NeuroImage 178 (2018) 332–345

Contents lists available at ScienceDirect

NeuroImage

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

What if? Neural activity underlying semantic and episodic counterfactual thinking

Natasha Parikh [a,b,*], Luka Ruzic [a,b], Gregory W. Stewart [a,b], R. Nathan Spreng [c,d], Felipe De Brigard [a,b,e,f]

[a] Department of Psychology and Neuroscience, Duke University, USA
[b] Center for Cognitive Neuroscience, Duke University, USA
[c] Laboratory of Brain and Cognition, Montreal Neurological Institute, Department of Neurology and Neurosurgery, McGill University, Canada
[d] Human Neuroscience Institute, Department of Human Development, Cornell University, USA
[e] Department of Philosophy, Duke University, USA
[f] Duke Institute for Brain Sciences, Duke University, USA

ARTICLE INFO

Keywords:
Counterfactual thinking
Episodic memory
Semantic memory
Plausibility
Partial least squares

ABSTRACT

Counterfactual thinking (CFT) is the process of mentally simulating alternative versions of known facts. In the past decade, cognitive neuroscientists have begun to uncover the neural underpinnings of CFT, particularly episodic CFT (eCFT), which activates regions in the default network (DN) also activated by episodic memory (eM) recall. However, the engagement of DN regions is different for distinct kinds of eCFT. More plausible counterfactuals and counterfactuals about oneself show stronger activity in DN regions compared to implausible and other- or object-focused counterfactuals. The current study sought to identify a source for this difference in DN activity. Specifically, self-focused counterfactuals may also be more plausible, suggesting that DN core regions are sensitive to the plausibility of a simulation. On the other hand, plausible and self-focused counterfactuals may involve more episodic information than implausible and other-focused counterfactuals, which would imply DN sensitivity to episodic information. In the current study, we compared episodic and semantic counterfactuals generated to be plausible or implausible against episodic and semantic memory reactivation using fMRI. Taking multivariate and univariate approaches, we found that the DN is engaged more during episodic simulations, including eM and all eCFT, than during semantic simulations. Semantic simulations engaged more inferior temporal and lateral occipital regions. The only region that showed strong plausibility effects was the hippocampus, which was significantly engaged for implausible CFT but not for plausible CFT, suggestive of binding more disparate information. Consequences of these findings for the cognitive neuroscience of mental simulation are discussed.

Introduction

The term *counterfactual*, coined by philosopher Nelson Goodman (1947), was introduced in reference to conditional statements whose antecedent is false, such as "If kangaroos had no tails, they would topple over" (Lewis, 1973, p. 1). Since, in fact, kangaroos do have tails, the antecedent of this statement asserts something that is contrary-to-fact or counterfactual. In the last four decades, a number of psychological studies have shed light on the cognitive processes that give rise to the thoughts expressed by counterfactual statements. However, until a few years ago, the scientific study of *counterfactual thinking* (CFT), as it is now known, was almost entirely confined to behavioral economics (Kahneman and Tversky, 1982), social psychology (Roese, 1997), and, more recently, cognitive science (Byrne, 2005).

Only in the past few years has cognitive neuroscience started to uncover the neural underpinnings of counterfactual thinking. Inspired by the striking commonalities found between *episodic memory* (eM) and episodic future thinking—i.e., thoughts about possible events that may occur to oneself in the future (Schacter et al., 2012; Spreng and Grady, 2010; Szpunar, 2010)—researchers began to compare these two kinds of mental simulations to *episodic counterfactual thinking* (eCFT): thoughts about alternative ways one's past personal events could have occurred but did not (De Brigard et al., 2016; De Brigard, Rodriguez and Montañés, 2017; De Brigard, Spreng, Mitchell and Schacter, 2015; De Brigard and

* Corresponding author. LSRC Box #90999, Duke University, Durham, NC, 27708, USA.
E-mail address: natasha.parikh@duke.edu (N. Parikh).

https://doi.org/10.1016/j.neuroimage.2018.05.053
Received 25 January 2018; Received in revised form 4 May 2018; Accepted 22 May 2018
Available online 25 May 2018
1053-8119/Published by Elsevier Inc.

Giovanello, 2012; Özbek et al., 2016, 2017; Stanley et al., 2017). In a pioneer study (Addis et al., 2009), participants were asked to remember *actual past* events, to mentally recombine elements from their memories into *possible future* events, or to mentally recombine them into *possible past* events, while undergoing fMRI. When compared to baseline, the results revealed substantial overlap across these three kinds of mental simulations in core regions of the brain's default network (DN), a set of functionally connected neural structures including ventral medial prefrontal cortex (vmPFC), posterior cingulate cortex (PCC), inferior parietal lobule (IPL), lateral temporal cortex (LTC), dorsal medial prefrontal cortex (dmPFC), and the hippocampal formation in the medial temporal lobe (MTL; Andrews-Hanna et al., 2014; Benoit and Schacter, 2015; Buckner et al., 2008; Schacter et al., 2012).

Although related, the possible past condition in Addis et al.'s study was importantly different from the mental operation we normally associate with eCFT. By randomly recombining fragments from past events, participants may have been asked to imagine not only events that did not happen, but also events that could not have happened at all. Nevertheless, the hypothesis that core regions of the DN were preferentially engaged during eCFT was further supported by two studies (De Brigard, Addis, Ford, Schacter and Giovanello, 2013; Van Hoeck et al., 2013) directly asking participants to imagine alternative ways in which past personal events could have occurred while undergoing fMRI. Further findings have refined the nature of this initial observation. De Brigard et al. (2013) showed that the engagement of the DN was modulated by the perceived plausibility of eCFT, such that episodic counterfactual events that were perceived as more plausible-to-occur recruited core regions of the DN to a greater extent than those that were perceived as less plausible. In a subsequent study (De Brigard et al., 2015), participants were asked to engage in eCFT about themselves, other people, or objects while undergoing fMRI. The results showed little involvement of DN regions during counterfactual thinking involving objects but substantial involvement of DN regions during eCFT involving people—both self and others. However, the degree to which DN was recruited during people-based eCFT was modulated by how closely related participants were to the protagonist of the mental simulation. Specifically, person-based eCFT involving oneself as well as familiar people were more likely to engage core regions of the DN relative to person-based eCFT that involved unfamiliar people.

