NeuroImage 90 (2014) 315–325

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

False belief and counterfactual reasoning in a social environment

Nicole Van Hoeck [*], Elizabet Begtas, Johan Steen, Jenny Kestemont, Marie Vandekerckhove, Frank Van Overwalle [*]

Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium

ARTICLE INFO

Article history:
Accepted 22 December 2013
Available online 30 December 2013

Keywords:
Counterfactual reasoning
Theory of mind
False belief
fMRI

ABSTRACT

Behavioral studies indicate that theory of mind and counterfactual reasoning are strongly related cognitive processes. In a neuroimaging study, we explored the common and distinct regions underlying these inference processes. We directly compared false belief reasoning (inferring an agent's false belief about an object's location or content) and counterfactual reasoning (inferring what the object's location or content would be if an agent had acted differently), both in contrast with a baseline condition of conditional reasoning (inferring what the true location or content of an object is). Results indicate that these three types of reasoning about social scenarios are supported by activations in the mentalizing network (left temporo-parietal junction and precuneus) and the executive control network (bilateral prefrontal cortex [PFC] and right inferior parietal lobule). In addition, representing a false belief or counterfactual state (both not directly observable in the external world) recruits additional activity in the executive control network (left dorsolateral PFC and parietal lobe). The results further suggest that counterfactual reasoning is a more complex cognitive process than false belief reasoning, showing stronger activation of the dorsomedial, left dorsolateral PFC, cerebellum and left temporal cortex.

Introduction

The power of imagination is an extraordinary ability that people use on a daily basis. It supports the ability to predict how someone else experiences the world (social mentalizing), what people's beliefs are, even if they differ from reality. It also provides us the capacities to vividly simulate our personal pasts, and how these past events could have turned out differently (counterfactual thinking). Obviously, both mental simulations have in common that they allow us to imagine complex scenes that differ from reality, as we immediately observe(d) it. When we attempt to infer the beliefs of others, we may hold a different belief about reality at the same time. Also, when we image an alternative past, we are still aware of what really happened. What are the neural commonalities and differences of holding these two mental simulations? Investigating this question is the aim of the current functional imaging study, in order to shed further light on the processes underlying mental simulation.

Thinking about the beliefs of others is referred to as *mentalizing* or *theory of mind*, "the implicit or explicit attribution of mental states to others and self (desires, beliefs) in order to explain and predict what they will do" (Frith and Frith, 2010, p. 289). It is therefore a crucial capacity for successful social interaction. A common way to examine theory of mind is by the use of a false belief task, which assesses one particular aspect of our mentalizing ability: the understanding that other people's actions are determined by their (false) beliefs and not by what actually happened. A prototypical version of this task portrays two protagonists, Sally and Anne (Baron-Cohen et al., 1985). In this story, Sally places a marble into her basket, and after Sally left the room, Anne moves the marble into her own box. Fully matured belief reasoning or mentalizing is reflected in the belief that, upon her return, Sally will look in her basked, although the participants clearly know the new location of the marble (Wellman et al., 2001).

Counterfactual reasoning refers to making inferences about how an event or state could have been different. For instance, we can create a counterfactual state of affairs by changing the true outcome of an event (e.g., passing the exam instead of failing it) and inferring how this could have come about (e.g., "If I had studied the correct chapters"). An important attribute of this type of reasoning is that we adhere to the "nearest possible world" constraint (e.g., Rafetseder et al., 2013; Van Hoeck et al., 2012). During counterfactual reasoning we create a possible event that is nearest to the real event. We don't create a totally new event in which everything is different or impossible (make-believe; e.g. "If I had the best memory in the world, I would have passed the exam") nor do we create a "general possible" event by applying typical regularities (basic conditional reasoning; e.g., studying long enough), because they may not always provide a better outcome (i.e., since you studied the wrong chapters; Rafetseder and Perner, 2010).

What do false belief and counterfactuals have in common? One common process is mentalizing. Like false beliefs, counterfactuals cannot directly be inferred from observation. Both thought processes involve integrating general knowledge and information in a mental

[*] Corresponding authors at: Department of Psychology, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium.
E-mail addresses: Nicole.Van.Hoeck@vub.ac.be (N. Van Hoeck), Frank.VanOverwalle@vub.ac.be (F. Van Overwalle).

1053-8119/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neuroimage.2013.12.043

representation about a specific action or situation away from reality. So instead of only reasoning about reality (the true state of affairs), in a false belief state or a counterfactual state, we mentally simulate an alternative model representing contrary-to-the-fact assumptions. Failure to simulate (i.e., generate and/or reason from) these alternative representations results in "realist errors" (i.e., answering false belief and counterfactual question based on reality).

Another common process is executive control. Both false belief and counterfactual reasoning require extra executive control because simulating an alternative state of affairs entails holding this representation in mind (working memory), updating it with what we know about reality, but suppressing the information about reality that does not apply in the alternative state (inhibition), and the cognitive flexibility to switch between the two representations. Recent research suggests that working memory, cognitive flexibility and inhibitory control each partially mediated the relationship between the performance of children on false belief and counterfactual tasks (Drayton et al., 2011; Guajardo et al., 2009). Several developmental studies further show that children's performance on false belief and counterfactual tasks is positively related (e.g., Perner et al., 2002, 2004; Riggs and Robinson, 1998) and is linked to executive skills that improve between ages 3 and 5, the same time that realist errors decline (e.g., Beck et al., 2009; Carlson et al., 2004; Guajardo et al., 2009; Hughes and Ensor, 2007; Müller et al., 2007).

To increase our understanding of the communalities and differences between false beliefs and counterfactuals, this study directly compares the neurological substrates involved in both inferences. Many neuroimaging studies have demonstrated that false belief reasoning engages a so-called *mentalizing* network, encompassing the temporo-parietal junction (TPJ), medial prefrontal cortex (mPFC), and the posterior cingulate (PCC) or precuneus (Carrington and Bailey, 2009; Lombardo et al., 2011; Mar, 2011; Schilbach et al., 2012; Spreng and Grady, 2010; Spreng and Mar, 2012; Spreng et al., 2009; Van Overwalle, 2009; Van Overwalle and Baetens, 2009). This mentalizing network overlaps considerably with regions recruited during mental simulations of episodic events (including past, future and counterfactual episodic thinking), engaging an *episodic memory network* that comprises additional memory-related regions (including the hippocampus, parahippocampal gyrus, and retrosplenial cortex; Buckner and Carroll, 2007; De Brigard et al., 2013; Hassabis and Maguire, 2007, 2009; Martinelli et al., 2013; Spreng and Grady, 2010; Spreng et al., 2009; Spreng and Mar, 2012; Summerfield et al., 2010; Van Hoeck et al., 2013). Hence, this mentalizing network supports mental simulations of social scenarios and thus might be common to false beliefs and social counterfactual reasoning. We concur with the idea put forward by Hassabis and Maguire (2007) that the underlying process shared by both inferences is concerned with scene construction, or the process of mentally generating and maintaining a complex and coherent event portraying goal-directed actions. In the present study, however, we expect memory-related activations to play a minor role, since we present short descriptions of non-personal events, rather than recalling and reimagining memories of one's past.

In addition, recent research suggests that the flexible retrieval and representation of different belief perspectives during false belief reasoning requires executive control functions that are subserved by the lateral prefrontal cortex (PFC) (Hartwright et al., 2012; Van der Meer et al., 2011). More generally, it has been suggested that executive control functions needed for goal-directed mental simulations (e.g., integrating and monitoring multiple information, updating and implementation of goal-directed behavior, working memory, etc.) are supported by an *executive control* network. This network comprises the dorsal anterior cingulate cortex (dACC) extending to the larger posterior medial frontal cortex (pmFC; detection of multiple or conflicting inputs) and the dorsolateral PFC and inferior parietal lobe (IPL; working memory, attention, integration and general support of cognitive operations) (e.g., Barbey et al., 2012; Botvinick et al., 2004; Cabeza et al., 2012; Chein and Schneider, 2012; Dodds et al., 2011; Miller and Cohen, 2001; Nelson et al., 2010; Schacter et al., 2012; Seeley et al., 2007; Spreng, 2012; Spreng and Grady, 2010; Vincent et al., 2008; Whitman et al., 2013).

To compare false belief and counterfactual inferences, we presented stories in which location or content of the same object was switched in the absence of a protagonist, followed by a question about the content or location tailored to each condition (see Appendix A). As a control condition, we included a basic conditional reasoning task that did not require the simulation of an alternative representation, as participants had to answer a reality-based question based on the (last) object's location or content (see also Rafetseder and Perner, 2010; Rafetseder et al., 2013). This basic conditional reasoning condition controls for a possible confound present in many previous studies measuring counterfactual reasoning, where answers to the counterfactual questions can be inferred by basic conditional reasoning alone. To illustrate, participants in a recent study were asked: "The motor is switched off today. If the motor had been switched on today,… would it have burned fuel?" (Kulakova et al., 2013, p. 267). To answer this question participants don't need to know what the actual situation is, only the general rule that a running motor burns fuel. The current study addresses this issue by ensuring that participants can only answer the counterfactual question correctly by reasoning counterfactually about the specific story and not by applying a general rule.

