

Counterfactual thinking: an fMRI study on changing the past for a better future

Nicole Van Hoeck, Ning Ma, Lisa Ampe, Kris Baetens, Marie Vandekerckhove, and Frank Van Overwalle

Department of Psychology, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussel, Belgium

Recent studies suggest that a brain network mainly associated with episodic memory has a more general function in imagining oneself in another time, place or perspective (e.g. episodic future thought, theory of mind, default mode). If this is true, counterfactual thinking (e.g. 'If I had left the office earlier, I wouldn't have missed my train.') should also activate this network. Present functional magnetic resonance imaging (fMRI) study explores the common and distinct neural activity of counterfactual and episodic thinking by directly comparing the imagining of upward counterfactuals (creating better outcomes for negative past events) with the re-experiencing of negative past events and the imagining of positive future events. Results confirm that episodic and counterfactual thinking share a common brain network, involving a core memory network (hippocampal area, temporal lobes, midline, and lateral parietal lobes) and prefrontal areas that might be related to mentalizing (medial prefrontal cortex) and performance monitoring (right prefrontal cortex). In contrast to episodic past and future thinking, counterfactual thinking recruits some of these areas more strongly and extensively, and additionally activates the bilateral inferior parietal lobe and posterior medial frontal cortex. We discuss these findings in view of recent fMRI evidence on the working of episodic memory and theory of mind.

Keywords: fMRI; counterfactual thinking; episodic memory

INTRODUCTION

'Without considering alternatives to reality, we must accept the past as having been inevitable and must believe that the future will be no different from the past. The generation of counterfactuals gives us the flexibility in thinking about possible futures and prepares us better for those futures' (Johnson and Sherman, 1990, p. 150).

Counterfactuals are thoughts about how (past) reality could have been different (e.g. 'If only I had stopped drinking earlier last night, I would not have felt this sick.'). Counterfactual thinking takes place in many social, emotional, and other aspects of our daily lives (e.g. Lebow, 2000; Bunzl, 2004; Roese and Morrison, 2009; Roese *et al.*, 2009; Yoon and Vargas, 2010; Rafetseder *et al.*, 2010). As the above citation illustrates, counterfactual thinking is closely related to episodic past and future thinking, which involve mentally traveling back or forth within a given context of time and space (e.g. Tulving, 2002; Szpunar and Mcdermott, 2008; Vandekerckhove, 2008). Recent studies suggest that episodic past and future thinking share common cognitive and neural processes (Buckner and Carroll, 2007; Spreng *et al.*, 2008). However, up till now, little is known about the processes common to counterfactual and episodic thinking, and this is especially true on the neurological level. The aim of this functional magnetic resonance imaging (fMRI) study is to fill this void, and to explore the common and unique neural processes underlying counterfactual thinking.

Counterfactual thinking

Counterfactual events are no mere fantasies; they could have actually replaced reality. Counterfactuals are activated when a problem is detected (e.g. Epstude and Roese, 2008; Barbey *et al.*, 2009). They generally take the form of conditionals ('If... then...') in which an alternative outcome to reality is achieved by a different (*additive*) or

absent (*subtractive*) action (Kahneman and Miller, 1986). These alternative events are better (*upward*) or worse (*downward*) than reality, which in turn will influence how we appraise and perceive the actual event (e.g. regret, envy, and relief) and will regulate our (future) behavior. Upward counterfactuals generate feelings of regret and lead to stronger behavioral intentions (e.g. 'If only I had checked the food in the oven more frequently, I would have noticed it started to burn. Next time, I will check the food more often.'). Whereas downward counterfactuals have mostly an affective regulation function (e.g. feeling relieved because it could have been worse; Markman *et al.*, 1993; Roese, 1994; Markman, *et al.*, 2008; Smallman and Roese, 2009; Epstude and Roese, 2011). That counterfactual thinking is closely related to problem solving abilities is supported by data suggesting that patients with impoverished decision-making and problem-solving skills, like frontal lobe patients, schizophrenic patients and patients with Parkinson disease, show counterfactual reasoning impairments (Hooker *et al.*, 2000; Mcnamara *et al.*, 2003; Camille *et al.*, 2004; Beldarrain *et al.*, 2005).

An episodic memory network

It is a reasonable assumption that counterfactuals share many cognitive and neurological processes with past and future memory. To construct a counterfactual, key elements from past experiences need to be remembered (like episodic past thinking) and, crucially, some elements need to be recombined so that a novel imagined scenario can be constructed (like episodic future thinking). An abundance of neurological studies during the last decade have demonstrated that episodic past and future thinking share several areas in the brain. Addis *et al.* (2009); Schacter and Addis, (2007) proposed a core memory brain network engaged during remembering and imaging of past and future events that includes the hippocampus, posterior cingulate/retrosplenial midline, inferior parietal lobule, lateral temporal cortices and the medial prefrontal cortex (Okuda *et al.*, 2003; Addis *et al.*, 2007; Buckner and Carroll, 2007; Addis and Schacter, 2008; Botzung *et al.*, 2008; D'Argembeau *et al.*, 2008; Schacter *et al.*, 2008; Szpunar *et al.*, 2009; Addis *et al.*, 2009; Andrews-Hanna *et al.*, 2010).

The hippocampus and its related para-hippocampal structures are primary memory structures responsible for the binding and encoding

Received 2 December 2011; Accepted 2 March 2012

Advance Access publication 7 March 2012

This research was supported by a GOA68 Grant from the Vrije Universiteit Brussel to the last two authors, and performed at GfMRI (Ghent Institute for Functional and Metabolic Imaging).

Correspondence should be addressed to Nicole Van Hoeck, Department of Psychology, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussel, Belgium. Email: nicole.van.hoeck@vub.ac.be

of multiple sensory-perceptual experiences from the past as well as for the vivid recovering of these experiences with rich detail (Addis *et al.*, 2007; Addis and Schacter, 2008; Moscovitch, 2008). It is generally assumed that the posterior cingulate/retrosplenial midline structures (including the precuneus and posterior cingulate cortex) are strongly engaged during autobiographic encoding and retrieval (Maddock *et al.*, 2001; Johnson *et al.*, 2002; Cavanna and Trimble, 2006; D'Argembeau *et al.*, 2008; Moran *et al.*, 2009). Semantic memory resides in the lateral temporal cortices (Vandekerckhove *et al.*, 2005; Addis *et al.*, 2007; Buckner and Carroll, 2007).

Mentalizing

Apart from the core memory network, there are other brain networks that show extensive overlap with this network, but appear to have distinct functionalities also outside memory. We suggest that these additional functionalities may play a crucial role in counterfactual reasoning.