What could be the reason for the differential engagement of DN regions during eCFT? One possibility is that person-based counterfactual thoughts that involve oneself or familiar people are perceived as more plausible than those that involve either unfamiliar people or objects; this would be consistent with the aforementioned result suggesting greater involvement of core regions of the DN during plausible relative to implausible eCFT (De Brigard et al., 2013). Another possibility is that the mental simulations generated during counterfactual thoughts involving objects or unfamiliar people as well as implausible events recruit information and processes from semantic memory (sM) to a greater degree than they do so from eM (Addis and Schacter, 2012; Irish and Piolino, 2016). Partial support for this hypothesis comes from a recent neuroimaging study whereby counterfactual thoughts about factual or semantic statements—i.e., *semantic counterfactual thinking* (sCFT)—correlate with activation in the cuneus and caudate (Kulakova et al., 2013), rather than core regions of the DN, such as the vmPFC, the PCC, or the MTL. However, this study did not manipulate plausibility, and it only compared sCFT against non-counterfactual hypothetical statements, not against eCFT.

The purpose of the current study is to uncover the precise relationship between the neural regions engaged during plausible and implausible eCFT and sCFT. Specifically, set up as a 2×3 within-subject design, the current study allows us to compare neural activity associated with eM and sM when they are reactivated during memory recollection, plausible counterfactual generation, or implausible counterfactual generation, to answer three questions. First, we seek to determine commonalities and differences in brain activity associated with both plausible and implausible eCFT and sCFT as compared to eM and sM. Second, we directly compare both plausible and implausible eCFT and sCFT to test whether there is differential engagement of DN activity as a function of plausibility, episodicity, or both. Finally, given previous results suggesting changes in hippocampal activity as a function of perceived plausibility and amount of episodic detail during episodic future thinking (Addis and Schacter, 2008; Spreng et al., 2015; Weiler et al., 2010), we seek to investigate changes in hippocampal activity as a function of both plausibility and episodicity during eCFT and sCFT.

To these ends, three strategies to analyze the fMRI data were planned. First, a data-driven mean-centered spatiotemporal partial least squares analysis (PLS; McIntosh et al., 2004) was employed to explore reliable commonalities and differences in brain activity during the mental simulation of the six conditions of interest. We predicted that this analysis would yield three distinct patterns of brain activity: 1) segregating memory (both episodic and semantic) from counterfactual thinking (both eCFT and sCFT); 2) segregating semantic simulations from episodic simulations; and 3) separating plausible from implausible simulations. Next, to directly explore differences in brain activity for plausible and implausible eCFT and sCFT, we conducted a mean-centered PLS analysis restricted to the four counterfactual thinking conditions. We hypothesized that, if perceived plausibility is the main factor driving the engagement of DN activation during counterfactual simulations, we should see greater activity in core regions of DN during plausible relative to implausible counterfactual thinking, whether episodic or semantic. However, if it is episodicity rather than perceived plausibility that is driving the DN activity, we should expect to see greater activation of core regions of the DN during plausible and implausible eCFT relative to both plausible and implausible sCFT. Finally, a region-of-interest (ROI) analysis was planned to explore differences in hippocampal activation as a function of perceived plausibility and episodicity during both eCFT and sCFT. This ROI analysis would allow us to explore whether increased activity in the hippocampus responds to the episodic nature of the CFT simulation, its perceived plausibility, or a combination of the two.

Materials and methods

Participants

Twenty-seven individuals completed both sessions of our study (Age $M = 22.15 \pm 2.98$, 14 females). Participants were right-handed, native English speakers, and had no history of psychiatric or neurological disorders. Participants were recruited using flyers posted on campus and on a Duke recruitment website, and they gave written consent according to the requirements of the Duke University Health System Institutional Review Board. Participants received monetary compensation for their time.

Pre-scan stimulus preparation

Prior to the study, we generated a list of 180 possible life events to cue participants' autobiographical memories. The list of these possible autobiographical episodes included common events that were specific to a spatiotemporal location and had clear outcomes that would have followed directly from the event or action (e.g. "You were in a snowball fight." or "You were caught in the rain without an umbrella."). These 180 events were also classified by two independent coders (NP and GWS) on whether they could be easily manipulated in an implausible manner (see Supplementary Information). For example, "You broke something that did not belong to you" was rated as easily manipulated, while "You accidentally pushed on a pull door" was classified as difficult to make into an implausible eCFT that was still credible. Additionally, 80 true and 80 false semantic statements were created (e.g., "Combining red paint and blue paint will create purple paint." and "Rhode Island is one of the largest states in the U.S."). To minimize the chances of including semantic statements that would cue episodic autobiographical

recollections, we avoided semantic statements referring to social or person-based facts. The resultant 160 statements were normed on Amazon's Mechanical Turk (MTurk; Buhrmester, Kwang and Gosling, 2011) according to four dimensions: familiarity, plausibility that the false statement could have been true, amount of visual detail, and degree of social content. After feedback from 60 participants, the 63 most familiar, least social, and least visually detailed statements were chosen for the semantic trials in the main experiment (see Supplementary Information).

Experimental sessions

This study consisted of two sessions. In Session 1, participants were asked to fill out a "possible experiences" questionnaire with the 180 generated autobiographical events. For each event, participants were asked to recall whether, to the best of their knowledge, they had experienced the displayed event in the past ten years and, if they had experience it, how sure they were of its occurrence, from 1 (very unsure) to 7 (very sure). If they indicated that they had experienced the given event, they were asked to provide a brief description of their specific memory. From these, 63 events with the highest ratings of confidence of occurrence were chosen for the episodic trials in the second session.

A week later, participants came back for the second, experimental session, which took place in an MRI scanner. This experimental session was structured as a 2 (Stimuli type: episodic, semantic) × 3 (Simulation type: memory, plausible counterfactual, implausible counterfactual) within-subject experimental design (Fig. 1). Thus, three of the conditions involved episodic autobiographical stimuli and three involved semantic stimuli. Within each memory type, the three simulation types were the same: a memory condition, a plausible counterfactual condition, and an implausible counterfactual condition. For each condition, participants saw a screen naming the condition type, the instruction for that condition, and a stimulus statement. In the Episodic Memory (eM) condition, participants saw the terms "Memory," "Remember?" and one of the 63 episodic statements they had identified as referring to an actual past personal event during Session 1 (e.g. "You voted in the last election"). They were asked to recall the memory, press a button once they had a clear image of it in their mind, and then continue to elaborate on it for the remainder of the 12-s simulation window. In the Episodic Plausible (Plausible eCFT) and Implausible (Implausible eCFT) Counterfactual conditions, participants saw the cues "Imagine Plausible" or "Imagine Implausible," respectively, followed by the instruction "What if?" and the negation of one of the episodic statements they had identified as an actual autobiographical event in Session 1. For example, if they had identified "You got caught in the rain" as an actual autobiographical event, they would see "You did not get caught in the rain" in the counterfactual condition. Participants were asked to imagine either a plausible or an implausible counterfactual scenario in which the outcome of their specific memory—i.e., getting caught in the rain—was different. As in the remember condition, participants were asked to press a button once they constructed the counterfactual simulation, and elaborated on it for the remainder of the 12 s.

