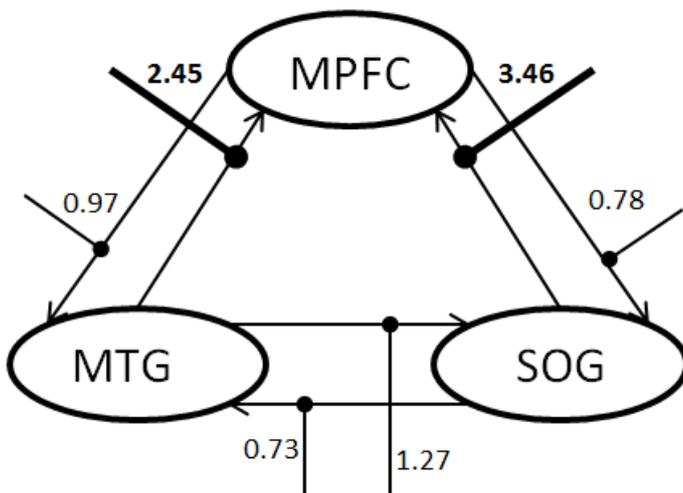
**(A)** Input and modulatory effects of Director present vs. absent**(B)** Modulatory effects of 3-object vs. 1-object

Figure 1 (black-and-white for print)

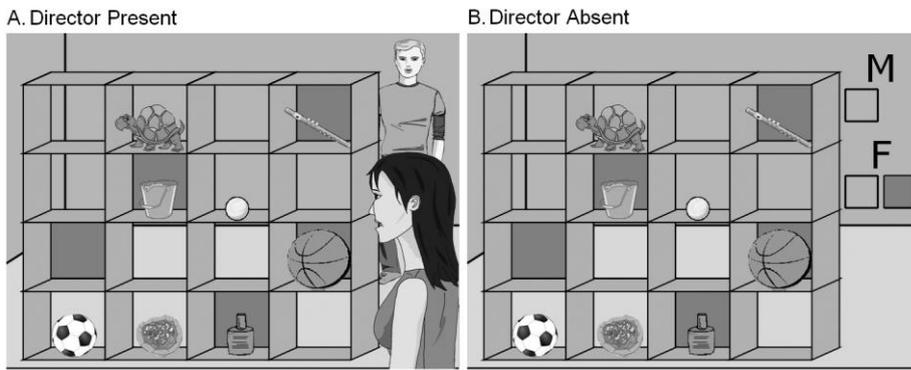
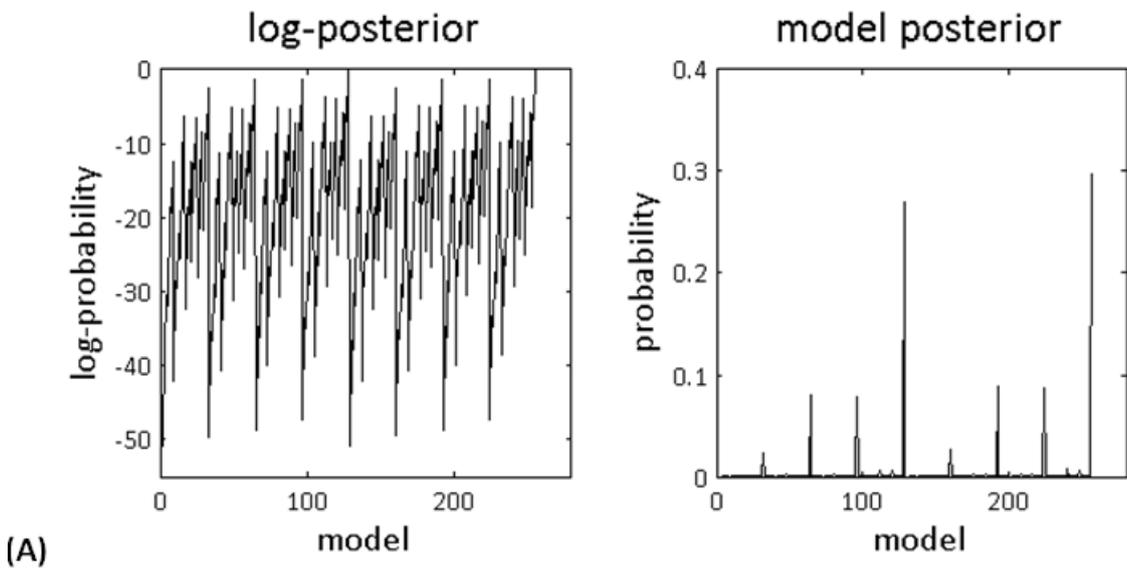
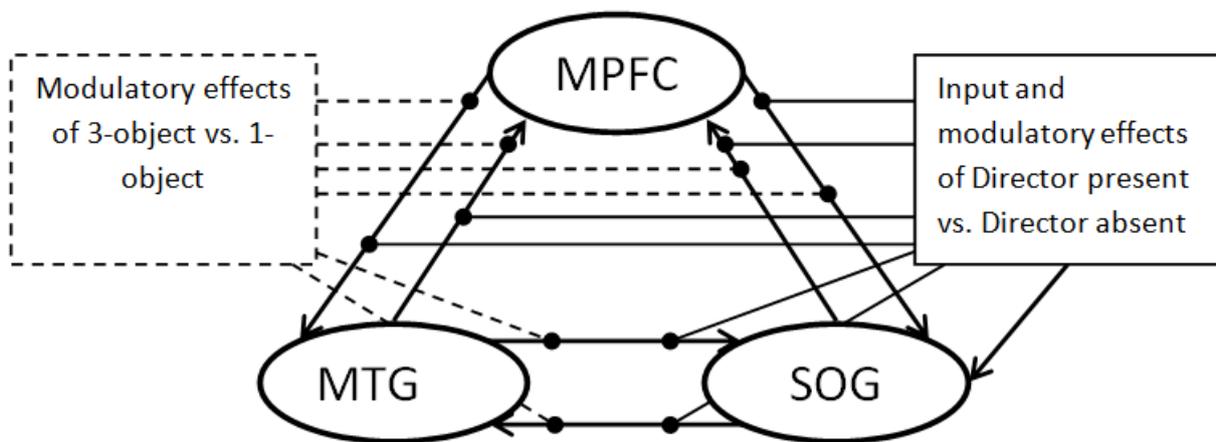