Although showing striking commonalities with the core memory network, the posterior and anterior midline structures as well as the parietal cortices are also strongly engaged during mere social perception and mentalizing (theory of mind) processes (Amodio and Frith, 2006; Van Overwalle, 2009; Van Overwalle and Baetens, 2009) as well as during spontaneous mind-wandering, an undirected resting state which has been termed the default mode (Raichle *et al.*, 2001; Stawarczyk *et al.*, 2011). Therefore, several researchers have suggested that this set of regions, commonly activated by episodic memory, theory of mind, and default mode, is involved in constructing a mental scene based on memory (Hassabis and Maguire, 2007; Andrews-Hanna *et al.*, 2010) and projecting oneself mentally in a situation different from here and now (e.g. Tulving, 2002; Buckner and Carroll, 2007; Spreng *et al.*, 2008; Vandekerckhove, 2008; Vandekerckhove and Panksepp, 2009). Since counterfactual thinking is a form of self-projection and mental scene construction, we may expect the same regions be activated.

Based on a meta-analysis of over 200 fMRI studies, Van Overwalle (2009) proposed that mentalizing about temporary goals, intentions, desires, or beliefs of oneself or of another person activates the temporo-parietal junction (TPJ; part of the inferior parietal lobule), while inferring more long-term social aspects of self or other like traits, norms, and preferences recruits the medial prefrontal cortex (mPFC). We predict these mentalizing areas will be more engaged during counterfactual reasoning in comparison to episodic past thinking. In order to recombine elements of the past to undo earlier failures, counterfactuals involve thoughts on the causal implications of agents and actions that prevent a positive outcome and the active behavioral involvement of the self or other agents to remedy this situation (Barbey *et al.*, 2009).

Executive functions and emotionality

Counterfactuals inherently contain an aspect of conflict, as the counterfactual event contradicts the factual past event. We assume that focusing on the distinction between the counterfactual and factual outcome engages conflict detection (posterior medial frontal cortex; Botvinick *et al.*, 2004) and eliminating the undesirable elements of the past and constructing an alternative course of action strongly requires executive functionality involving performance monitoring, adaptive control, conflict resolution and physical causal reasoning (bilateral PFC; Ridderinkhof *et al.*, 2004; Neumann *et al.*, 2008; Van Overwalle, 2009; Barbey and Patterson, 2011).

Counterfactuals not only generate cognitive conflict, but also their emotional counterpart in the form of feelings of regret or relief. In making a decision, people often have expectations about the (un)pleasurable value of the likely outcome, and afterwards compare

these expectations with the real outcome. Neurological research demonstrates that these processes involve the orbito-frontal cortex (OFC; O'Doherty *et al.*, 2001; Ursu and Carter, 2005; Coricelli *et al.*, 2005, 2007; Fujiwara *et al.*, 2008; Ursu *et al.*, 2008; Chua *et al.*, 2009). Therefore, we predict that counterfactual thinking involves processing the value of event outcomes which engages the OFC (Barbey *et al.*, 2009).

Current task and specific predictions

To explore the role of counterfactual thinking, in the present experiment, participants had to re-experience personal negative past events and imagine for each a self-involved counterfactual event in which they or someone else would undo the negative outcome and turn it into a positive event. Note that these counterfactual past events were no pure hypotheticals which were not likely to have ever happened (Addis *et al.*, 2009), rather they could have easily replaced the factual past event. To control for imagination and positive valence of the counterfactual, they were also invited to imagine a personal positive event that would likely happen in the future. The positive outcome of both the counterfactual and future event also supports the notion that these events are not a recasting of the past, but a recombination of past details. As noted earlier, we make several predictions. Given that counterfactual thinking involves reasoning about the past to create a better outcome, our main prediction is that counterfactuals are supported by a set of brain areas common to all three conditions: the core memory network (e.g. Addis *et al.*, 2009) and the social mentalizing network (Van Overwalle, 2009). Aside from commonalities, we also expect that counterfactuals recruit distinct brain areas in comparison with past and future thinking (Barbey *et al.*, 2009). First, we hypothesize that counterfactual thinking will generate more activation in the conflict detection and resolution network (pmFC and lateral PFC). Secondly, we expect that counterfactuals will generate greater activation in the mentalizing network (TPJ and mPFC), because of its strong focus on alternative actions and possible related thought of the intentions and goals of oneself or others. Finally, we also expect counterfactuals to recruit areas involved in the processing the value of event outcomes (OFC).

MATERIALS AND METHODS

Participants

Fourteen healthy, right-handed adults (six males, eight females, $M_{\text{age}} = 21$ years, $s.d. = 22$ months, range = 19–25 years) with no prior history of neurological or psychiatric impairment participated in the study. One participant was removed because of considerable movements during scanning, large standard deviation of global mean intensity ($s.d. = 9.6$), and insufficient trials completed. 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€ for their participation.

Pre-scanning and stimulus material

One week before scanning, participants were asked to retrieve 30 distinct negative events that had taken place in their own past and to describe these events as concretely as possible. The events were not constrained to have happened in a certain timeframe (e.g. week, month, year, etc.). Events were, however, required to be temporally and contextually specific to guarantee a vivid recollection: past research has shown that events situated in a specific time and context recruit episodic memory, allowing the vivid recollection of these events in contrast to repeated events which elicit semantic memory (Wheeler

et al., 1997; Vandekerckhove and Panksepp, 2009). Participants also had to mention explicitly what the perceived cause of the negative outcome was for each event. Of the 30 events provided by the participants, we selected the 20 (plus one for the practice run) most temporally and contextually specific ones to develop personalized trials for the experimental scanning (e.g. a story about a car accident, caused by talking on a cell phone, during a holiday in France). From each selected event, three specific cue words were extracted (e.g. car–cell phone–holiday) to promote a vivid recollection of the event during scanning (Conway *et al.*, 2003).

We did not ask the participants to give a list of counterfactual or future plans before scanning, which otherwise could result in mere recall of these counterfactual or future plans rather than the online simulation of a new event during scanning.