The semantic conditions paralleled the episodic conditions. The Semantic Memory (sM) condition displayed one of the 63 chosen semantic statements with the instructions "Fact" and "True?" presented concurrently. The participant had to first answer if the statement was true by pressing a button once they recognized whether or not it stated a fact, and then were asked to mentally elaborate over that fact for the remainder of the 12-s simulation window. If the statement was false, they were asked

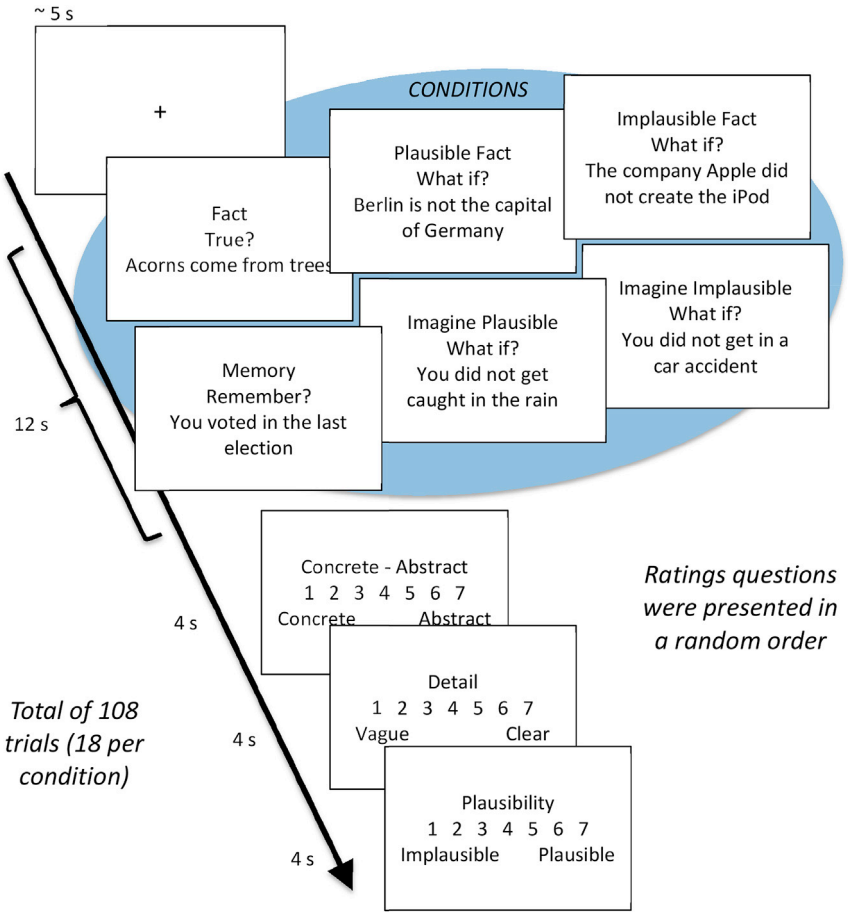

Fig. 1. Summary of a trial in experimental paradigm, Session 2. This session was conducted in an fMRI scanner, broken up into 6 runs of 18 trials each (3 of each condition in each run).

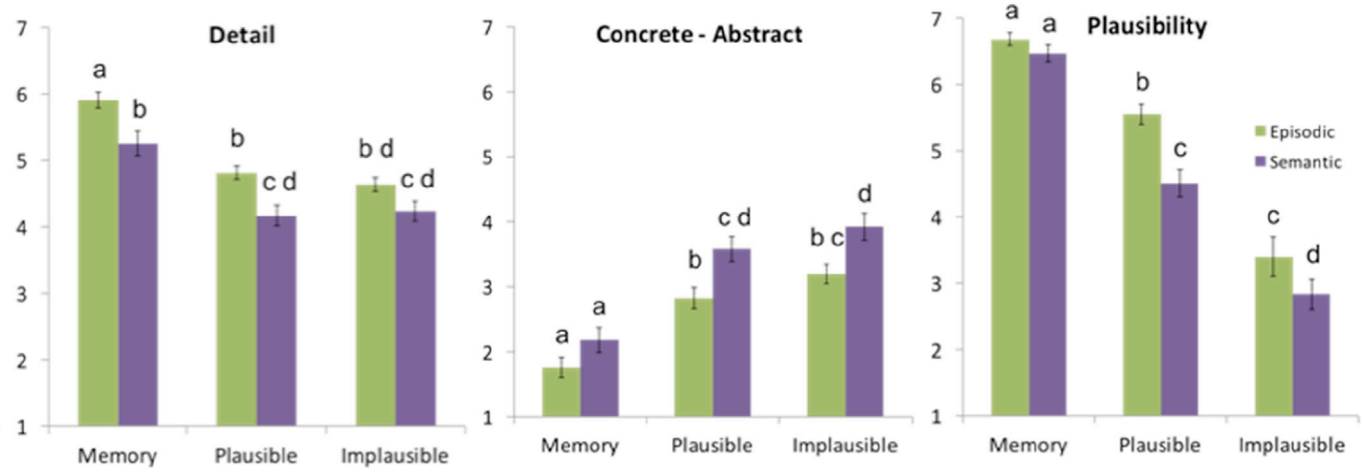

Fig. 2. Behavioral ratings of Level of Detail, Concreteness, and Plausibility. Scales range from 1 to 7. For each graph, unique letter labels represent significant differences between bars ($p < .05$, corrected). Error bars represent SEM.

to come up with the true version of the fact and mentally elaborate over that statement. For the Semantic Plausible (Plausible sCFT) and Implausible (Implausible sCFT) Counterfactual conditions, participants see "Plausible Fact" or "Implausible Fact," "What if?" and a true fact phrased negatively (e.g. "Washington D.C. is not the capital of the U.S."). Participants were instructed to imagine an alternative version of the fact that was either plausible or implausible, respectively. Once they created a counterfactual, they pressed a button and thought about the new version for the remainder of the 12 s.