We hypothesize that all three conditions – which require the representation of goal-directed action by human protagonists – engage the mentalizing network (e.g., Mar, 2011; Van Overwalle, 2009; Van Overwalle and Baetens, 2009). In line with recent work on episodic counterfactual thinking (Van Hoeck et al., 2013) and on the inhibition of self-representations in false belief reasoning (Hartwright et al., 2012; Van der Meer et al., 2011) as described earlier, we further predict that mental simulation of false belief and counterfactuals requires additional cognitive control supported by an executive control network (e.g., Botvinick et al., 2004; Schacter et al., 2012; Spreng, 2012).

Given that the majority of the empirical evidence points to communalities, we do not expect major differences between false belief and counterfactuals. Nevertheless, for false beliefs, we might expect stronger involvement of the TPJ as part of the mentalizing network, in line with the widely accepted notion that this region has specialized involvement in belief reasoning (Saxe and Powell, 2006; Young et al., 2010). For counterfactuals, we might expect stronger activation in the lateral temporal lobe. Although this region is commonly engaged by different forms of mental simulations, studies indicate it is stronger engaged while simulating a new or possible event, in contrast to a factual event (e.g., Addis et al., 2007, 2009; De Brigard et al., 2013; Nieuwland, 2012; Van Hoeck et al., 2013). This may additionally trigger activation in the lateral temporal lobe due to incorporating semantic knowledge into the construction of a novel/hypothetical event (Irish et al., 2012; Schacter et al., 2012; Viard et al., 2011).

Material and methods

Participants

Twenty healthy, right-handed adults (four males and sixteen females, $M_{age} = 22$ years, range = 19–29 years) who did not report a prior history of neurological or psychiatric impairment participated in the study. One participant was removed from analysis because of anatomical irregularities. All participants were native Dutch speakers who had normal or corrected-to-normal vision. They gave written informed consent, in a manner approved by the Medical Ethics Committee at the Hospital of University of Gent and the Free University of Brussels, and were paid 10 euro for their participation.

Stimuli

Participants read vignettes in which the object's content or location was switched in the absence of a protagonist. To illustrate, one of the vignettes describes how Jonas moves his wallet from his trousers to the windowsill, to be then moved again in Jonas' absence by Marion to the kitchen table after paying for pizzas. Next, questions were asked about the object's location or content: during false beliefs — the location expected by the absent protagonist (e.g., where Jonas expects his wallet to be if/when he returns?); during counterfactuals — the location if one of the protagonists had acted differently (e.g., where Jonas' wallet would be, if he had laid out money for pizzas); during basic conditional reasoning — a reality-based question on the last location (see also Appendix A).

The structure and the content of all the stimulus material were carefully controlled. The same 24 short social vignettes were used for all three conditions. Twelve vignettes portrayed the story of an agent changing an object's location while another agent was absent, and the other twelve vignettes described an agent changing an object's content while the other agent was absent. Agent's gender was also balanced across the vignettes. Every vignette was 40 words long and encompassed 3 sentences; a different object location or content in each sentence.

These vignettes were followed by a question tailored to each condition. The false belief and counterfactual questions (experimental conditions) were always 14 words long and the basic conditional questions (control condition) 11 words long. Each question had a similar "If …, then …" structure in the subjunctive past tense (counterfactual condition) or indicative present tense (false belief & basic conditional condition). The first part of the question referred to an action (the first agent returning or one of the agents acting counterfactually). The second part of the question asked for the (expected) location or content of the object. All the questions were phrased in an affirmative way (no negations). The three different object locations or contents in each vignette ensured that participants needed to reason about each specific vignette and question to infer the correct answer.

Procedure

Immediately prior to scanning, the Delis–Kaplan Color–Word Inference Test was administered (Delis et al., 2006). Aside from basic components like color and word naming, it assess response inhibition capacities (naming the ink color and not the word) and cognitive flexibility (switching between naming the color and naming the word).

During the structural scanning, participants received written instructions about the experiment. They first completed a practice session, so that they were familiar with the different types of tasks and with the response box. The experiment itself included just one run. All stimuli were presented in black text on a white background and projected on a screen viewed by participants on a mirror incorporated into the head-coil. E-Prime software (Psychology Software Tools, Inc., Pittsburgh, PA) was used for the presentation and timing of stimuli.

Twenty four False Belief, 24 Counterfactual Reasoning, and 24 Basic Conditional reasoning trials (3 conditions × 24 vignettes = 72 trials) were presented across the entire scanning session. Each trial (see Fig. 1) started with a fixation cross (duration = 2 s) followed by a vignette (self-paced). Participants pressed a button on the response box when they finished reading the vignette. Another fixation cross appeared (jittered between 1 and 4 s), this time followed by a question (self-paced). Participants were explicitly instructed only to press a button once they knew the answer to the question. After a third fixation cross (jittered between 1 and 4 s), three answering options (self-paced) were presented. Participants pressed the button that corresponded to their answer. A fourth fixation cross (duration = 0.5 s) preceded a difficulty rating scale (not difficult–slightly difficult–difficult–very difficult).

The question-phase of each trial was the critical phase that differed between conditions. The specific question was depended on the condition (see Appendix A), while all other trial phases were the same for each condition. During the false belief trials, participants had to answer questions concerning the false belief of the first agent (who was not aware of the change of location or content of an object). Participants had to reason where this agent expected an object was located or what the content of an object was. During the counterfactual trials, participants were required to reason where an object would be located or what the content of an object would be if one of the agents had acted differently. The basic conditional questions asked the participants where the object is located or what the content of the object is when the first agent returns.

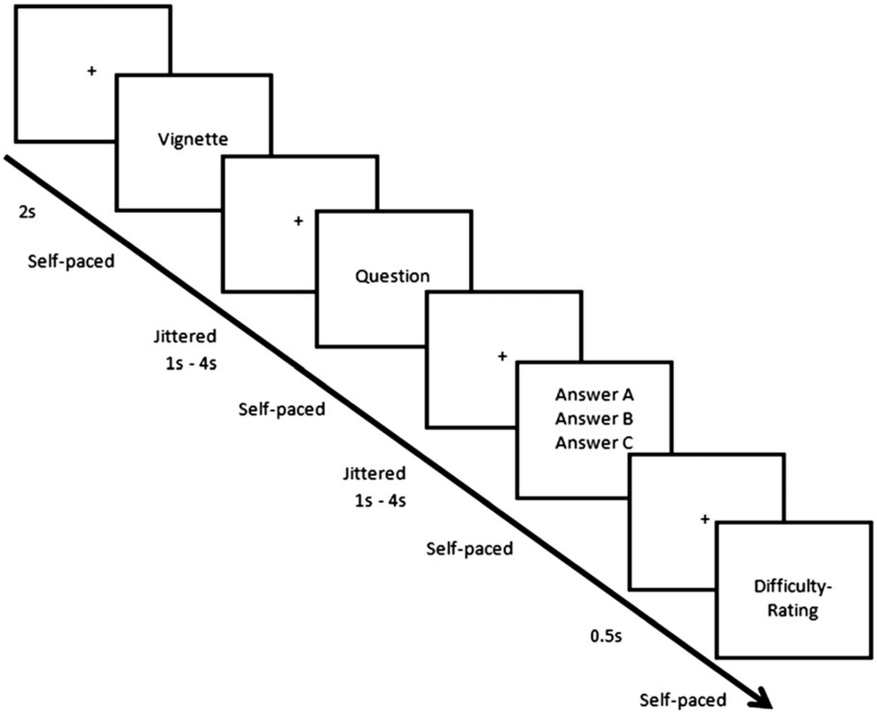

Fig. 1. Design.

Since we had 12 change-of-location and 12 change-of-content vignettes, repeated for each of the 3 conditions, we pseudo-randomized the order of the trials. The 72 trials were split into 3 consecutive blocks of 24 trials, with the order of the 3 blocks being pseudo-randomized. Specifically, the same vignette was never repeated within one block of 24 trials and there were at least 12 trials between repetitions of the same vignette (in the following block). Each block contained the same amount of trials for each condition (3 conditions × 9 trials = 24 trials/block). Half (4 or 5) of the 9 condition-trials in each block contained a change-of-location vignette and the remaining trials, a change-of-content vignette. In addition, half of the counterfactual questions referred to the agent who was not aware of the change of location/content, while in the remaining trials the counterfactual questions referred to the other agent (who had changed the location/content).

During the last 2 min (approximate) of scanning we included a reading localizer baseline task, adapted from Pinel et al. (2007). Participants were instructed to read 15 sentences, each containing between 5 and 8 words and not referring to any mentalizing or causal content. Once participant had read a sentence they were required to press a button, which was followed by a fixation cross (jittered between 1 and 4 s). Post scanning, participants' working memory was assessed by the Reading Span test (Van den Noort et al., 2008). Due to technical difficulties (i.e., participants sometimes pressed the response button too early during stimulus presentation, resulting in the automatic skipping of this trial), this measure was compromised and not included in the analyses.