Experimental task

Twenty past, 20 future and 20 counterfactual event trials were presented randomly across the entire scanning session. Each of the 20 cue combinations (three cues per trial) collected from the pre-scanning event descriptions were used for all types of events. Each trial started with a fixation cross that was presented for a mean duration of 4 s (jittered between 2 and 6 s). This was followed by an instruction slide that announced the task condition and it comprised three lines: (i) task ('negative past', 'positive future' or 'negative > positive' for the past, future and counterfactual conditions, respectively); (ii) three personal cue words (e.g. car–cell phone–holiday) which referred to a specific personal negative event from the participant's past; (iii) explicit task instructions that were adapted to the specific story (e.g. 'Imagine/recall what happened when you got in a car accident during your holiday in France', 'Imagine a future car trip that goes well', 'Imagine what would have happened if you had paid more attention to the road during the car trip in France' for the past, future and counterfactual conditions, respectively). The counterfactual instruction was always an upward and additive counterfactual-antecedent (i.e. a change of action that could have resulted into a different event-outcome; this new outcome had to be simulated and was therefore not provided). The antecedent was provided to ensure that this condition was not substantially more difficult than the past and future conditions; since increased difficulty of processing could by itself increase activation in areas that support this process.

Control task

As an explicit baseline task, 20 semantic memory trials were randomly interspersed through the scanning session. These tasks followed the same sequence as the experimental task. On the instruction slide, the first line described the task ('semantic retrieval'). The second line presented the same three cue words used in the experimental tasks. The third line stated the explicit instruction ('Imagine as much detail as possible about the semantic meaning of each word'). This control task measured semantic memory, and thus allows us to differentiate between processes related to semantic memory (control trials) and episodic memory (experimental trials).

Procedure

Immediately prior to scanning, participants had to read their description of the 20 past events (plus 1 practice event) that were selected. After reading the descriptions, it was explained to the participants that during scanning they had to recall these negative past events, imagine similar positive future events, or imagine counterfactual events (better outcomes than the past reality). The recalled or simulated events were not required to have happened in a specific time frame (e.g. week, month, year, etc.). Events were, however, required to be temporally

and contextually specific, i.e. participants were instructed to recall or imagine where and when a certain event did/would take place, which person was/would be present, and so on. Participants had to imagine the events as vividly as possible. In a semantic baseline condition, they had to recall as much detail as possible about the semantic meaning of the three cue words.

Participants started with a practice run to make them familiar with the different types of tasks and with the response box. The experiment itself consisted of 1 run, interrupted by a 30 s pause after a block of 20 trials. When participants were finished with reading the instructions, they were required to press a button to start the imagining phase (in case they did not press the button, this phase was skipped and the next event would start automatically after 20 s). In this second, critical phase of the task, the instruction were still visible, however, the background was gray indicating that during this phase, participants had to imagine the event as indicated in the instruction. To prevent mind-wandering, which would bias the results by activating the default system, this phase was self-paced (i.e. participants pressed a button when they finished imagining the event; Stawarczyk *et al.*, 2011). After an elapse of 25 s (without button pressed), the next phase started automatically. In a last phase, participants had to rate the vividness (amount of detail) of the imagined event (or semantic meaning of the cue words) by using one of the four buttons on the response box, which indicated either that they were able to imagine *little*, *few*, *many* or a *great many* details. This phase was included to make sure the events did not significantly differ in vividness.

After scanning, participants finished a cued recall task to check if the participants were able to perform the imagining tasks. They were provided with 70% of the trials (instructions and cues; 14 randomly chosen out of 20 trials per condition) and the task was to describe the event they imagined during scanning.

Imaging procedure

Images were collected with a 3T Magnetom Trio MRI scanner system (Siemens medical Systems, Erlangen, Germany), using an 8-channel radiofrequency head coil. Stimuli were presented in black text on a white background and 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.psnet.com/eprime; Psychology Software Tools) under Windows XP. Foam cushions were placed within the head coil to minimize head movements. We first collect a high-resolution T1-weighted structural scan (MP-RAGE) followed by one functional run of volume acquisitions (30 axial slices; 4 mm thick; 1 mm skip). Functional scanning used a gradient-echo echo-planar pulse sequence (TR = 2 s; TE = 33 ms; 3.5 × 3.5 × 4.0 mm in-plane resolution).

Image processing and statistical analysis

The fMRI data were preprocessed using SPM5 (Wellcome Department of Cognitive Neurology, London, UK). For each functional run, data were pre-processed 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 were 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 subjects had >10% outliers. No correlations between motion and experimental design or global signal and experimental design were identified.

Statistical analyses were performed using the general linear model of SPM8 (Wellcome Department of Cognitive Neurology, London, UK). The data were already preprocessed with SPM5 when we switched to SPM8 in order to apply peak-level correction instead of voxel-level correction (Chumbley *et al.*, 2010). The event-related design is modeled using a canonical hemodynamic response function and its temporal derivative. We only modeled the imagining phase, as an event with constant default duration (parameter set to 0). Six directions of motion parameters from the realignment step as well as outlier time points (defined by ART) were included as nuisance regressors in the first level analyses. In the second level analyses, we used a one-way within-subjects analysis of variance (ANOVA with one factor and four levels). We computed contrasts between each experimental condition and the control condition, as well as between the experimental conditions. A peak-based whole-brain statistical threshold of $P \leq 0.001$ was used for all comparisons. In a second general linear model, the trial-based vividness ratings were included as a parametric modulator for each event type, followed by a contrast analysis using a random-effect model to identify regions which were modulated by the vividness ratings in the three experimental conditions (past, future and counterfactual).

We report statistical contrasts 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 ($64 \times 64 \times 30$ 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 re-sampled 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 repeated measure ANOVA of the vividness ratings (Table 1) revealed that the mean vividness ratings differed significantly between conditions [$F(3,36) = 8.925$, $P < 0.001$]. Pair-wise comparisons revealed that the mean vividness ratings of the counterfactual condition were significantly lower ($P < 0.05$, Bonferroni corrected) than the past and the semantic condition, and marginally lower ($P = 0.06$) than the future condition. A non-parametric test, Friedman's one-way ANOVA, with median vividness ratings revealed the same results.

In addition, a one-way repeated measure ANOVA on the mean duration (Table 1) spend on imagining revealed no significant difference between conditions [$F(1.49, 17.88) = 1.99$; $P = 0.173$; Greenhouse–Geisser corrected].