Each trial began with one of these six simulation instruction screens. After the simulation, three rating questions followed in a randomized order (see Fig. 1). Participants were asked to rate their mental simulation on its level of detail (1 = vague to 7 = detailed), how concrete vs. abstract it was (1 = concrete, 7 = abstract), and how plausible it was (1 = implausible to 7 = plausible). Since the events in the episodic and semantic memory conditions were veridical, at least to the best of the participant's knowledge, they were asked to score their corresponding mental simulations as a 7 on the plausibility scale. Trials were separated by a fixation cross "+" that was jittered around 4 s. Each run contained 18 trials (3 of each condition, randomized across the block), and six runs were completed in the scanner. Participants completed one full practice run before entering the scanner. Each run was approximately 7 min long.

Post-scan interview

Immediately after the scanning session, participants completed a post-scan questionnaire about their counterfactual thoughts. Participants gave a brief description of each counterfactual they created in the scanner. To aid with this task, they saw a list of the stimulus and associated condition types from each trial to verify compliance.

Scanning parameters

The scanning session was completed in a 3T GE MR750 Scanner. The scan started with a localizer and a high-resolution T1-weighted structural scan (162 1 mm slices, TR = 8.16 ms, TE = 3.18 ms). These were followed by six runs of functional scans using a whole brain, gradient-echo, spiral sequence (TR = 2 s, TE = 30 ms, FOV = 240 mm, Matrix Size = 64 × 64, Flip Angle = 80°). Slices were acquired in an interleaved fashion (36 × 3.8 mm slices). Foam padding was put around the participants' heads inside the head coil to ensure minimal head movement. The experimental task was projected into the scanner and viewed by the participants from a mirror placed above the head coil. Stimuli were presented in black lettering on a white background using E-Prime software (Psychology Software Tools, Inc., Pittsburgh, PA). Responses were made on two 4-button MR-compatible response boxes held in each hand. An extra 8 s of fixation (4 TRs) were included at the beginning of each run that were dropped from analysis.

Data preprocessing and analyses

Ratings responses and response time (RT) for construction were analyzed for each condition using analyses of variance (ANOVAs), t-tests, and Pearson correlations. Participants who did not press a button to indicate a clear separation between the construction and elaboration periods of the mental simulation (n = 5) were excluded from RT analyses. Results were corrected for sphericity using Greenhouse-Geisser corrections where appropriate, multiple hypothesis testing used Holm's method, and effect sizes for t-tests were calculated using Cohen's d_z (Lakens, 2013). Functional fMRI data were preprocessed in SPM8 (Wellcome Department of Cognitive Neurology, London, UK) implemented in MATLAB (Mathworks, Natick, MA). Images were realigned, segmented using the SPM add-on VBM8 (Structural Brain Mapping Group, Jena, Germany), co-registered, re-sliced and normalized to the MNI template (2 × 2 × 2 mm), and smoothed with a 6-mm full-width at half maximum isotropic Gaussian kernel.

Preprocessed neuroimaging data were analyzed using a spatiotemporal partial least squares (PLS) analysis toolbox implemented in MATLAB. PLS is a multivariate approach that extracts whole brain activity correlated with task. A strength of the approach is that PLS allows for the simultaneous analysis of multiple conditions across all voxels at once, to assess both patterns of dissociation and convergence. In particular, we employ data-driven, mean-centered PLS analyses. This type of analysis returns a number of orthogonal components (Latent Variables, LVs) that explain correlations between the brain data and the experimental design, with each subsequent LV explaining less of the variance seen in the data. For each LV, conditions receive a "brain score," which is a whole brain composite activity value, that is used to compare conditions based upon overlapping or non-overlapping confidence intervals determined from a bootstrap analysis. These brain scores can be used to indicate the level with which each condition correlates with the brain regions identified by the LV. We ran two mean-centered PLS analyses, one that included all conditions, and an analysis without the two memory simulation conditions to look at interactions within the counterfactual simulation conditions only. Activity from the 14 s (i.e., 7 TRs) following the simulation instruction screen was captured from each trial and binned into appropriate conditions. Activity during these periods was normalized relative to the first TR of the trial. The significance of each LV as a whole was determined by permutation testing, using 500 permutations. In a second, independent step, the reliability of the saliences for the brain voxels