Imaging procedure

Images were collected with a 3 Tesla Magnetom Trio MRI scanner system (Siemens medical Systems, Erlangen, Germany), using an 8-channel radiofrequency head coil. Stimuli were projected onto a screen at the end of the magnet bore that participants viewed by way of a mirror mounted on the head coil. Stimulus presentation was controlled by E-Prime 2.0 (www.pstnet.com/eprime; Psychology Software Tools) under Windows XP. Foam cushions were placed within the head coil to minimize head movements. We first collected a high-resolution T1-weighted structural scan (MP-RAGE) followed by one functional run of 922 volume acquisitions (30 axial slices; 4 mm thick; 1 mm skip). Functional scanning used a gradient-echo echoplanar pulse sequence (Repetition Time (TR) = 2 s; Echo-Time (TE) = 33 mm; 3.5 mm × 3.5 mm × 4.0 mm resolution).

Image processing and statistical analysis

The fMRI data were preprocessed and analyzed using SPM8 (Statistical Parametric Mapping; The Wellcome Trust Centre for NeuroImaging, London, UK). For each functional run, data were preprocessed to remove sources of noise and artifact. Functional data were corrected for differences in acquisition time between slices for each whole-brain volume, realigned within and across runs to correct for head movement, and co-registered with each participant's anatomical data. Functional data were then transformed into a standard anatomical space (2 mm isotropic voxels) based on the ICBM 152 brain template (Montreal Neurological Institute), which approximates Talairach and Tournoux atlas space. Normalized data were then spatially smoothed (6 mm full-width-at-half-maximum [FWHM]) using a Gaussian kernel. Finally, realigned data were examined, using the Artifact Detection Tool software package (ART; www.nitrc.org/projects/artifact_detect), for excessive motion artifacts and for correlations between motion or global mean signal and any of the conditions. Outliers where identified in the temporal difference series by assessing between scan differences (Z-threshold: 3.0, scan to scan movement threshold: 0.5 mm; rotation threshold: 0.02 radians). These outliers were omitted in the analysis by including a single regressor for each outlier. None of the participants had >10% outliers. No correlations between motion and experimental design or global signal and experimental design were identified.

Statistical analyses involved a first-level single participant mixed epoch-related design, modeled using a canonical hemodynamic response function (GLM). We used a default high-pass filter of 128, and serial correlations were accounted for by the default autoregressive AR(1) model. The design included a regressor for each condition time-locked at the presentation of the question (duration set to individual duration of each question-phase), with the movement and nuisance artifact regressors, and the other trial-phases as regressors of no-interest.

At second-level, we computed several random effect general linear models. A peak-based statistical threshold of $p \leq 0.001$ (uncorrected) was used for all comparisons. In a first model, statistical comparisons between conditions were computed using a one-way within-participants analysis of variance (ANOVA with one factor and four levels) on the parameter estimates associated with each trial type. In addition, we computed two conjunction analysis, one combining the contrasts of all conditions > reading baseline (i.e., false belief > reading localizer, counterfactual > reading localizer and basic conditional > reading localizer) and one combining the two contrasts of the experimental > control conditions (i.e., false belief > basic conditional and counterfactual > basic conditional).

To estimate the covariation between executive functioning and activity in the two experimental conditions, in a second model, we conducted two One Sample T-tests with the false belief > basic conditional contrast parameter (generated at the 1st level) and one executive function covariate (single score per participant) for each *T*-test, and similar two One Sample T-tests with the counterfactual > basic conditional contrast parameter. We tested two executive function covariates at a time: response inhibition (the scaled contrast score of Condition 3–Condition 1 of the Delis–Kaplan Word–Color Inference Task), and cognitive flexibility (the scaled contrast score of Condition 4–Condition 3 in the Delis–Kaplan Word–Color Inference Task). The resulting parameter estimate represents the magnitude of correlation between task-specific activations and the participant-specific covariate measure.

We report statistical comparisons after correction for multiple comparisons using the non-parametric test statistic developed by Slotnick et al. (2003). This procedure enforces a cluster extent threshold by Monte-Carlo simulations of fMRI activation of the entire functional image matrix (64 × 64 × 30 voxels), assuming a corrected type I error voxel activation probability of 0.05 and smoothing with a 3D 6-mm FWHM Gaussian kernel. After 2000 simulations, to yield a corrected $p < 0.05$, the cluster extent was determined at 31 contiguous resampled voxels. We also report which of these clusters were significant after FDR correction at cluster level ($p < 0.05$) and at peak level ($p < 0.05$).

Results

Behavioral results

A one way within-participant ANOVA with the total accurate answers per condition confirmed that the participant's accuracy did not significantly differ between conditions ($F(2, 36) = .66$, $p = .52$).[1] This is not surprising since the average accuracy in each condition is almost perfect (counterfactual: $M = 23.11$, $SD = .22$; false belief: $M = 23.11$, $SD = .24$; basic conditional: $M = 23.37$, $SD = .21$) and

[1] As has been pointed out by Dixon (2008), the assumption of normality (in ANOVA) might often not be met when dealing with proportions, especially when accuracy percentages reach ceiling levels (as is the case here). Moreover, standard errors for effect estimates in within-subject designs tend to be inflated when subject variability is high (Dixon, 2008). As this limitation illustrates the lack of power of traditional repeated measures ANOVAs compared to more modern methods such as generalized linear mixed-effect models (Dixon, 2008; Jaeger, 2008), the latter modeling approach was additionally adopted in order to provide an analysis that is more sensitive to potential condition effects. A mixed logit model with a main effect for condition and random intercepts for subjects and items (Baayen et al., 2008) was fit, but did not reveal any significant differences in accuracy between conditions, $\chi^2(2, N = 19) = 1.22$, $p = 0.54$.

the questions are perceived as easy (overall median = 1 = not difficult) in all the conditions.

The average Question-phase duration (i.e., time to read the question and infer the answer in seconds; counterfactual: $M = 4.72$, $SD = 1.80$; false belief: $M = 4.4$, $SD = 1.92$; basic conditional: $M = 3.73$, $SD = 1.51$) differed however significantly between conditions (one way within-participant ANOVA, $F(2, 36) = 17.96$, $p < .001$). Pairwise comparisons (Bonferroni corrected) indicated that the duration of the basic conditional question-phase was significantly shorter than the false belief and counterfactual question-phase ($p < .001$). The reaction time of the false belief and counterfactual question did not differ significantly ($p = .104$). Aside from the fact that the basic conditional question is 2 words shorter, the shorter response time might be due also to the fact that the reasoning process in the basic conditional condition does not require the simulation and/or manipulation of two state-of-affairs (contrary to the other two conditions), which is supported by language research comparing statements about facts (i.e., indicative mood) versus counterfactual possibilities (i.e., subjunctive mood) (Ferguson, 2012; Santamarı et al., 2005). However, none of the executive function measures (Inhibition: $M = 11.84$, $SD = 2.58$; Switching: $M = 10.37$, $SD = 2.41$) or the difficulty scores correlate significantly with the duration of the question-phase in any of the conditions.

fMRI results

Overlap between false belief, counterfactual and basic conditional reasoning

To estimate the activations common to all three experimental conditions, we computed a conjunction analysis including the false belief > reading localizer, counterfactual > reading localizer and basic conditional > reading localizer contrasts (Table 1 & Fig. 2). The results demonstrate that, as predicted, all three conditions activate regions from the mentalizing network (precuneus and left TPJ) and the executive control network (bilateral PFC, dACC & right IPL). There are a number of additional regions which may reflect increased semantic processing (middle temporal gyrus) as well as increased mental imagery (lingual gyrus, cuneus, left occipital gyrus), perhaps due to the emphasis on the change of locations/contents described in the vignettes in comparison with mere reading (e.g., Kulakova et al., 2013; Spreng et al., 2009).

Overlap between false belief and counterfactual in contrast to conditional reasoning

We next estimated the overlap between the two experimental conditions in comparison with basic conditional reasoning from reality, by computing the conjunction of false belief > basic conditional and counterfactual > basic conditional contrasts (Table 1 & Fig. 2). In support of the prediction that the simulation of an alternative (mental) representation away from reality requires executive processing, the conjunction shows activation of the left IPL. There was additional activation in the occipital cortex (left middle occipital gyrus & cuneus) again pointing to the possibility of increased mental imagery.

False beliefs and counterfactuals

Separate contrasts comparing the two belief conditions each against the basic conditional condition (Table 1 & Fig. 3) confirm the results from the conjunction. Interestingly, they further reveal that both false beliefs and counterfactuals activate the dorsolateral PFC from the executive control network, with counterfactuals more robustly than false beliefs. Counterfactuals, but not false beliefs, activated additional regions from the executive network (pmFC & right IPL, and the mentalizing network (dorsal mPFC). There was also activation in the right cerebellum. More importantly, confirming our prediction, counterfactuals recruited the left inferior temporal gyrus implicated in semantic processing.