Table 1 Mean vividness ratings and mean duration of each condition

	Past	Future	Counterfactual	Semantic
Vividness ratings, M (SD)	2.97 (0.43)	2.85 (0.43)	2.52 (0.24)	3.00 (0.40)
Duration, in seconds, M (SD)	18.98 (4.34)	18.27 (4.56)	18.68 (3.96)	20.29 (5.47)

The cued recall task confirmed that participants were able to recall all the 14 requested events for each condition, except in the counterfactual condition where the average recall was 13 out of 14. This suggests that the participants were able to image almost all requested events during scanning.

fMRI results

Regions common to past, future and counterfactual thinking

To examine the neural response patterns common to past, future and counterfactual thinking, a conjunction analysis (comprising three contrasts of each experimental condition > semantic baseline condition) was performed (Table 2 and Figure 1). Consistent with our prediction, this analysis confirmed that all three conditions activate the core memory network (hippocampus, posterior midline structures, parietal lobule and temporal lobe), and areas related to mentalizing (bilateral TPJ and mPFC) and executive control (right PFC). In addition, we also observed common activation of the bilateral cerebellum and bilateral pre-motor cortex (PMC).

Counterfactual thinking

To explore which regions are particularly involved in counterfactual thinking, we contrasted this condition against the past and future conditions (Table 2 and Figure 1). The Counterfactual > Past and Counter > Future contrasts confirmed our prediction that counterfactuals generate greater activation in conflict detection and resolution related areas (left PFC and pmFC), in the bilateral inferior parietal (IPL), right temporal pole (TP), left middle temporal gyrus (MTG) and left cerebellum. In contrast to predictions, we did not observe stronger activation in the TPJ or in the orbital PFC (related to processing feelings of regret and emotion regulatory processes). In addition, in the Counterfactual > Past contrast, we found the predicted activation in the mPFC (mentalizing about self and other) and also activation extending to the right hemisphere (in the MTG, PFC and cerebellum).

Future thinking

We also explored the regions that are particularly involved in future thinking (Table 2 and Figure 2). The Future > Past and Future > Counterfactual contrasts revealed common activation in regions involved in memory (precuneus, left MTG), executive control processes (left PFC), as well as in the medial OFC and right pre-central gyrus. Furthermore, the Future > Past contrast revealed additional activation in the ventral mPFC, left hippocampus, bilateral IPL, right PFC, and the right cerebellum. The Future > Counterfactual contrast revealed additional activation in the middle cingulate cortex, insula, as well as in the left precentral, the right post-central gyrus, and the left OFC.

Past thinking

In line with previous research (e.g. Addis *et al.*, 2007), the Past > Future and Past > Counterfactual contrast did not reveal greater activation in any region.

Parametric modulation by vividness ratings

To identify the regions related to the vividness of the simulated events, we conducted a parametric modulation analysis. This analysis did not reveal any significant activation in the Future or Counterfactual conditions. However, during past thinking, vividness was related to increased activation in the right cuneus (extending to cerebellum; coordinates 22 –58 20, $t = 4.42$).

Table 2 Whole brain analysis

	Conjunction Past + Future + Counter > Semantic				Counter > Past				Counter > Future				Future > Past				Future > Counter									
	x	y	z	Size	T	x	y	z	Size	T	x	y	z	Size	T	x	y	z	Size	T	x	y	z	Size	T	
Frontal																										
vmPFC	-10	54	16	1219 ^c	6.42 ^{ab}	10	54	20	49	4.13	-10	58	14	1879 ^c	5.43 ^b	-8	44	-14	480	5.11 ^b	-26	34	-14	63	4.45	
dmpFC	12	50	26	1219 ^c	6.06 ^{ab}	6	46	48	631	4.39	-8	42	-14	1879 ^c	5.55 ^b	-20	34	44	243	5.07 ^b	-20	34	44	243	5.07 ^b	
mOFC																										
ACC	6	36	2	42	4.21	-12	28	62	1031 ^c	5.21 ^b	12	26	58	78	4.42	-32	16	54	1940	6.13 ^{ab}	-32	28	26	235	4.62 ^b	
pmFC																										
L OFC																										
L PFC																										
L PFC																										
R PFC	22	26	40	70	4.20	-20	56	32	1031 ^c	6.54 ^{ab}	-22	56	32	52	4.91	20	34	48	383	5.25 ^b	42	-18	56	69	4.46	
L PMC	-40	10	52	65	4.42	-38	18	50	183	4.97 ^b																
R PMC	40	10	46	64	4.90	34	24	38	112	4.18																
L Pre-central gyrus																										
R Pre-central gyrus																										
R Post-central gyrus																										
Temporal																										
Lingual gyrus	14	-44	0	47	4.10																					
R insula																										
L hippocampus	-28	-8	-24	81 ^c	3.59																					
L Sub-gyral	-36	-6	-20	81 ^c	4.35																					
	-44	-16	-18	48	4.34																					
R temporal pole	44	12	-40	779 ^c	5.90 ^{ab}	40	12	-44	130	5.26 ^b	46	8	-46	105	4.72											
L MTG	-52	4	-26	108	5.83 ^a	-58	-8	-30	60	4.71																
L MTG	-56	-24	-16	134	4.18	-56	-30	-12	349	4.52 ^b	-60	-28	-6	66	4.77											
R MTG	46	2	-30	779 ^c	6.17 ^{ab}	62	-14	-22	62	4.27																
R MTG						68	-28	-4	35	4.22																
Parietal																										
Middle cingulate cortex	-2	-20	46	175	4.39 ^b																					
Precuneus	10	-50	36	2780 ^c	6.45																					
Posterior cingulate	4	-62	22	2780 ^c	6.50 ^{ab}																					
L TPJ extending to IPL	-58	-56	36	620	5.76 ^{ab}																					
R TPJ	46	-56	24	610	5.48 ^b																					
L IPL																										
R IPL																										
Cerebellum																										
L cerebellum	-8	-48	-48	61	4.72	-48	-76	-34	437	5.49 ^{ab}	-28	-74	-26	60	4.11											
R cerebellum	8	-46	-50	59	4.73	2	-56	-48	49	3.95																
	24	-78	-30	343	6.21 ^{ab}	36	-60	-34	48	3.84																
						28	-82	-38	462	5.81 ^{ab}																

Coordinates 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 seemingly having an equal size cluster, belong to one and the same cluster. Coordinates refer to the MNI (Montreal Neurological Institute) stereotaxic space. We consider x-coordinates within -15 to +15 mm range as medial. R = Right; L = Left; vmPFC = ventromedial prefrontal cortex; dmpFC = dorso-medial Prefrontal Cortex; mOFC = medial OrbitoFrontal Cortex; pmFC = posterior medial frontal cortex; PMC = premotor cortex; MTG = middle temporal gyrus; TPJ = temporo-parietal junction; IPL = inferior parietal lobule. The contrasts Past > Counterfactual and Past > Future did not show any significant activation.