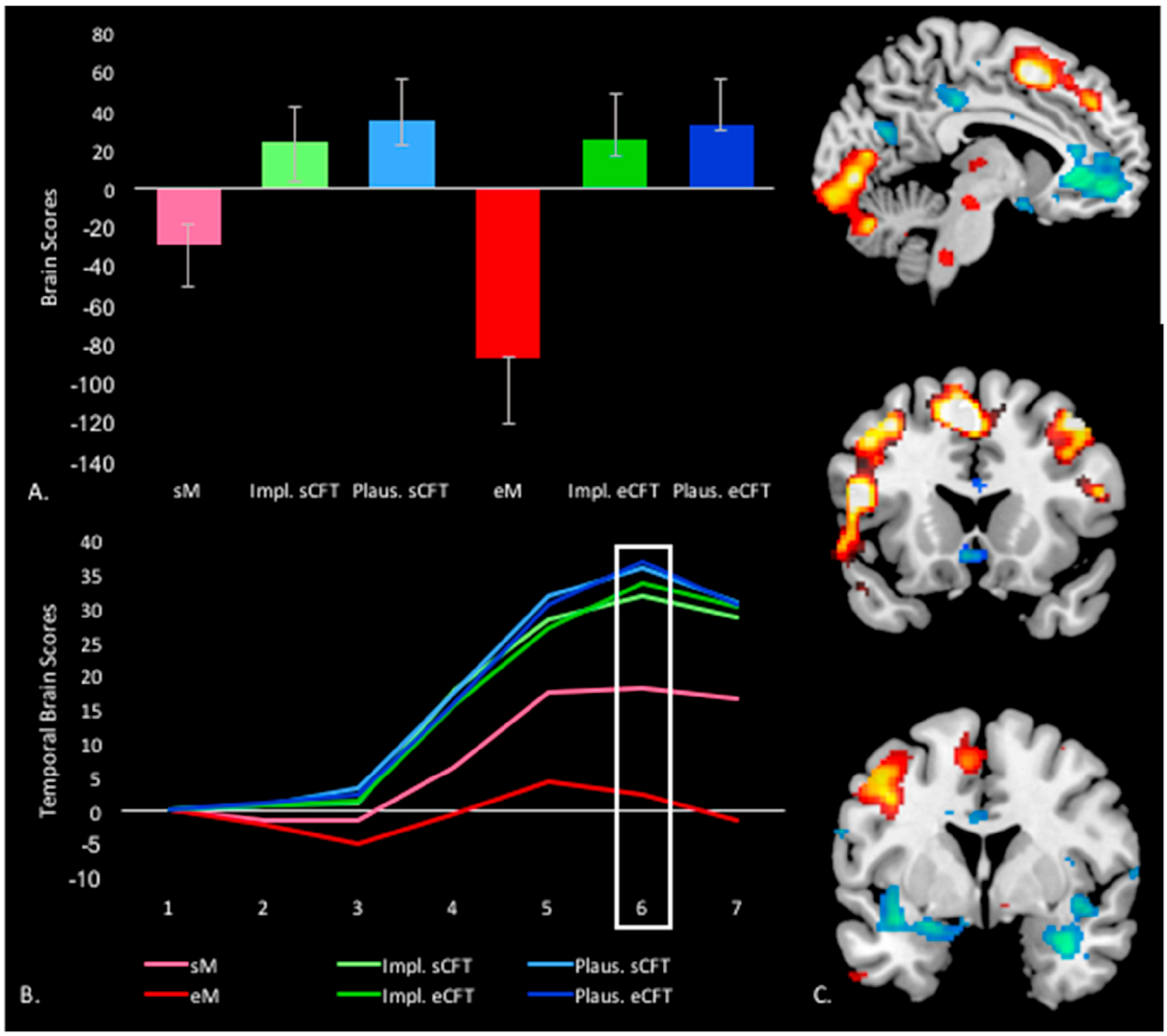

Fig. 3. Results from Latent Variable 1 (LV1) from the mean-centered PLS analysis. A) Plot of brain scores for each condition. Error bars represent confidence intervals from bootstrapping. B) Plot of temporal brain scores indicating weighed average of activation across all voxels in all participants during the length of the mental simulation. The white box represents the TR chosen for maximal separation between conditions. C) Regions with negative saliences in brain scores, depicted here in cool colors, were engaged during eM and sM, whereas regions with positive saliences, depicted in warm colors, indicate common activity for counterfactual simulations. Conventions: sM = Semantic Memory; Impl. = Implausible; Plaus. = Plausible; sCFT = semantic counterfactual thinking; eCFT = episodic counterfactual thinking; eM = episodic memory.

across subjects, characterizing each pattern identified by a LV, was determined by bootstrap resampling, using 100 iterations, to estimate the standard errors for each voxel. Clusters larger than 10 voxels (40 mm^3) with a ratio of the salience to the bootstrap standard error values (i.e., the "bootstrap ratio"; BSR) greater than 3.2 ($p < .002$) were reported, consistent with our prior work (see De Brigard et al., 2015). The local maximum for each cluster was defined as the voxel with a BSR higher than any other voxel in a 2-cm cube centered on that voxel. PLS identifies whole-brain patterns of activity in a single analytic step, thus, no correction for multiple comparisons is required. Reported clusters were extracted from the TR that displayed the maximal separation between conditions. Chosen TRs are displayed as a white box around each LV's Temporal Brain score plot (see Figs. 3B, 4B, 5B, and 6B). Regions with the highest BSR and key regions of interest in each LV were entered into NeuroSynth (Yarkoni et al., 2011) to determine functional connectivity and compared to a previously reported 7-network parcellation (Yeo et al., 2011). Voxel intensities for these regions were also extracted to confirm that activity within each region during the conditions followed the general pattern of the LV (see Supplementary Information).

Additionally, a planned ROI analysis was conducted for activity in the hippocampus, employing coordinates extracted using Harvard-Oxford hippocampal masks distributed by FSL (left hippocampus: 952 voxels; right hippocampus: 960 voxels; all voxels were $2 \times 2 \times 2$ mm in MNI space; Desikan et al., 2006). To that end, a general linear model (GLM) analysis on the fMRI data was first conducted on SPM8. Boxcar regressors, convolved with the hemodynamic response, were made to model the time between stimulus presentation and the button press indicating the 6 s average of the *simulation construction period* (e.g., Addis et al., 2009; Addis et al., 2007) for both memories and counterfactuals. The model also included nuisance regressors coding for each run, linear drift across time during each run, head movement parameters (x, y, z, roll, pitch, and yaw), and their mean-centered squares, derivatives, and squared derivatives for each run. Group-level analyses were completed using robust regression (Wager et al., 2005), and the results were corrected to a False Discovery Rate (FDR) threshold of $p < .05$, using both voxel ($p < .005$) and cluster thresholding ($p < .05$). After the univariate analysis (see Supplementary Information for further details), the mean activity across voxels in each hippocampal mask was extracted from the