We then compared the two experimental conditions directly. A counterfactual > false belief contrast confirmed our prediction of greater activation during counterfactuals in the left temporal cortex (semantic processing), and further revealed stronger activation in the dorsolateral PFC from the executive network (Table 1 & Fig. 3). The opposite false belief > counterfactual comparison did not yield any results. Taken together, this is generally consistent with our hypothesis that differences between the two conditions are rather limited. However, counterfactual reasoning engaged executive control and semantic processes to a larger extent.

Post-hoc we also regrouped the counterfactual trials into a Change of Location condition (12 trials) and a Change of Content condition (12 trials). One can assume that during a Counterfactual Change of Location ("where") requires more visual-spatial processing (scene reconstruction) than Counterfactual Change of Content ("what"). Results of a one-way within-participant ANOVA seem to support this (Table 2). The Counterfactual Change of Location condition activated the posterior regions supporting the process of mentally generating and maintaining a complex and coherent scene: bilateral parahippocampal and fusiform gyrus, left precuneus, bilateral inferior parietal lobule and middle occipital gyrus. The Counterfactual Change of Content revealed activity in the left anterior cingulate, left SMA & middle cingulate cortex, and left postcentral gyrus. These regions have been linked to control process (inhibition and managing multiple representations) during counterfactual comprehension (Urrutia et al., 2012; see also next heading).

Parametric analysis of executive processes

With respect to executive control, the parametric analysis of the false belief > basic conditional contrast (Tables 3 & 4) revealed a positive correlation between the Inhibition measure and areas of the executive network (lateral PFC and right IPL) as well as the right superior temporal gyrus. The lateral PFC was also correlated negatively with the Flexibility measure.

The parametric analysis of the counterfactual > basic conditional contrast (Tables 3 & 4) showed a positive correlation with the Inhibition measure and negative correlation with Flexibility measures in a plethora of regions, including regions of the executive (lateral PFC, IPL) and mentalizing (right TPJ) networks.

Discussion

On a daily basis we make inferences about states that are not directly observable in the external world, like false belief reasoning (inferring an agent's false belief about an object's location or content) and counterfactual reasoning (inferring what the location or content of object would be if an agent had acted differently). Developmental, cognitive and neurological studies all indicate that false belief and counterfactual reasoning are related to each other. However, until present no functional MRI had investigated their common and distinct brain regions. The current study compared the neural underpinnings of these two types of reasoning processes in a social environment and contrasted this to basic conditional reasoning (inferring what the [current] object's location or content is).

This study confirms that false belief, counterfactual and basic conditional reasoning, in contrast to reading, engage processes supporting goal-directed mental simulations in a social context: mentalizing (left TPJ & precuneus) and executive control functions (bilateral PFC & right IPL). We also found increased activation in regions related to mental imagery (lingual gyrus, cuneus, left occipital gyrus) and semantic processing (bilateral middle temporal gyrus). Although not predicted, this increased engagement of simulation processes (Kulakova et al., 2013; Spreng et al., 2009) is understandable because it is most probably due to the emphasis in our vignettes on changes in locations/contents in contrast to our reading localizer which measures general reading processes. In summary, the results of this study provide neurological support for the notion that false belief, counterfactual and basic conditional reasoning in a social context rely on representing the relevant components of the situation (generating a mental representation) and executive processes that control the generation of such representation.

Table 1
Whole brain analysis.

	False belief (FB) > basic conditional (BC)					Counterfactual (CF) > basic conditional					Conjunction: FB > BC + CF > BC					Counterfactual > false belief					Conjunction: FB > RL + CF > RL + BC > RL				
	x	*y*	*z*	*T*	*Voxels*	*x*	*y*	*z*	*T*	*Voxels*	*x*	*y*	*z*	*T*	*Voxels*	*x*	*y*	*z*	*T*	*Voxels*	*x*	*y*	*z*	*T*	*Voxels*
dmPFC						−4	50	48	5.58[ab]	378															
pmFC						−14	28	58	4.38	98															
Left dlPFC	−20	34	40	4.00	38	−38	30	46	3.94	36						−38	30	46	5.81[a]	44	−38	4	52	6.38[ab]	1597[c]
Left vlPFC																					−40	24	26	5.23[ab]	1597[c]
Right vlPFC																					42	24	26	4.11	189
Left dACC																					−6	14	50	4.35	92
Left temporal pole																−40	16	−26	4.05	77					
Left inferior temporal gyrus						−50	−14	−26	4.77	128															
Left fusiform gyrus																−34	−38	−20	4.86	60					
Left middle temporal gyrus																−54	−54	2	4.22	54	−48	−24	−16	3.98	40
																					−52	−42	0	4.26	167
Right middle temporal gyrus																					54	−10	−18	5.05	177
Left temporo-parietal junction																					−48	−54	18	4.72[b]	370
Left inferior parietal lobe	−50	−68	42	3.95	67	−52	−62	42	4.98[b]	356	−50	−68	42	3.95	64										
Right inferior parietal lobe						54	−62	38	3.98	145											36	−66	44	4.67	194
Precuneus																					−4	−64	50	8.42[ab]	3949
Lingual Gyrus & Cuneus	14	−92	0	4.11[b]	271																14	−90	−6	6.06[ab]	257
Cuneus/calcarine gyrus						16	−92	10	5.35 [ab]	653	16	−92	2	4.07[b]	242						−10	−94	−8	6.67[ab]	437[c]
Left middle occipital gyrus	−20	−98	8	5.73[ab]	498	−20	−98	8	7.17[ab]	929	−20	−98	8	5.73[ab]	487										
Left inferior occipital gyrus																					−22	−90	−10	5.87[ab]	437[c]
Right cerebellum						28	−82	−32	3.68	58															

Coordinates in the MNI (Montreal Neurological Institute) stereotactic space of the peak voxel within each cluster as indicated by the highest T-score. The reported clusters survive a whole-brain threshold of $p < 0.001$ and are significant after correction for multiple comparisons according to the Slotnick test statistic (cluster size > 30). Regions denoted by [a] or [b] are also significant after FDR correction at peak [a] or cluster [b] level (SPM8, $p < 0.05$). Regions denoted by [c] and having an equal size cluster belong to one and the same cluster. FB = false belief reasoning; BC = basic conditional reasoning; CF = counterfactual reasoning; RL = reading localizer; PFC = prefrontal cortex; dmPFC = dorsomedial PFC; dlPFC = dorsolateral PFC; vlPFC = ventrolateral PFC; dACC = dorsal anterior cingulate cortex; IPL = inferior parietal lobule. The contrast false belief > counterfactual did not show any significant activation.

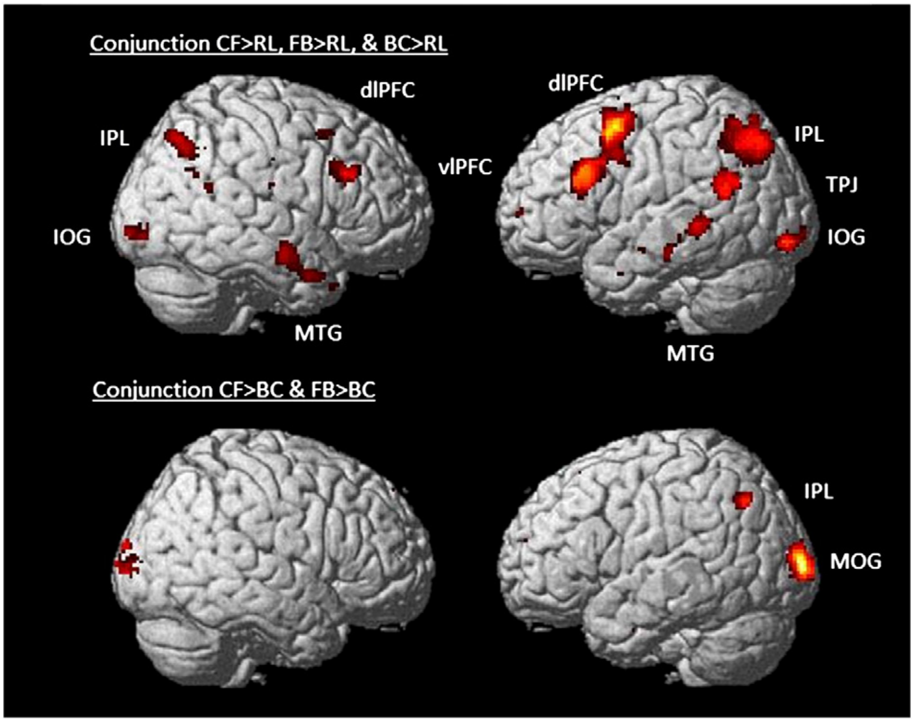

Fig. 2. Conjunction analyses (from the whole brain analyses, $p < .001$ uncorrected, cluster size > 30 voxels). CF = counterfactual reasoning; FB = false belief; BC = basic conditional; RL = reading localizer; vlPFC = ventrolateral prefrontal cortex, dlPFC = dorsolateral prefrontal cortex, TPJ = temporo-parietal junction, IPL = inferior parietal lobule, MTG = middle temporal gyrus, IOG = inferior occipital gyrus, MOG = middle occipital gyrus.