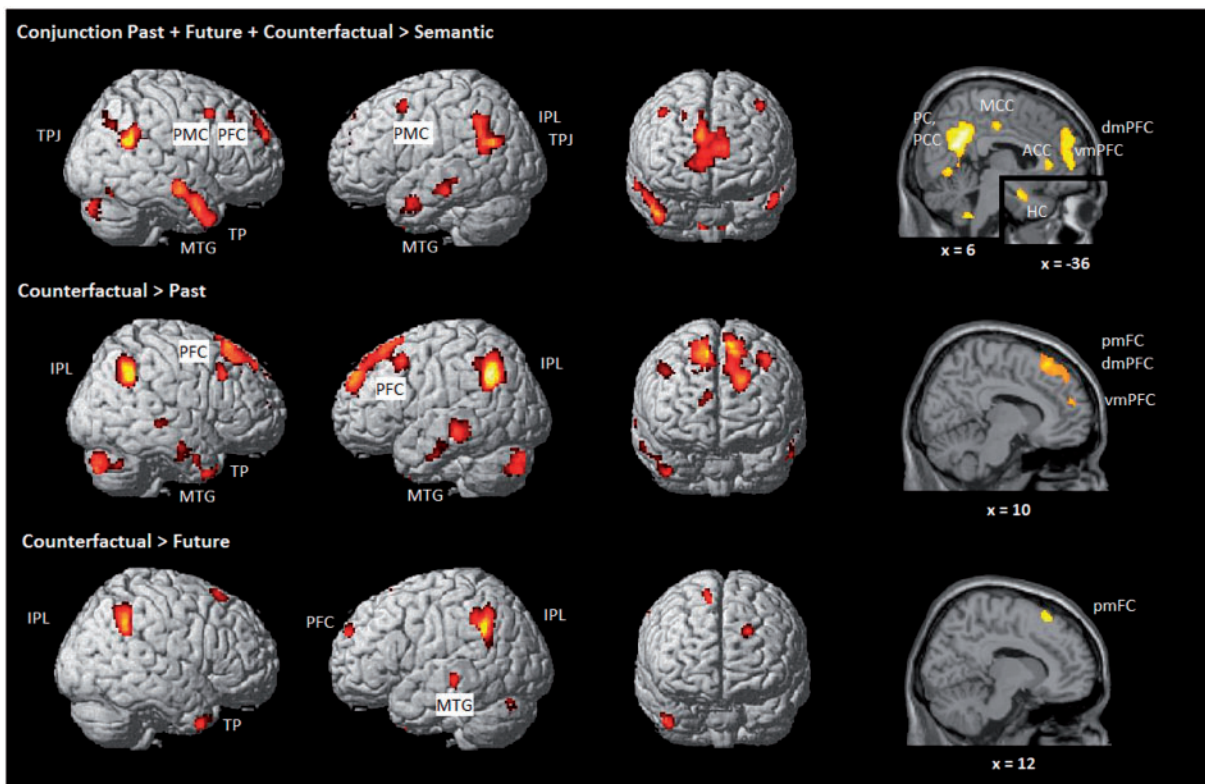


Fig. 1 Counterfactual, past, and future thinking. Conjunction and contrasts of counterfactual thinking with past and future. Whole-brain activation thresholded at $P < 0.001$ (corrected according to the Slotnick test statistic: cluster size > 30).

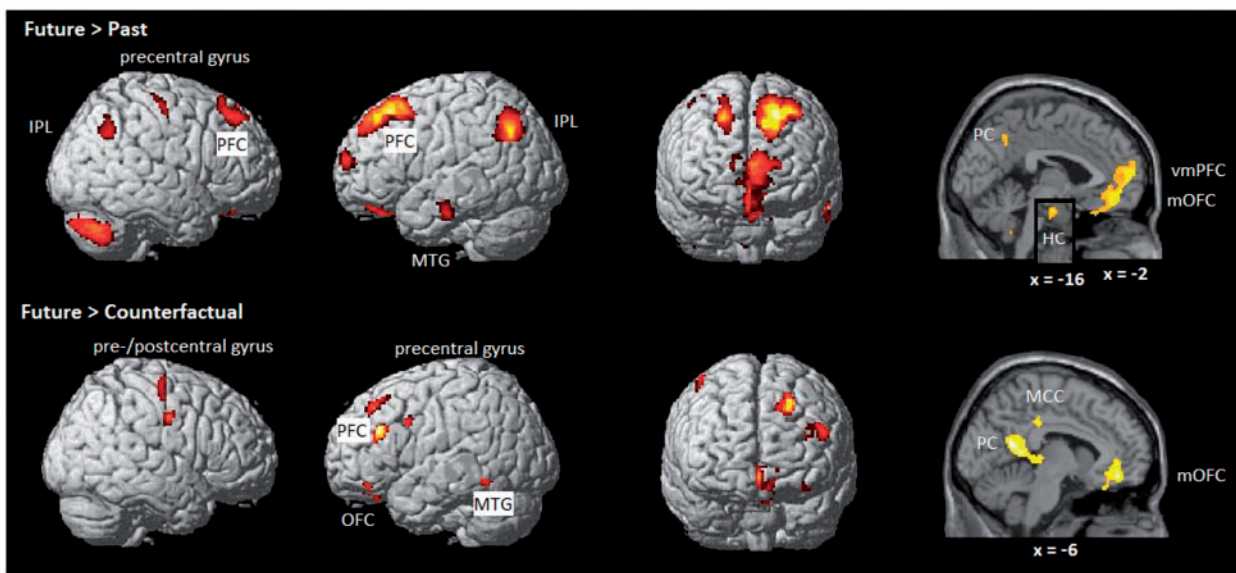


Fig. 2 Contrast of future thinking with counterfactual and past thinking. Whole-brain activation thresholded at $P < 0.001$ (corrected according to the Slotnick test statistic: cluster size > 30).

DISCUSSION

Reflecting on the past and imagining the future has become an important research topic in the last few years (e.g. Okuda *et al.*, 2003; Addis *et al.*, 2007; Hassabis and Maguire, 2007; Schacter *et al.*, 2007; Botzung *et al.*, 2008; Szpunar *et al.*, 2009; Vandekerckhove and Panksepp, 2009; D’Argembeau *et al.*, 2010). Recent studies demonstrate that a set of regions in the brain, previously linked with the

recollection of episodic past events, is also recruited by thinking about a personal future event (episodic future thinking), during default mode and theory of mind. This resulted in the suggestion that this set of regions is involved in constructing a mental scene based on memory (Andrews-Hanna *et al.*, 2010) and projecting oneself mentally in a situation different from here and now (e.g. Buckner and Carroll, 2007; Spreng *et al.*, 2008)

Since counterfactual thinking is a form of self-projection and mental scene construction, we reasoned that it should also activate these same regions. This study confirmed this. Episodic past, future and counterfactual thinking activate common regions typically related to memory processes (hippocampus, posterior midline structures, parietal lobule and temporal lobe), to mentalizing about intentions and goals of oneself or others (TPJ and mPFC), and to adaptive control (PFC). This is the first study that documented this overlap between counterfactual and episodic past and future thinking.