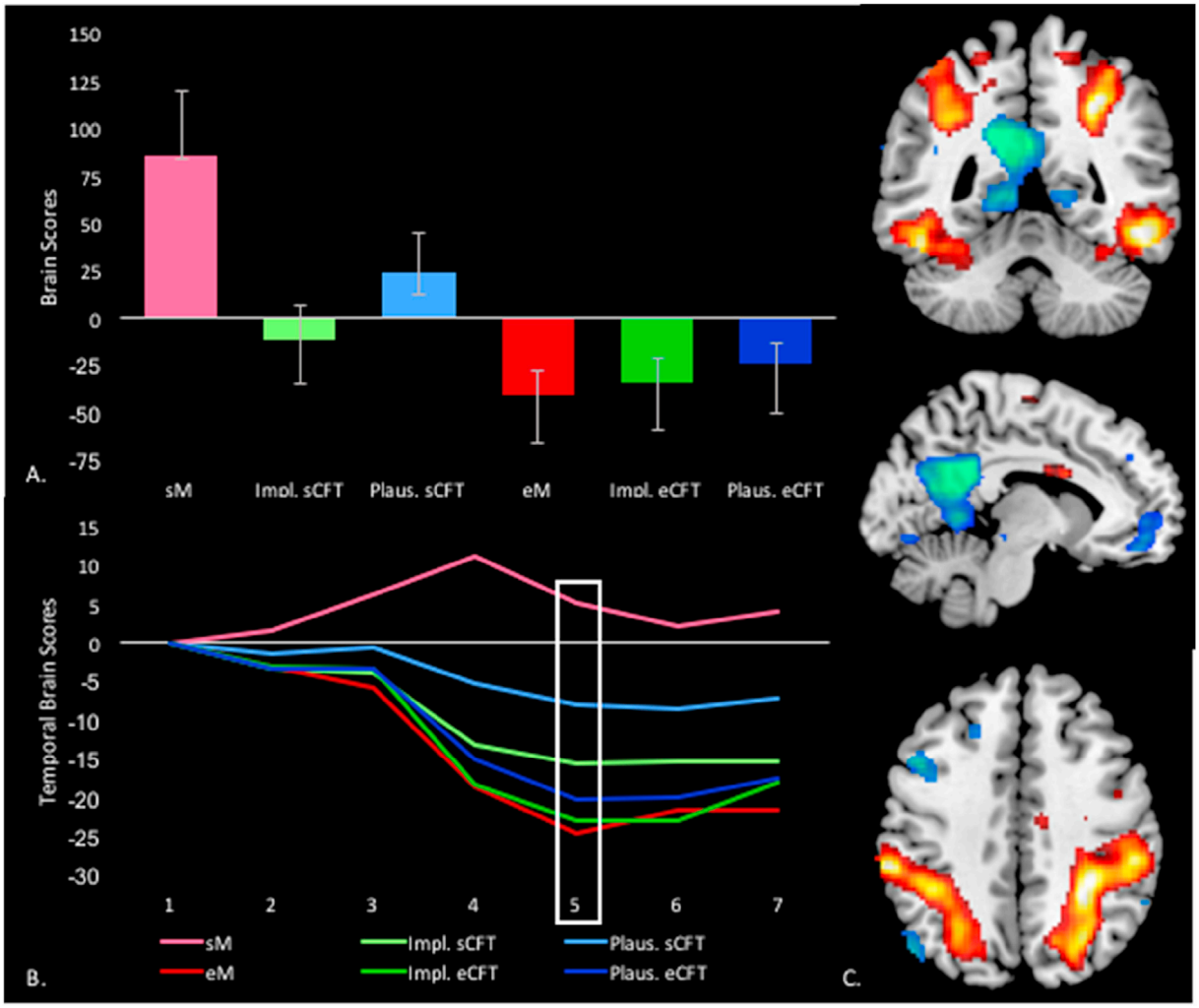

Fig. 4. Results from Latent Variable 2 (LV2) from the mean-centered PLS analysis. A) Plot of brain scores for each condition. Error bars represent confidence intervals from bootstrapping. B) Plot of temporal brain scores indicating weighed average of activation across all voxels in all participants during the length of the mental simulation. The white box represents the TR chosen for maximal separation between conditions. C) Regions with negative saliences in brain scores, depicted here in cool colors, were engaged during eM, Impl. and Plaus. eCFT, while regions with positive saliences, shown here in warm colors, were engaged during sM and Plaus. sCFT. Conventions: sM = Semantic Memory; Impl. = Implausible; Plaus. = Plausible; sCFT = semantic counterfactual thinking; eCFT = episodic counterfactual thinking; eM = episodic memory.

subject-level contrast maps. These were then entered into a group regression analysis using an iteratively reweighted least squares (IRLS) approach. Thus, outliers were weighted less heavily than central data points during analysis. This procedure was conducted separately for the left and right hippocampi.

Results

Behavioral results

An initial two-way ANOVA of memory type (episodic, semantic) by simulation type (memory recall, plausible CFT, implausible CFT) was run on the data from each of the ratings questions. For all three ratings (concreteness, level of detail, and plausibility), there were main effects of memory type (*Concreteness*: $F(1, 26) = 16.3$, $p < .001$, $\eta_p^2 = 0.39$; *Detail*: $F(1, 26) = 18.4$, $p < .001$, $\eta_p^2 = 0.41$; *Plausibility*: $F(1, 26) = 47.5$, $p < .001$, $\eta_p^2 = 0.65$) and simulation type (*Concreteness*: $F(1.69, 43.99) = 79.7$, $p < .001$, $\eta_p^2 = 0.75$; *Detail*: $F(2, 52) = 52.7$, $p < .001$, $\eta_p^2 = 0.67$; *Plausibility*: $F(2, 52) = 76.1$, $p < 0.001$, $\eta_p^2 = 0.75$). Specifically, simulations based on episodic memory were more concrete ($t(26) = -4.02$, $p < .001$, Cohen's $d_z = 0.77$), more detailed ($t(26) = 4.27$, $p < .001$, Cohen's $d_z = 0.82$), and more plausible ($t(26) = 6.91$, $p < .001$, Cohen's $d_z = 1.33$) than those based on semantic memory. Furthermore, memory recall received higher concrete ratings than creating plausible CFT ($t(26) = -9.74$, $p < .001$, Cohen's $d_z = 1.87$), which in turn received higher concrete ratings than creating implausible CFT ($t(26) = -3.26$, $p = .003$, Cohen's $d_z = 0.63$). Simulating a memory involved more detail than simulating either plausible ($t(26) = 8.91$, $p < .001$, Cohen's $d_z = 1.71$) or implausible CFT ($t(26) = 7.46$, $p < .001$, Cohen's $d_z = 1.44$). As expected, recalled memories were rated as most plausible, followed by plausible CFT ($t(26) = 9.30$, $p < .001$, Cohen's $d_z = 1.79$) and then implausible CFT ($t(26) = 5.88$, $p < .001$, Cohen's $d_z = 1.13$).