When we compare false belief and counterfactual reasoning against basic conditional reasoning, we further detected increased activity in the executive control network (dorsolateral PFC and IPL; e.g., Barbey et al., 2012; Schacter et al., 2012; Spreng, 2012), although not always in the same cluster or hemisphere. This increased activity confirms our prediction that false belief as well as counterfactual inferences, in comparison to basic conditional inferences, relies heavier on executive processes recruited by the integration and control of cognitive representations. Furthermore, the covariation analysis demonstrated that several executive regions in the frontal and parietal lobule correlated positively with false belief activations, and even more so with counterfactual activations.

This study was not able to provide support for a specialized role of the right temporo-parietal junction (TJP) in reasoning about other people's beliefs (e.g., Saxe and Powell, 2006; Young et al., 2010). We did not observe stronger activation of the right TPJ during false belief reasoning in comparison with the other conditions. We did find left TPJ activation in the conjunction of the three conditions compared to general reading. This confirms the major tenet that the TPJ is essentially involved in social cognition (see meta-analyses by Kubit and Jack, 2013; Mar, 2011; Mars et al., 2012).

An important contribution of the current study is that it suggests that counterfactual reasoning about adding an alternative action to a social scenario is a more complex process than false belief reasoning. First, during counterfactual reasoning, we observed stronger activation of the dorsomedial PFC, a mentalizing region supporting the formation of higher construals, that is, abstractions away from the observable reality (Baetens et al., 2013). Second, in line with our previous study (Van Hoeck et al., 2013), the left dorsal lateral PFC (i.e., executive control network) and the left temporal cortex (i.e., semantic processing) were also stronger activated. Knight and Grabowecky (1995) describe a patient with damage in the dorsolateral PFC whose most marked behavior was a “complete absence of counterfactual expression” (p. 1367). They suggest that impairments in this region lead to deficits in simulation and reality monitoring. Combined with the left temporal cortex, it seems to assist in the generation and integration of elements in a detailed counterfactual representation (Holland et al., 2011; Summerfield et al., 2010). Third, unlike false belief activations, numerous counterfactual task-dependent activations in the frontal and parietal lobule (executive control network) correlated with behavioral measures of executive control reflecting response inhibition and cognitive flexibility. Fourth, there was also enhanced recruitment of the posterior cerebellum. This is in accordance with recent work on episodic counterfactual thinking (Van Hoeck et al., 2013) and a recent review by Van Overwalle and coworkers (Van Overwalle et al., 2013) demonstrating that cerebellar activity increases when there is greater complexity and abstraction in social cognitive judgments. The cerebellum, presumably serves a domain-general cognitive function in the support of executive and semantic processes in complex higher social cognition.

We extensively controlled the materials of the tasks (see method section), and the average difficulty rating, accuracy and time needed to infer the answer did not differ significantly between the false belief and counterfactual reasoning task. Consequently, this study indicates that the simulation of an additive counterfactual social representation (entailing adding a new element to the scenario; versus subtractive counterfactuals, which remove an element from the scenario) is indeed more complex than simulating a false belief representation. To answer a false belief question correctly, one has to simulate a second representation (while the first one represents reality), which only needs to represent the object's location or content right before the protagonist left the scene. However, answering our counterfactual question correctly requires inferring which change in location or content was likely to have been canceled out because of the counterfactual action, which can only be inferred by representing all the components of each specific vignette and question and additional causal reasoning (longer inference length). This is surprisingly analogous to the processes associated with the role of semantic (temporal lobe), executive (PFC & IPL) and mentalizing regions (dmPFC) during self-reflective autobiographical reasoning (i.e., extracting meaning and integrating conceptual knowledge) that goes beyond mere autobiographical retrieval (D'Argembeau et al., 2013). Therefore, the present study indicates that creating specific and detailed counterfactual representations requires additional cognitive

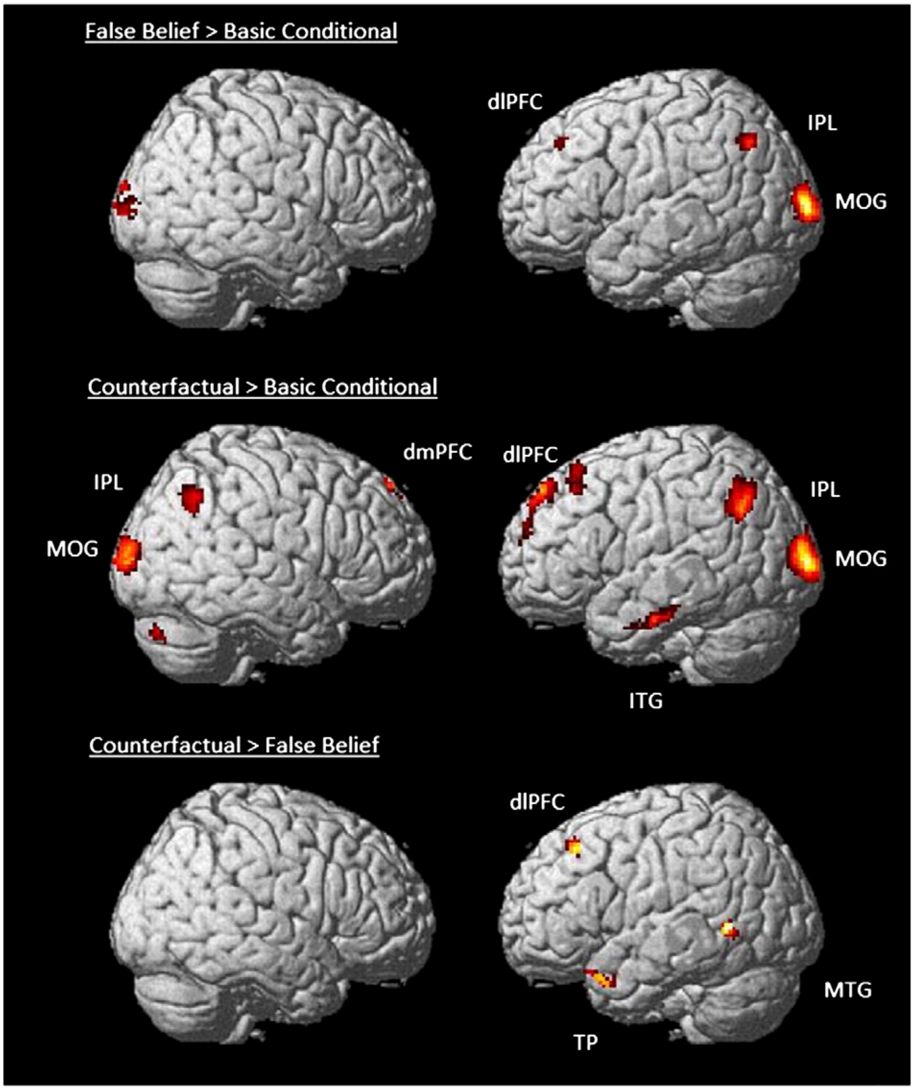
False Belief > Basic Conditional
dlPFC
IPL
MOG
Counterfactual > Basic Conditional
IPL
dmPFC
MOG
dlPFC
IPL
MOG
ITG
Counterfactual > False Belief
dlPFC
MTG
TP

Fig. 3. Contrast analyses (from the whole brain analyses, $p < .001$ uncorrected, cluster size > 30 voxels). dmPFC = dorsomedial prefrontal cortex, dlPFC = dorsolateral prefrontal cortex, IPL = inferior parietal lobule, ITG = inferior temporal gyrus, TP = temporal pole, MOG = middle occipital gyrus, MTG = middle temporal gyrus.

control and semantic processing. This also supports the recent developmental findings suggesting that counterfactual skills develop later than false belief, between the age of 6 and 12 (Rafetseder et al., 2010, 2013). However, future research should investigate if the difference in inference length is indeed inherent to counterfactual versus false belief reasoning, which otherwise limits the interpretation of the results. Additionally, we cannot rule out that linguistic differences add to the complexity of the reasoning process in counterfactual

Table 2
Counterfactual Change of Location versus Change of Content.

	Counterfactual									
	Location > content					Content > location				
	x	*y*	*z*	*T*	*Voxels*	*x*	*y*	*z*	*T*	*Voxels*
Left anterior cingulate cortex						−10	40	14	5.36	34
Left SMA & middle cingulate cortex						−4	6	46	5.29	199
Left postcentral gyrus						−50	−30	50	4.08	50
Right parahippocampal & fusiform gyrus	34	−34	−14	5.03	99					
Left parahippocampal & fusiform gyrus	−30	−36	−16	6.01[b]	257					
Left precuneus &cuneus	−20	−48	10	5.07	33					
	−20	−56	24	4.08	75					
Right inferior parietal lobe	52	−70	34	4.43	93					
Left inferior parietal lobe	−38	−80	42	4.58	152					
Right middle occipital gyrus	−16	−98	−2	4.49	47					

Coordinates in the MNI (Montreal Neurological Institute) stereotactic space of the peak voxel within each cluster as indicated by the highest T-score. The reported clusters survive a whole-brain threshold of $p < 0.001$ and are significant after correction for multiple comparisons according to the Slotnick test statistic (cluster size > 30). Regions denoted by [b] are also significant after FDR correction at cluster level (SPM8, $p < 0.05$).