Another contribution of this study is that it confirmed our prediction that counterfactual thinking increases activation in areas related to conflict detection, action monitoring, adaptive control and causal reasoning (pmFC and PFC). During counterfactual thinking one has to consider two opposing events at the same time (the factual and counterfactual event), search for an appropriate way to alter a past action, and predict how this could affect other aspects of the event. All these mental processes require intensive controlled monitoring. Hence, counterfactuals requires more executive control than the mere re-experiencing or recasting of the past, and reflects an active process of recombining past elements into a novel event.

In line with our predictions, counterfactual thinking also activated the mPFC stronger than episodic past thinking. The mPFC is involved in mentalizing about long-term social aspects, like stable characteristics of self or other (e.g. traits or preferences), and long-term intentions and personal goals (Van Overwalle, 2009; D'Argembeau *et al.*, 2010). In contrast to our predictions, counterfactual thinking did not increase activation in the TPJ (social mentalizing about temporary intentions, goals and beliefs; Van Overwalle, 2009). Rather, a more dorsal activation of the bilateral inferior parietal lobule was observed. We speculate that this area might reflect the monitoring of the recollection of contextual details (Nelson *et al.*, 2010), or might be related to expectancy violation (O'Connor *et al.*, 2010) or the computation of action-reward contingency (Liljeholm *et al.*, 2011). All these functions might support the simulation of a counterfactual event.

Our hypothesis that the OFC would be recruited by counterfactual thinking (in particular by monitoring an event outcome) was not confirmed. This is might be the result of the specific implementation of counterfactuals in this experiment, where the participants knew in advance that the counterfactual involved a failed event and the focus was not on processing the value of an outcome, but on imagining the counterfactual event. Chandrasekhar *et al.* (2010) pointed out that when people are sure an event will turn out negative they will experience little to no regret, and indeed the medial OFC was not activated under such conditions. Likewise, Nicole *et al.* (2011) found no OFC activation when there was no anticipated risk of regret.

During counterfactual thinking, we also observed stronger activation in the temporal lobe (middle temporal gyrus and right temporal pole) and the cerebellum. These finding are, however, not surprising since these regions have been mentioned in many memory studies (e.g. Addis *et al.*, 2009). It has been suggested that the right temporal pole has a role of binding semantic and emotional information across domains (Vandekerckhove *et al.*, 2005; Willems *et al.*, 2010), while the cerebellum, is involved in episodic memory and emotion regulation (e.g. Strick *et al.*, 2009). The results suggest that these processes are perhaps more involved during counterfactual thinking.

Although we made no specific predictions concerning future thinking, it is interesting to see that our results confirm previous studies by demonstrating that imagining the future activates some regions relevant for memory and cognitive control more strongly than episodic past thinking (Addis *et al.*, 2007; Szpunar *et al.*, 2007; Abraham *et al.*, 2008; Schacter *et al.*, 2008). Also of interest is that episodic future thinking engages the pre-cuneus more strongly, but counterfactual thinking does not. A possible explanation is that—although future

and counterfactual thinking almost always involved self-related events—in this study counterfactual instructions invited to consider interventions by other persons besides the self. We leave it to future research whether this or another explanation is correct. Another interesting region more strongly activated by future thinking is the hippocampus. Addis *et al.* (2009) suggested that the hippocampus has an important role in the flexible recombination of past detail during imagining. However, counterfactual thinking did not recruit the hippocampus more than past thinking. This might suggest that recombining elements from the past is less flexible in counterfactual thinking than in future thinking. Indeed, counterfactual thoughts are grounded in past reality and are therefore perhaps restricted in content (i.e. limited recombination possibilities). During future thinking one might have more freedom in what one can imagine (i.e. more flexible recombination possibilities). Future studies should investigate if future thinking entails a more intensive or flexible recombination process than counterfactual thinking. Another explanation for the weaker activation of the hippocampus during counterfactual thinking is that participants in this study were given specific suggestions on how to resolve the undesirable past in the counterfactual condition, while in the future condition, the resolution was unspecified. This may have left ambiguity in how to fill in the imagination of the future events episodes, requiring more flexible reintegration of past details and executive effort. This might also explain why regions involved in executive control processes (PFC) are stronger activated during future thinking. We also found activations unique to future thinking in the precentral and post-central gyrus, and the orbital PFC.

CONCLUSION

Counterfactual thinking recruited several brain regions common with past and future thinking. These regions have been related in past research to memory, social mentalizing and executive functions, and might be generally involved in constructing a mental scene based on memory and/or projecting oneself mentally in a situation different from here and now. Of most interest, counterfactual thinking, in contrast to episodic past and future thinking, engages a network of brain areas related to conflict detection, action monitoring, adaptive control, and physical causality (pmFC and lateral PFC) as well as areas related to memory (right temporal pole, left middle temporal gyrus and left cerebellum). Additionally, counterfactual thinking strongly recruits the inferior parietal lobule. In contrast to past thinking only, counterfactual thinking elicits more mentalizing about self and other (mPFC).

REFERENCES

- Abraham, A., Schubotz, R.L., von Cramon, D.Y. (2008). Thinking about the future versus the past in personal and non-personal contexts. *Brain Research*, 1233, 106–19.
- Addis, D.R., Schacter, D.L. (2008). Constructive episodic simulation: temporal distance and detail of past and future events modulate hippocampal engagement. *Hippocampus*, 18, 227–37.
- 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), 2222–38.
- 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), 1363–77.
- Amodio, D.M., Frith, C.D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nature Reviews Neuroscience*, 7(4), 268–77.
- Andrews-Hanna, J.R., Reidler, J.S., Sepulcre, J., Poulin, R., Buckner, R.L. (2010). Functional-anatomic fractionation of the brain's default network. *Neuron*, 65(4), 550–62.
- Barbey, A.K., Patterson, R. (2011). Architecture of explanatory inference in the human prefrontal cortex. *Frontiers in Psychology*, 2, 1–9.
- Barbey, A.K., Krueger, F., Grafman, J. (2009). Structured event complexes in the medial prefrontal cortex support counterfactual representations for future planning. *Philosophical transactions of the Royal Society of London: Biological sciences*, 364, 1291–300.