The ANOVA also revealed an interaction of memory type and simulation type for the Plausibility rating ($F(1.96, 51.05) = 11.8$, $p < .001$, $\eta_p^2 = 0.31$). Follow-up, full simple effects analyses revealed that this interaction was driven by increased differences in plausibility ratings between episodic and semantic memory types when simulating plausible and implausible CFT (*plausible eCFT vs. plausible sCFT:* $t(26) = 6.26$, $p < .001$, Cohen's $d_z = 1.20$; *implausible eCFT vs. implausible sCFT:* $t(26) = 4.10$, $p < .001$, Cohen's $d_z = 0.790$), compared to more similar plausibility ratings across episodic and semantic memory types when

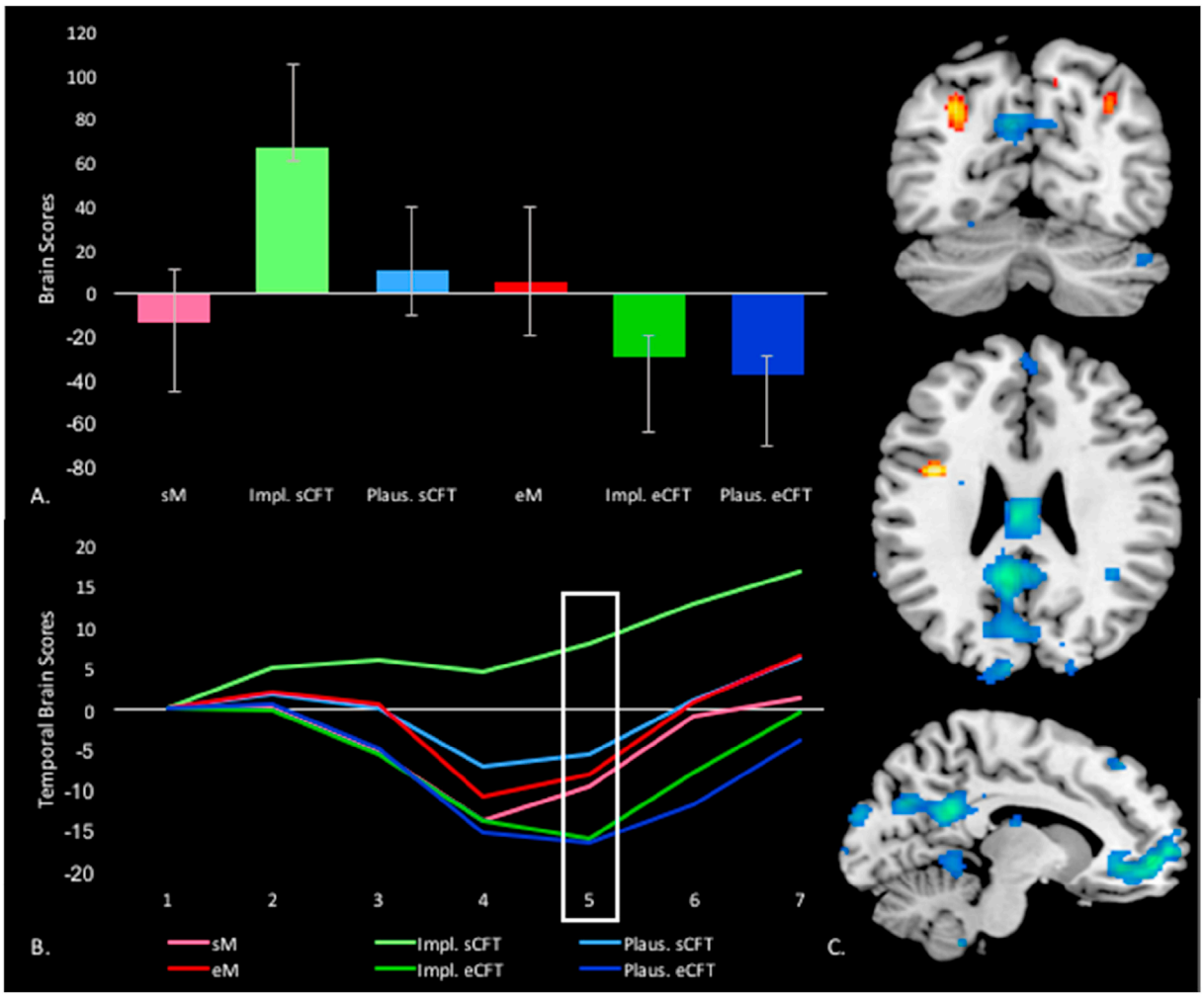

Fig. 5. Results from Latent Variable 3 (LV3) from the mean-centered PLS analysis. A) Plot of brain scores for each condition. Error bars represent confidence intervals from bootstrapping. B) Plot of temporal brain scores indicating weighed average of activation across all voxels in all participants during the length of the mental simulation. The white box represents the TR chosen for maximal separation between conditions. C) Regions with negative saliences in brain scores, depicted here in cool colors, were engaged during Plaus. and Impl. eCFT, while regions with positive saliences, shown here in warm colors, were engaged during Impl. sCFT. Conventions: sM = Semantic Memory; Impl. = Implausible; Plaus. = Plausible; sCFT = semantic counterfactual thinking; eCFT = episodic counterfactual thinking; eM = episodic memory.

recalling the memory (eM vs. sM: $t(26) = 2.86$, $p = .083$, Cohen's $d_z = 0.55$). For a graphical summary of the behavioral results and simple effects analyses, see Fig. 2.

Response time (RT) results

A similar two-way ANOVA of memory type (episodic, semantic) by simulation type (memory recall, plausible CFT, implausible CFT) was run on the simulation construction RT data. The analysis revealed main effects of memory type ($F(1, 21) = 10.3$, $p = .004$, $\eta_p^2 = 0.33$) and simulation type ($F(1.24, 26.00) = 103.1$, $p < .001$, $\eta_p^2 = 0.83$). Semantic simulations (M = 4.06 s) were constructed faster than episodic simulations ($M = 4.36$ s; $t(21) = 3.22$, $p = .017$, Cohen's $d_z = 0.69$). Furthermore, simulations involving remembering an actual memory ($M = 3.13$ s) were constructed faster than plausible CFT simulations ($M = 4.64$ s; $t(21) = 11.3$, $p < .001$, Cohen's $d_z = 2.40$). which were constructed faster than implausible CFT simulations ($M = 4.86$ s; $t(21) = 3.12$, $p = .017$, Cohen's $d_z = 0.67$).

There was also a memory by simulation interaction ($F(2, 42) = 5.21$, $p = .010$, QUOTE $\eta_p^2 = 0.20$). Follow-up simple effects analyses revealed that this interaction was driven by a large difference in eM and sM construction RTs (*mean difference* = 0.66 s, $t(21) = 4.05$, $p = .003$, Cohen's $d_z = 0.86$) compared to plausible eCFT and sCFT construction RTs (*mean difference* = 0.19 s, $t(21) = 1.12$, $p = .554$, Cohen's $d_z = 0.24$) and implausible eCFT and sCFT construction RTs (*mean difference* = 0.07 s, $t(21) = 0.75$, $p = .554$, Cohen's $d_z = 0.16$).