Figure 2 (black-and-white for print)

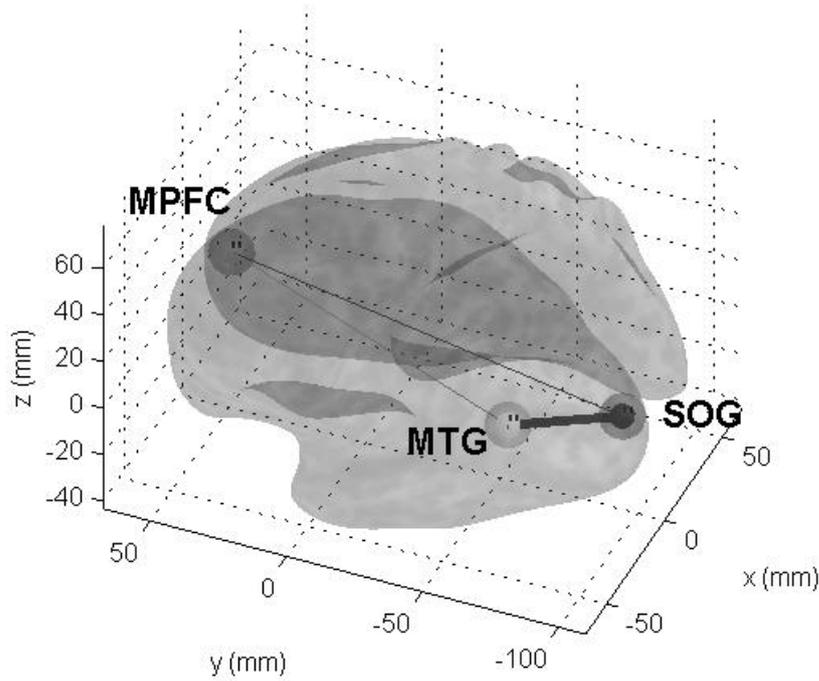


(A)