Table 3
Covariate analysis: positive correlation.

	Response inhibition									
	False belief > basic conditional					Counterfactual > basic conditional				
	x	*y*	*z*	*T*	*Voxels*	*x*	*y*	*z*	*T*	*Voxels*
Right OFC						46	52	−8	4.77	32
Right vlPFC	56	24	18	6.11	98	44	22	24	5.38	87
Left pmFC/dlPFC						−20	−2	60	5.65	58
Right pmFC/dlPFC						22	20	42	4.70	90
Right precentral gyrus						48	2	34	5.52	123
Right precentral gyrus/SMA						12	−14	74	5.44	223
Left thalamus						−14	−18	−2	4.52	58
Right superior temporal gyrus	36	−60	28	5.68	43					
Right fusiform gyrus						42	−48	−14	4.51	58
Left inferior temporal gyrus/fusiform gyrus						−44	−56	−10	10.06[a]	185
Right precuneus						18	−48	38	6.50[b]	302
Right temporo-parietal junction						58	−60	26	4.29	31
Left inferior parietal lobe						−42	−46	34	4.92	56
Right inferior parietal lobe	58	−48	40	4.09	55					
Right superior parietal lobe						38	−54	56	4.13	48
Left cerebellum						−14	−68	−38	5.23	195

Coordinates in the MNI (Montreal Neurological Institute) stereotactic space of the peak voxel within each cluster as indicated by the highest T-score. The reported clusters survive a whole-brain threshold of $p < 0.001$ and are significant after correction for multiple comparisons according to the Slotnick test statistic (cluster size > 30). Regions denoted by [a] or [b] are also significant after FDR correction at peak [a] or cluster [b] level (SPM8, $p < 0.05$). vlPFC = ventrolateral prefrontal cortex; dlPFC = dorsolateral prefrontal cortex. The contrasts with cognitive flexibility covariate did not show any significant activation.

conditions. Even though all Dutch questions had the same "Als … [antentedent], dan … [consequent]." structure, the "als" could have had two meanings in the false belief and basic conditional conditions. The questions in these conditions were formulated in the indicative present mood (e.g. "If A is B, what is C?"; see Appendix A) and thus this "als" could have been interpreted as a conditional expression ("if") or a temporal expression ("when"); we cannot rule out either. Regardless, in both these conditions the antecedent reflects a factual action (Johnson-Laird and Byrne, 2002). The counterfactual questions on the other hand were formulated in the subjunctive past mood (e.g., "If A had been B, what would C be?", see Appendix A), thus involving a conditional "if" which doesn't reflect a factual action but an action that is imagined, but never took place. Moreover, the use of additive versus subtractive (negations) counterfactuals might also have contributed to the complexity of the reasoning process. In addition, unlike counterfactuals, it might be the case that participants processed the false belief automatically during the vignette phase and kept this online to answer a possible false belief question afterwards (Apperly et al., 2006; Kovács et al., 2010). Future research should investigate this further.

We did not detect activation of the hippocampus frequently observed during autobiographical simulation. This is not surprising, because the lack of activity in many memory-related regions was predicted by the fact that the simulation process in all these three conditions did not require retrieving or recombining personal relevant information; instead information was presented in impersonal vignettes that were readily available to the participants (e.g., Addis and Schacter, 2012). Previous neurological studies confirm that not all false belief and counterfactual tasks necessarily require hippocampus involvement (Kulakova et al., 2013; Nieuwland, 2012; Rabin et al., 2012; Urrutia et al., 2012; Van Hoeck et al., 2013). In addition, the post-hoc analysis of Counterfactual Change of Location (i.e., altering the 'where') versus Counterfactual Change of Content (i.e., altering the 'what') supports the idea that not all types of counterfactual reasoning require the same amount of scene (re)construction.

Conclusion

The current research indicates that false belief, counterfactual and basic conditional reasoning engages regions supporting mentalizing, executive control, and mental imagery. It also provides support for the notion that false belief and counterfactual reasoning both entail generating an additional internal representation, depicting the other person's

Table 4
Covariate analysis: negative correlation.

	Cognitive flexibility									
	False belief > basic conditional					Counterfactual > basic conditional				
	x	*y*	*z*	*T*	*Voxels*	*x*	*y*	*z*	*T*	*Voxels*
dmPFC						14	60	26	4.29	32
						2	48	26	4.83	33
dlPFC	46	26	44	4.91	36	44	26	40	4.27	32
Right middle cingulate cortex						4	−4	30	4.94	69
Right posterior cingulate cortex						6	−36	22	4.56	50
Right paracentral lobule						6	−34	64	4.60	76
Right middle temporal gyrus						58	−66	−2	5.68	53
						34	−66	16	4.94	138
Precuneus						2	−50	38	5.28[b]	418
						−14	−54	36	4.69	33
Right temporo-parietal junction						62	−56	24	6.30	239
Left middle occipital gyrus/temporal gyrus						−36	−70	22	5.29	203

Coordinates in the MNI (Montreal Neurological Institute) stereotactic space of the peak voxel within each cluster as indicated by the highest T-score. The reported clusters survive a whole-brain threshold of $p < 0.001$ and are significant after correction for multiple comparisons according to the Slotnick test statistic (cluster size > 30). Regions denoted by [b] are also significant after FDR correction at cluster level (SPM8, $p < 0.05$). vlPFC = ventrolateral PFC; dlPFC = dorsolateral PFC; dmPFC = dorsomedial PFC. The contrasts with response inhibition covariate did not show any significant activation.

(false) belief state or the counterfactual state. In other words, thinking about the expectations of another person or about how an event could have turned out if something had changed, requires us to simulate a second representation, aside from representing the true state of affairs (basic conditional reasoning). This requires stronger engagement of neural regions supporting the integration of cognitive representations (left inferior parietal lobe & lateral prefrontal cortex) and mental imagery (occipital cortex). In contrast to false belief reasoning, the generation and integration of elements in a detailed counterfactual representation involves additional engagement of the dorsomedial and left dorsolateral prefrontal cortex, the cerebellum and the left temporal lobe. This suggests that counterfactual reasoning reflects a higher level, more complex and abstract cognitive process, recruiting a more extensive network of brain activation.

Acknowledgments

This research was funded by the Vrije Universiteit Brussel, projects GOA68 and SRP15.

Appendix A. Example vignette and questions

(Best possible translation from Dutch to English)

Example Vignette:

Jonas has a pizza delivered and takes his wallet out of his jacket. He places it on the windowsill and takes a shower. In the meanwhile Marion takes money out of Jonas' wallet to pay for the pizzas and places his wallet on the kitchen table.

False Belief question:

If/When Jonas steps out of the shower, then where does he expect his wallet to be?

Counterfactual question[2]:

If Jonas had laid out money for the pizzas, then where would Jonas' wallet be?

Basic conditional reasoning question:

If/When Jonas steps out of the shower, then where is his wallet?

References

Addis, D.R., Schacter, D.L., 2012. The hippocampus and imagining the future: where do we stand? Front. Hum. Neurosci. 5, 1–15. http://dx.doi.org/10.3389/fnhum.2011.00173.

Addis, D.R., Wong, A.T., Schacter, D.L., 2007. Remembering the past and imagining the future: common and distinct neural substrates during event construction and elaboration. Neuropsychologia 45 (7), 1363–1377. http://dx.doi.org/10.1016/j.neuropsychologia.2006.10.016.

Addis, D.R., Pan, L., Vu, M.-A., Laiser, N., Schacter, D.L., 2009. Constructive episodic simulation of the future and the past: distinct subsystems of a core brain network mediate imagining and remembering. Neuropsychologia 47 (11), 2222–2238. http://dx.doi.org/10.1016/j.neuropsychologia.2008.10.026.

Apperly, I.A., Riggs, K.J., Simpson, A., Chiavarino, C., Samson, D., 2006. Is belief reasoning automatic? Psychol. Sci. 17 (10), 841–844. http://dx.doi.org/10.1111/j.1467-9280.2006.01791.x.

Baayen, R.H., Davidson, D.J., Bates, D.M., 2008. Mixed-effects modeling with crossed random effects for subjects and items. J. Mem. Lang. 59 (4), 390–412. http://dx.doi.org/10.1016/j.jml.2007.12.005.

Baetens, K., Ma, N., Steen, J., Van Overwalle, F., 2013. Involvement of the mentalizing network in social and non-social high construal. Soc. Cogn. Affect. Neurosci. http://dx.doi.org/10.1093/scan/nst048.