- Beldarrain, M.G., Garcia-Monco, J.C., Astigarraga, E., Gonzalez, A., Grafman, J. (2005). Only spontaneous counterfactual thinking is impaired in patients with prefrontal cortex lesions. *Cognitive Brain Research*, 24(3), 723–6.
- Botvinick, M.M., Cohen, J.D., Carter, C.S. (2004). Conflict monitoring and anterior cingulate cortex: an update. *Trends in Cognitive Sciences*, 8(12), 539–46.
- Botzung, A., Denkova, E., Manning, L. (2008). Experiencing past and future personal events: functional neuroimaging evidence on the neural bases of mental time travel. *Brain and Cognition*, 66(2), 202–12.
- Buckner, R.L., Carroll, D.C. (2007). Self-projection and the brain. *Trends in Cognitive Sciences*, 11(2), 49–57.
- Bunzl, M. (2004). Counterfactual history: a user's guide. *The American Historical Review*, 109(3), 845–58.
- Camille, N., Coricelli, G., Sallet, J., Pradat-Diehl, P., Duhamel, J.-R., Sirigu, A. (2004). The involvement of the orbitofrontal cortex in the experience of regret. *Science*, 304, 1167–70.
- Cavanna, A.E., Trimble, M.R. (2006). The precuneus: a review of its functional anatomy and behavioural correlates. *Brain*, 129, 564–83.
- Chandrasekhar, P.V.S., Capra, C.M., Moore, S., Noussair, C., Berns, G.S. (2008). Neurobiological regret and rejoice functions for aversive outcomes. *NeuroImage*, 39(3), 1472–84.
- Chua, H.F., Gonzalez, R., Taylor, S.F., Welsh, R.C., Liberzon, I. (2009). Decision-related loss: regret and disappointment. *NeuroImage*, 47(4), 2031–40.
- Chumbley, J., Worsley, K., Flandin, G., Friston, K. (2010). Topological FDR for neuroimaging. *NeuroImage*, 49(4), 3057–64.
- Conway, M.A., Pleydell-Pearce, C.W., Whitecross, S.E., Sharpe, H. (2003). Neurophysiological correlates of memory for experienced and imagined events. *Neuropsychologia*, 41(3), 334–40.
- Coricelli, G., Critchley, H.D., Joffily, M., O'Doherty, J.P., Sirigu, A., Dolan, R.J. (2005). Regret and its avoidance: a neuroimaging study of choice behavior. *Nature Neuroscience*, 8(9), 1255–6.
- Coricelli, G., Dolan, R.J., Sirigu, A. (2007). Brain, emotion and decision making: the paradigmatic example of regret. *Trends in Cognitive Sciences*, 11(6), 258–265.
- D'Argembeau, A., Stawarczyk, D., Majerus, S., Collette, F., Van der Linden, M., Salmon, E. (2010). Modulation of medial prefrontal and inferior parietal cortices when thinking about past, present, and future selves. *Social Neuroscience*, 5(2), 187–200.
- D'Argembeau, A., Xue, G., Lu, Z.-L., Van der Linden, M., Bechara, A. (2008). Neural correlates of envisioning emotional events in the near and far future. *NeuroImage*, 40(1), 398–407.
- Epstude, K., Roese, N.J. (2008). The functional theory of counterfactual thinking. *Personality and Social Psychology Review*, 12(2), 168–92.
- Epstude, K., Roese, N.J. (2011). When goal pursuit fails. *Social Psychology*, 42(1), 19–27.
- Hassabis, D., Maguire, E.A. (2007). Deconstructing episodic memory with construction. *Trends in Cognitive Sciences*, 11(7), 299–306.
- Fujiwara, J., Tobler, P.N., Taira, N., Iijima, T., Tsutsui, K.-I. (2008). Personality-dependent dissociation of absolute and relative loss processing in orbitofrontal cortex. *European Journal of Neuroscience*, 27(6), 1547–52.
- Hooker, C., Roese, N.J., Park, S. (2000). Impoverished counterfactual thinking is associated with schizophrenia. *Psychiatry*, 63(4), 326–35.
- Johnson, S.C., Baxter, L.C., Wilder, L.S., Pipe, J.G., Heiserman, J.E., Prigatano, G.P. (2002). Neural correlates of self-reflection. *Brain*, 125, 1808–14.
- Johnson, M.K., Sherman, S.J. (1990). Constructing and reconstructing the past and the future in the present. In: Higgins, E.T., Sorrention, R.M., editors. *Handbook of Motivation and Cognition: Foundations of Social Behavior*, Vol. 2, New York: Guilford Press, pp. 482–526.
- Kahneman, D., Miller, D.T. (1986). Norm theory: comparing reality to its alternatives. *Psychological Review*, 93(2), 136–53.
- Lebow, R.N. (2000). What's so different about a counterfactual? *World Politics*, 52, 550–85.
- Liljeholm, M., Tricomi, E., O'Doherty, J.P., Balleine, B.W. (2011). Neural correlates of instrumental contingency learning: differential effects of action-reward conjunction and disjunction. *The Journal of Neuroscience*, 31(7), 2474–80.
- Maddock, R.J., Garrett, A.S., Buonocore, M.H. (2001). Remembering familiar people: the posterior cingulate cortex and autobiographical memory retrieval. *Neuroscience*, 104(3), 667–76.
- Markman, K.D., Gavanski, I., Sherman, S.J., McMullen, M.N. (1993). The mental simulation of better and worse possible worlds. *Journal of Experimental Social Psychology*, 29, 87–109.
- Markman, K.D., McMullen, M.N., Elizaga, R. (2008). Counterfactual thinking, persistence, and performance: A test of the reflection and evaluation model. *Journal of Experimental Social Psychology*, 44(2), 421–8.
- Mcnamara, P., Durso, R., Brown, A., Lynch, A. (2003). Counterfactual cognitive deficit in persons with Parkinson's disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, 74, 1065–70.
- Moran, J.M., Heatherton, T.F., Kelley, W.M. (2009). Modulation of cortical midline structures by implicit and explicit self-relevance evaluation. *Social Neuroscience*, 4(3), 197–211.
- Moscovitch, M. (2008). The hippocampus as a “stupid,” domain-specific module: Implications for theories of recent and remote memory, and of imagination. *Canadian Journal of Experimental Psychology*, 62(1), 62–79.
- Neumann, J., von Cramon, D., Lohmann, G. (2008). Model-based clustering of meta-analytic functional imaging data. *Human Brain Mapping*, 29, 177–92.
- Nelson, S.M., Cohen, A.L., Power, J.D., et al. (2010). A parcellation scheme for human left lateral parietal cortex. *Neuron*, 67, 157–70.