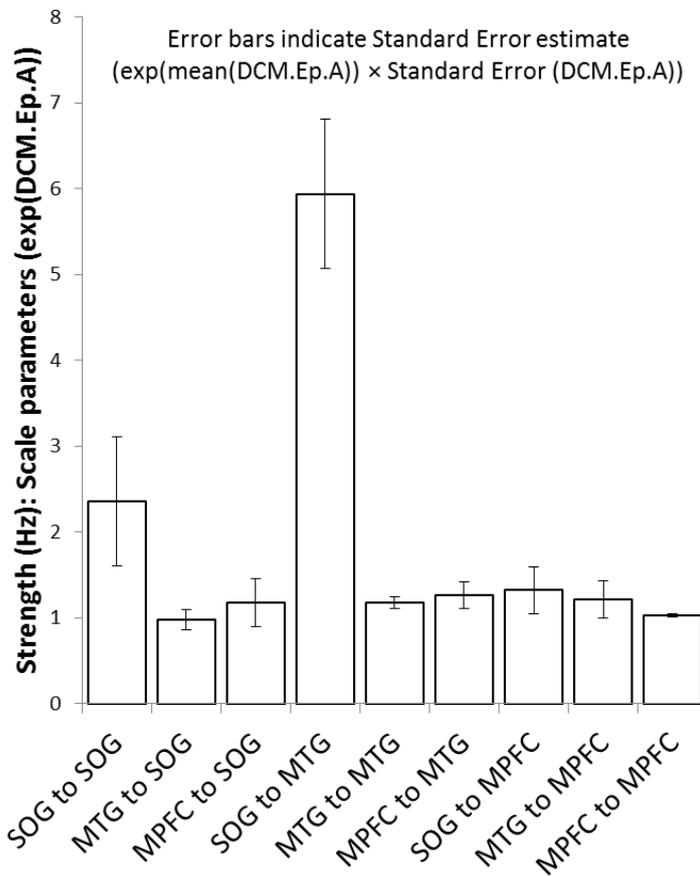
(B)

Figure 3 (black-and-white for print)



(A)

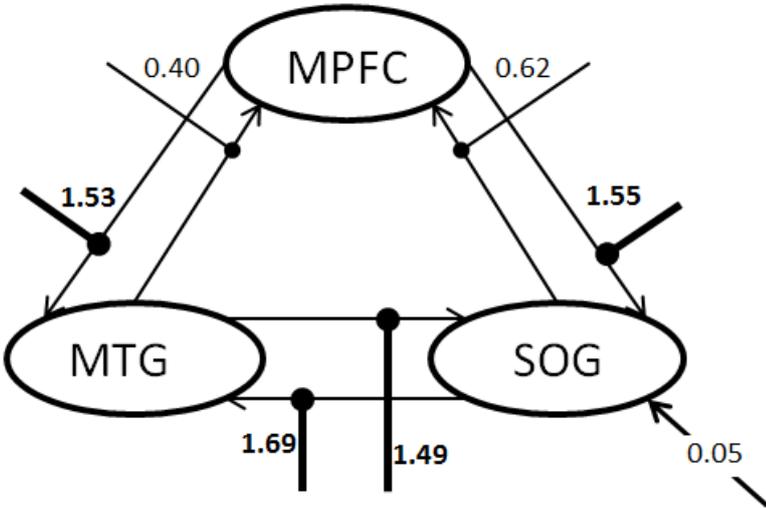
**Fixed connectivity parameter estimates**



(B)

Figure 4 (black-and-white for print)

(A) Input and modulatory effects of Director present vs. absent



(B) Modulatory effects of 3-object vs. 1-object