Barbey, A.K., Colom, R., Solomon, J., Krueger, F., Forbes, C., Grafman, J., 2012. An integrative architecture for general intelligence and executive function revealed by lesion mapping. Brain 135, 1154–1164. http://dx.doi.org/10.1093/brain/aws021.

Baron-Cohen, S., Leslie, A.M., Frith, U., 1985. Does the autistic child have a "theory of mind"? Cognition 21 (1), 37–46 (Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9775957).

Beck, S.R., Riggs, K.J., Gorniak, S.L., 2009. Relating developments in children's counterfactual thinking and executive functions. Think. Reason. 15 (4), 337–354. http://dx.doi.org/10.1080/13546780903135904.

Botvinick, M.M., Cohen, J.D., Carter, C.S., 2004. Conflict monitoring and anterior cingulate cortex: an update. Trends Cogn. Sci. 8 (12), 539–546. http://dx.doi.org/10.1016/j.tics.2004.10.003.

Buckner, R.L., Carroll, D.C., 2007. Self-projection and the brain. Trends Cogn. Sci. 11 (2), 49–57. http://dx.doi.org/10.1016/j.tics.2006.11.004.

Cabeza, R., Ciaramelli, E., Moscovitch, M., 2012. Cognitive contributions of the ventral parietal cortex: an integrative theoretical account. Trends Cogn. Sci. 16 (6), 338–352. http://dx.doi.org/10.1016/j.tics.2012.04.008.

Carlson, S.M., Moses, L.J., Claxton, L.J., 2004. Individual differences in executive functioning and theory of mind: an investigation of inhibitory control and planning ability. J. Exp. Child Psychol. 87 (4), 299–319. http://dx.doi.org/10.1016/j.jecp.2004.01.002.

Carrington, S.J., Bailey, A.J., 2009. Are there theory of mind regions in the brain? A review of the neuroimaging literature. Hum. Brain Mapp. 30 (8), 2313–2335. http://dx.doi.org/10.1002/hbm.20671.

Chein, J.M., Schneider, W., 2012. The brain's learning and control architecture. Curr. Dir. Psychol. Sci. 21 (2), 78–84. http://dx.doi.org/10.1177/0963721411434977.

D'Argembeau, A., Cassol, H., Phillips, C., Balteau, E., Salmon, E., Van der Linden, M., 2013. Brains creating stories of selves: the neural basis of autobiographical reasoning. Soc. Cogn. Affect. Neurosci. http://dx.doi.org/10.1093/scan/nst028.

De Brigard, F., Addis, D.R., Ford, J.H., Schacter, D.L., Giovanello, K.S., 2013. Remembering what could have happened: neural correlates of episodic counterfactual thinking. Neuropsychologia 51 (12), 2401–2414. http://dx.doi.org/10.1016/j.neuropsychologia.2013.01.015.

Delis, D.C., Kaplan, E., Kramer, J.H., 2006. Test review. Appl. Neuropsychol. 13 (4), 275–279.

Dixon, P., 2008. Models of accuracy in repeated-measures designs. J. Mem. Lang. 59 (4), 447–456. http://dx.doi.org/10.1016/j.jml.2007.11.004.

Dodds, C.M., Morein-Zamir, S., Robbins, T.W., 2011. Dissociating inhibition, attention, and response control in the frontoparietal network using functional magnetic resonance imaging. Cereb. Cortex 21 (5), 1155–1165. http://dx.doi.org/10.1093/cercor/bhq187.

Drayton, S., Turley-Ames, K.J., Guajardo, N.R., 2011. Counterfactual thinking and false belief: the role of executive function. J. Exp. Child Psychol. 108 (3), 532–548. http://dx.doi.org/10.1016/j.jecp.2010.09.007.

Ferguson, H.J., 2012. Eye movements reveal rapid concurrent access to factual and counterfactual interpretations of the world. Q. J. Exp. Psychol. 65 (5), 939–961. http://dx.doi.org/10.1080/17470218.2011.637632.

Frith, C.D., Frith, U., 2010. Mechanisms of social cognition. Annu. Rev. Psychol. http://dx.doi.org/10.1146/annurev-psych-120710-100449.

Guajardo, N.R., Parker, J., Turley-Ames, K., 2009. Associations among false belief understanding, counterfactual reasoning, and executive function. Br. J. Dev. Psychol. 27 (3), 681–702. http://dx.doi.org/10.1348/026151008X357886.

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. http://dx.doi.org/10.1016/j.neuroimage.2012.03.012.

Hassabis, D., Maguire, E.A., 2007. Deconstructing episodic memory with construction. Trends Cogn. Sci. 11 (7), 299–306. http://dx.doi.org/10.1016/j.tics.2007.05.001.

Hassabis, D., Maguire, E.A., 2009. The construction system of the brain. Philos. Trans. R. Soc. Lond. B Biol. Sci. 364 (1521), 1263–1271. http://dx.doi.org/10.1098/rstb.2008.0296.

Holland, A.C., Addis, D.R., Kensinger, E.A., 2011. The neural correlates of specific versus general autobiographical memory construction and elaboration. Neuropsychologia 49 (12), 3164–3177. http://dx.doi.org/10.1016/j.neuropsychologia.2011.07.015.

Hughes, C., Ensor, R., 2007. Executive function and theory of mind: predictive relations from ages 2 to 4. Dev. Psychol. 43 (6), 1447–1459. http://dx.doi.org/10.1037/0012-1649.43.6.1447.

Irish, M., Addis, D.R., Hodges, J.R., Piguet, O., 2012. Considering the role of semantic memory in episodic future thinking: evidence from semantic dementia. Brain 135, 2178–2191. http://dx.doi.org/10.1093/brain/aws119.

Jaeger, T.F., 2008. Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. J. Mem. Lang. 59 (4), 434–446. http://dx.doi.org/10.1016/j.jml.2007.11.007.

Johnson-Laird, P.N., Byrne, R.M.J., 2002. Conditionals: a theory of meaning, pragmatics, and inference. Psychol. Rev. 109 (4), 646–678. http://dx.doi.org/10.1037//0033-295X.109.4.646.

Knight, R.T., Grabowecky, M., 1995. Escape from linear time: prefrontal cortex and conscious experience. In: Gazzaniga, M.S. (Ed.), The Cognitive Neuroscience. MIT Press, Cambridge, MA, pp. 1357–1371.

Kovács, Á.M., Téglás, E., Endress, A.D., 2010. The social sense: susceptibility to others' beliefs in human infants and adults. Science 330 (6012), 1830–1834. http://dx.doi.org/10.1126/science.1190792.

Kubit, B., Jack, A.I., 2013. Rethinking the role of the rTPJ in attention and social cognition in light of the opposing domains hypothesis: findings from an ALE-based meta-analysis and resting-state functional connectivity. Front. Hum. Neurosci. 7, 1–18. http://dx.doi.org/10.3389/fnhum.2013.00323.

Kulakova, E., Aichhorn, M., Schurz, M., Kronbichler, M., Perner, J., 2013. Processing counterfactual and hypothetical conditionals: an fMRI investigation. NeuroImage 72, 265–271. http://dx.doi.org/10.1016/j.neuroimage.2013.01.060.

[2] This one particular example was also the only story in which the counterfactual question might have two possible correct answers. Although it was not explicitly formulated in the story participants might have assumed that if Jonas had laid out money for the pizzas he would have put his wallet back into his jacket instead of the windowsill. This possible confusion increased then the complexity of this question. The mean response latency (time needed to infer the answer in the question phase) of the counterfactual condition in this particular story indeed exceeds 2 (not 3) standard deviations from the overall mean response latency for the counterfactual condition. A closer inspection of the data informs us that this is due to a longer response latency of three subjects. The mean response latency of the counterfactual condition in the other stories did not exceed 2 standard deviations. We could neither detect in these stories the possibility of assuming two correct answers.

Lombardo, M.V., Chakrabarti, B., Bullmore, E.T., Baron-Cohen, S., 2011. Specialization of right temporo-parietal junction for mentalizing and its relation to social impairments in autism. NeuroImage 56 (3), 1832–1838. http://dx.doi.org/10.1016/j.neuroimage.2011.02.067.

Mar, R.A., 2011. The neural bases of social cognition and story comprehension. Annu. Rev. Psychol. 62, 103–134. http://dx.doi.org/10.1146/annurev-psych-120709-145406.

Mars, R.B., Neubert, F.-X., Noonan, M.P., Sallet, J., Toni, I., Rushworth, M.F.S., 2012. On the relationship between the "default mode network" and the "social brain". Front. Hum. Neurosci. 6 (June), 189. http://dx.doi.org/10.3389/fnhum.2012.00189.

Martinelli, P., Sperduti, M., Piolino, P., 2013. Neural substrates of the self-memory system: new insights from a meta-analysis. Hum. Brain Mapp. 34 (7), 1515–1529. http://dx.doi.org/10.1002/hbm.22008.

Miller, E.K., Cohen, J.D., 2001. An integrative theory of prefrontal cortex function. Annu. Rev. Neurosci. 24, 167–202. http://dx.doi.org/10.1146/annurev.neuro.24.1.167.