- Nicolle, A., Bach, D.R., Driver, J., Dolan, R.J. (2011). A role for the striatum in regret-related choice repetition. *Journal of Cognitive Neuroscience*, 23(4), 845–56.
- Okuda, J., Fujii, T., Ohtake, H., et al. (2003). Thinking of the future and past: the roles of the frontal pole and the medial temporal lobes. *NeuroImage*, 19(4), 1369–80.
- O'Connor, A.R., Han, S., Dobbins, I.G. (2010). The inferior parietal lobe and recognition memory: expectancy violation or successful retrieval? *The Journal of Neuroscience*, 30(8), 2924–2934.
- O'Doherty, J., Kringelbach, M.L., Rolls, E.T., Hornak, J., Andrews, C. (2001). Abstract reward and punishment representations in the human orbitofrontal cortex. *Nature Neuroscience*, 4(1), 95–102.
- Rafetseder, E., Cristi-Vargas, R., Perner, J. (2010). Counterfactual reasoning: developing a sense of “Nearest Possible World”. *Child Development*, 81(1), 376–89.
- Raichle, M.E., MacLeod, A.M., Snyder, A.Z., Powers, W.J., Gusnard, D.A., Shulman, G.L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences of the United States of America*, 98(2), 676–82.
- Ridderinkhof, K.R., Ullsperger, M., Crone, E.A., Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science*, 306, 443–7.
- Roese, N.J. (1994). The functional basis of counterfactual thinking. *Journal of Personality and Social Psychology*, 66(5), 805–18.
- Roese, N.J., Morrison, M. (2009). The psychology of counterfactual thinking. *Historical Social Research*, 34(2), 16–26.
- Roese, N.J., Epstude, K., Fessel, F., et al. (2009). Repetitive regret, depression, and anxiety: findings from a nationally representative survey. *Journal of Social and Clinical Psychology*, 28(6), 671–88.
- Schacter, D.L., Addis, D.R. (2007). The cognitive neuroscience of constructive memory: remembering the past and imagining the future. *Philosophical Transactions of the Royal Society of London: Biological sciences*, 362(1481), 773–86.
- Schacter, D.L., Addis, D.R., Buckner, R.L. (2007). Remembering the past to imagine the future: the prospective brain. *Nature Reviews. Neuroscience*, 8(9), 657–61.
- Schacter, D.L., Addis, D.R., Buckner, R.L. (2008). Episodic simulation of future events: concepts, data, and applications. *Annals of the New York Academy of Sciences*, 1124, 39–60.
- 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. *Cognitive Brain Research*, 17, 75–82.
- Smallman, R., Roese, N.J. (2009). Counterfactual thinking facilitates behavioral intentions. *Journal of Experimental Social Psychology*, 45(4), 845–52.
- Spreng, R.N., Mar, R.A., Kim, A.S.N. (2008). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative meta-analysis. *Journal of Cognitive Neuroscience*, 21(3), 489–510.
- Stawarczyk, D., Majerus, S., Maquet, P., D'Argembeau, A. (2011). Neural correlates of ongoing conscious experience: Both task-unrelatedness and stimulus-independence are related to default network activity. *PLoS ONE*, 6(2), 1–14.
- Strick, P.L., Dum, R.P., Fiez, J.A. (2009). Cerebellum and nonmotor function. *Annual Review of Neuroscience*, 32, 413–34.
- Szpunar, K.K., Chan, J.C.K., McDermott, K.B. (2009). Contextual processing in episodic future thought. *Cerebral Cortex*, 19(7), 1539–1548.
- Szpunar, K.K., McDermott, K.B. (2008). Episodic memory: an evolving concept. In: Sweat, D., Menzel, R., Eichenbaum, H., Roediger, H.L. III, editors. *Cognitive Psychology of Memory. Vol. [2] of Learning and Memory: A Comprehensive Reference*, Vol. 4 vols., Oxford: Elsevier, pp. 491–510.
- Szpunar, K.K., Watson, J.M., McDermott, K.B. (2007). Neural substrates of envisioning the future. *Proceedings of the National Academy of Sciences of the United States of America*, 104(2), 642–647.
- Tulving, E. (1989). Remembering and knowing the past. *American Scientist*, 77, 361–7.
- Tulving, E. (2002). Episodic memory: from mind to brain. *Annual Review of Psychology*, 53, 1–25.
- Ursu, S., Carter, C.S. (2005). Outcome representations, counterfactual comparisons and the human orbitofrontal cortex: implications for neuroimaging studies of decision-making. *Cognitive Brain Research*, 23, 51–60.
- Ursu, S., Clark, K.A., Stenger, V.A., Carter, C.S. (2008). Distinguishing expected negative outcomes from preparatory control in the human orbitofrontal cortex. *Brain Research*, 1227, 110–9.
- Van Overwalle, F. (2009). Social cognition and the brain: a meta-analysis. *Human Brain Mapping*, 30(3), 829–58.
- Van Overwalle, F., Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *NeuroImage*, 48, 564–84.
- Vandekerckhove, M.M.P. (2008). Memory, autoeocentric consciousness and the self: consciousness as a continuum of stages. *Self and Identity*, 8(1), 4–23.

- Vandekerckhove, M.M.P., Panksepp, J. (2009). The flow of anoetic to noetic and auto-noetic consciousness: a vision of unknowing (anoetic) and knowing (noetic) consciousness in the remembrance of things past and imagined futures. *Consciousness and Cognition*, 18(4), 1018–28.
- Vandekerckhove, M.M.P., Markowitsch, H.J., Mertens, M., Woermann, F.G. (2005). Bi-hemispheric engagement in the retrieval of autobiographical episodes. *Behavioural Neurology*, 16, 203–10.
- Wheeler, M.A., Stuss, D.T., Tulving, E. (1997). Toward a theory of episodic memory: the forntal lobes and auto-noetic consciousness. *Psychological Bulletin*, 121(3), 331–354.
- Willems, R.M., Clevis, K., Hagoort, P. (2010). Add a picture for suspense: neural correlates of the interaction between language and visual information in the perception of fear. *Social Cognitive and Affective Neuroscience*, 6(4), 404–16.
- Yoon, S., Vargas, P.T. (2010). Feeling happier when paying more: dysfunctional counter-factual thinking in consumer affect. *Psychology & Marketing*, 27(12), 1075–100.