Müller, U., Miller, M.R., Michalczyk, K., Karapinka, A., 2007. False belief understanding: the influence of person, grammatical mood, counterfactual reasoning and working memory. Br. J. Dev. Psychol. 25 (4), 615–632. http://dx.doi.org/10.1348/026151007X182962.

Nelson, S.M., Cohen, A.L., Power, J.D., Wig, G.S., Miezin, F.M., Wheeler, M.E., Petersen, S.E., 2010. A parcellation scheme for human left lateral parietal cortex. Neuron 67 (1), 156–170. http://dx.doi.org/10.1016/j.neuron.2010.05.025.

Nieuwland, M.S., 2012. Establishing propositional truth-value in counterfactual and real-world contexts during sentence comprehension: differential sensitivity of the left and right inferior frontal gyri. NeuroImage 59 (4), 3433–3440. http://dx.doi.org/10.1016/j.neuroimage.2011.11.018.

Perner, J., Lang, B., Kloo, D., 2002. Theory of mind and self-control: more than a common problem of inhibition. Child Dev. 73 (3), 752–767 (Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12038549).

Perner, J., Sprung, M., Steinkogler, B., 2004. Counterfactual conditionals and false belief: a developmental dissociation. Cogn. Dev. 19 (2), 179–201. http://dx.doi.org/10.1016/j.cogdev.2003.12.001.

Pinel, P., Thirion, B., Meriaux, S., Jobert, A., Serres, J., Le Bihan, D., Dehaene, S., 2007. Fast reproducible identification and large-scale databasing of individual functional cognitive networks. BMC Neurosci. 8, 91. http://dx.doi.org/10.1186/1471-2202-8-91.

Rabin, J.S., Braverman, A., Gilboa, A., Stuss, D.T., Rosenbaum, R.S., 2012. Theory of mind development can withstand compromised episodic memory development. Neuropsychologia 50 (14), 3781–3785. http://dx.doi.org/10.1016/j.neuropsychologia.2012.10.016.

Rafetseder, E., Perner, J., 2010. Is reasoning from counterfactual antecedents evidence for counterfactual reasoning? Think. Reason. 16 (2), 131–155. http://dx.doi.org/10.1080/13546783.2010.488074.

Rafetseder, E., Cristi-Vargas, R., Perner, J., 2010. Counterfactual reasoning: developing a sense of "nearest possible world". Child Dev. 81 (1), 376–389.

Rafetseder, E., Schwitalla, M., Perner, J., 2013. Counterfactual reasoning: from childhood to adulthood. J. Exp. Child Psychol. 114 (3), 389–404. http://dx.doi.org/10.1016/j.jecp.2012.10.010.

Riggs, K.J., Robinson, E.J., 1998. Are errors in false belief tasks symptomatic of a broader difficulty with counterfactuality? Cogn. Dev. 13, 73–90.

Santamarı, C., Espino, O., Byrne, R.M.J., 2005. Counterfactual and semifactual conditionals prime alternative possibilities. 31 (5), 1149–1154. http://dx.doi.org/10.1037/0278-7393.31.5.1149.

Saxe, R.R., Powell, L.J., 2006. It's the thought that counts: specific brain regions for one component of theory of mind. Psychol. Sci. 17 (8), 692–699. http://dx.doi.org/10.1111/j.1467-9280.2006.01768.x.

Schacter, D.L., Addis, D.R., Hassabis, D., Martin, V.C., Spreng, R.N., Szpunar, K.K., 2012. The future of memory: remembering, imagining, and the brain. Neuron 76 (4), 677–694. http://dx.doi.org/10.1016/j.neuron.2012.11.001.

Schilbach, L., Bzdok, D., Timmermans, B., Fox, P.T., Laird, A.R., Vogeley, K., Eickhoff, S.B., 2012. Introspective minds: using ALE meta-analyses to study commonalities in the neural correlates of emotional processing, social & unconstrained cognition. PLoS One 7 (2), e30920. http://dx.doi.org/10.1371/journal.pone.0030920.

Seeley, W.W., Menon, V., Schatzberg, A.F., Keller, J., Glover, G.H., Kenna, H., Greicius, M.D., 2007. Dissociable intrinsic connectivity networks for salience processing and executive control. J. Neurosci. 27 (9), 2349–2356. http://dx.doi.org/10.1523/JNEUROSCI.5587-06.2007.

Slotnick, S.D., Moo, L.R., Segal, J.B., Hart, J., 2003. Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. Cogn. Brain Res. 17 (1), 75–82. http://dx.doi.org/10.1016/S0926-6410(03)00082-X.

Spreng, R.N., 2012. The fallacy of a "task-negative" network. Front. Psychol. 3, 145. http://dx.doi.org/10.3389/fpsyg.2012.00145.

Spreng, R.N., Grady, C.L., 2010. Patterns of brain activity supporting autobiographical memory, prospection, and theory of mind, and their relationship to the default mode network. J. Cogn. Neurosci. 22 (6), 1112–1123. http://dx.doi.org/10.1162/jocn.2009.21282.

Spreng, R.N., Mar, R.A., 2012. I remember you: a role for memory in social cognition and the functional neuroanatomy of their interaction. Brain Res. 1428, 43–50. http://dx.doi.org/10.1016/j.brainres.2010.12.024.

Spreng, R.N., Mar, R.A., Kim, A.S.N., 2009. The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative meta-analysis. J. Cogn. Neurosci. 21 (3), 489–510.

Summerfield, J.J., Hassabis, D., Maguire, E.A., 2010. Differential engagement of brain regions within a "core" network during scene construction. Neuropsychologia 48 (5), 1501–1509. http://dx.doi.org/10.1016/j.neuropsychologia.2010.01.022.

Urrutia, M., Gennari, S.P., de Vega, M., 2012. Counterfactuals in action: an fMRI study of counterfactual sentences describing physical effort. Neuropsychologia 50 (14), 3663–3672. http://dx.doi.org/10.1016/j.neuropsychologia.2012.09.004.

Van den Noort, M., Bosch, P., Haverkort, M., Hugdahl, K., 2008. A standard computerized version of the reading span test in different languages. Eur. J. Psychol. Assess. 24 (1), 35–42. http://dx.doi.org/10.1027/1015-5759.24.1.35.

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. http://dx.doi.org/10.1016/j.neuroimage.2011.03.053.

Van Hoeck, N., Revlin, R., Dieussaert, K., Schaeken, W., 2012. The development of countefactual reasoning in belief revision. Psychol. Belg. 52 (4), 407–434. http://dx.doi.org/10.1093/scan/nss031.

Van Hoeck, N., Ma, N., Ampe, L., Baetens, K., Vandekerckhove, M., Van Overwalle, F., 2013. Counterfactual thinking: an fMRI study on changing the past for a better future. Soc. Cogn. Affect. Neurosci. 8 (5), 556–564. http://dx.doi.org/10.1093/scan/nss031.

Van Overwalle, F., 2009. Social cognition and the brain: a meta-analysis. Hum. Brain Mapp. 30 (3), 829–858. http://dx.doi.org/10.1002/hbm.20547.

Van Overwalle, F., Baetens, K., 2009. Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. NeuroImage 48, 564–584. http://dx.doi.org/10.1016/j.neuroimage.2009.06.009.

Van Overwalle, F., Baetens, K., Mariën, P., Vandekerckhove, M., 2013. Social cognition and the cerebellum: a meta-analysis of over 350 fMRI studies. NeuroImage. http://dx.doi.org/10.1016/j.neuroimage.2013.09.033.

Viard, A., Chételat, G., Lebreton, K., Desgranges, B., Landeau, B., de La Sayette, V., Piolino, P., 2011. Mental time travel into the past and the future in healthy aged adults: an fMRI study. Brain Cogn. 75 (1), 1–9. http://dx.doi.org/10.1016/j.bandc.2010.10.009.

Vincent, J.L., Kahn, I., Snyder, A.Z., Raichle, M.E., Buckner, R.L., 2008. Evidence for a frontoparietal control system revealed by intrinsic functional connectivity. J. Neurophysiol. 100 (6), 3328–3342. http://dx.doi.org/10.1152/jn.90355.2008.

Wellman, H.M., Cross, D., Watson, J., 2001. Meta-analysis of theory-of-mind development: the truth about false belief. Child Dev. 72 (3), 655–684 (Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11405571).

Whitman, J.C., Metzak, P.D., Lavigne, K.M., Woodward, T.S., 2013. Functional connectivity in a frontoparietal network involving the dorsal anterior cingulate cortex underlies decisions to accept a hypothesis. Neuropsychologia 51 (6), 1132–1141. http://dx.doi.org/10.1016/j.neuropsychologia.2013.02.016.

Young, L., Dodell-Feder, D., Saxe, R.R., 2010. What gets the attention of the temporo-parietal junction? An fMRI investigation of attention and theory of mind. Neuropsychologia 48 (9), 2658–2664. http://dx.doi.org/10.1016/j.neuropsychologia.2010.05.012.