Data-driven PLS analysis

To explore differences and commonalities across all six conditions, a mean-centered PLS analysis was conducted with the fMRI data. This analysis revealed three significant latent variables. The first latent variable (LV1; $p = 0.004$, accounting for 32.6% of the crossblock variance) separated both memory conditions (i.e., sM + eM) from the counterfactual simulation conditions (Fig. 3A). During the window of maximal differentiation in temporal brain scores (TR 6; Fig. 3B), regions associated with memory simulation included bilateral anterior cingulate cortex (ACC; a region sharing functional connectivity with other DN regions, Yeo et al., 2011) as well as PCC (DN), right hippocampus and amygdala (DN), bilateral precuneus (ventral attention network), and left insula (ventral attention and somatomotor networks). Regions associated with counterfactual simulation included left dmPFC (ventral attention network), bilateral dorsolateral PFC (dlPFC; DN and frontoparietal

network), left ventrolateral PFC (vlPFC; DN and frontoparietal network), and left temporal lobe (DN and limbic network; Fig. 3C; Table 1). The second latent variable (LV2; $p < 0.001$, accounting for 31.1% of the crossblock variance) separated two semantic conditions (sM and Plausible sCFT) from the three episodic conditions (eM, Implausible eCFT, and Plausible eCFT; Fig. 4A). During the window of maximal differentiation in temporal brain scores (TR 5; Fig. 4B), the semantic conditions were associated with activity in areas including bilateral lateral PFC (dorsal attention and frontoparietal networks), left cingulate cortex (dorsal attention and frontoparietal networks), bilateral lateral frontal pole (dorsal attention and frontoparietal networks), and right amygdala and hippocampus (limbic network and DN). Episodic simulations were associated with activity in bilateral vmPFC (BA 10; DN), left dorsolateral and ventrolateral PFC (DN), left PCC and paracingulate gyrus (DN), bilateral lingual gyri (DN and visual network), and bilateral temporal cortex (DN; Fig. 4C; Table 2). Finally, the third latent variable (LV3; $p = 0.026$, accounting for 19.4% of the crossblock variance) separated the Implausible sCFT from both Plausible and Implausible eCFT (Fig. 5A). During the window of maximal differentiation in temporal brain scores (TR 5, Fig. 5B), Implausible sCFT were associated with activity in bilateral lateral PFC (frontoparietal and dorsal attention networks), bilateral superior occipital cortex (dorsal attention network), and left inferior temporal gyrus (frontoparietal and dorsal attention networks). By

Table 1
fMRI clusters associated with LV1, Memory vs. CFT, from the PLS analysis. Conventions: BA = Brodmann's area in which the peak of the cluster lies; Hemi. = cerebral hemisphere (L – left or R – right); MNI = Montreal Neurological Institute standard space; BSR = bootstrap ratio derived from the partial least squares analysis; ACC/PCC = anterior/posterior cingulate cortex.

Region(s)	BA	Hemi.	Peak MNI Coordinates			# of Voxels	BSR
			X	Y	Z		
All CFT > Memory							
Superior Frontal Gyrus	6	L	−8	10	54	1829	8.3277
Lingual Gyrus	18	L	−8	−84	−2	5098	6.8822
Inferior Frontal Gyrus	44	L	−52	16	16	2954	6.8553
Substantia Nigra	–	L	−12	−20	−14	81	6.0022
Middle Temporal Gyrus	22	L	−50	−38	−2	278	5.5184
Middle Frontal Gyrus	8	R	44	16	48	1128	5.4761
Frontal Gyrus/ Frontal Pole	11	R	48	48	−16	232	5.2449
Cerebellum	–	L	−22	−34	−32	20	5.2199
		R	16	−36	−40	82	4.9313
		L	−34	−66	−36	148	3.9573
Inferior Parietal Lobule (IPL)	40	R	44	−62	40	281	5.2113
		L	−50	−46	44	17	3.551
Inferior Temporal Gyrus	20	L	−54	−4	−38	38	4.8477
Fusiform Gyrus	37	L	−40	−56	−22	69	4.6746
Thalamus	–	R	10	−6	0	105	4.6434
		L	−14	−22	4	123	4.0632
Lateral Occipital Cortex	39	L	−50	−64	26	56	4.2352
Precuneus	7	R	2	−70	54	20	3.9299
Lingual Gyrus/ Intracalcarine Cortex	30	R	14	−64	4	25	3.7738
Memory > All CFT							
ACC	32	L	−2	36	−2	1529	−10.3654
	33	R	2	12	24	139	−4.6804
	24	L	0	−8	34	67	−4.4045
Superior Temporal Gyrus	38	R	36	2	−24	295	−6.0272
	42	R	66	−18	8	1381	−5.3559
Insula	13	L	−40	−14	−6	1297	−5.3793
Superior Parietal Lobule	7	L	−26	−50	62	121	−4.8547
Precentral Gyrus	6	L	−64	0	24	37	−4.7813
PCC	31	L	−6	−30	40	268	−4.5677
Precuneus	31	L	−8	−66	22	186	−4.3458
	5	R	10	−44	56	14	−3.686
Cuneus	7	R	18	−80	40	19	−4.3221
	19	L	−20	−88	38	12	−3.4815
Superior Occipital Gyrus	19	L	−38	−88	26	30	−3.8542
Supramarginal Gyrus	2	L	−66	−24	28	19	−3.7206
Hippocampus	28	R	18	−8	−24	11	−3.5509
Planum Temporale	41	L	−36	−32	12	27	−3.5316

Table 2
fMRI clusters associated with LV2, Episodic vs. Semantic, from the PLS analysis. Conventions: BA = Brodmann's area in which the peak of the cluster lies; Hemi. = cerebral hemisphere (L – left or R – right); MNI = Montreal Neurological Institute standard space; BSR = bootstrap ratio derived from the partial least squares analysis; ACC/PCC = anterior/posterior cingulate cortex.

Region(s)	BA	Hemi.	Peak MNI Coordinates			# of Voxels	BSR
			X	Y	Z		
Semantic > Episodic							
Inferior Temporal Gyrus	37	R	48	−54	−14	2040	11.0051
		L	−48	−64	−10	2137	9.7597
Supramarginal Gyrus	40	L	−58	−34	44	4014	8.8362
Superior Parietal Lobule	7	R	30	−48	44	5825	8.8135
Inferior Frontal Gyrus	9	L	−42	−2	22	1933	8.2134
		R	48	10	26	1066	7.715
Lateral Occipital Cortex	19	L	−28	−74	30	166	6.8939
Frontal Pole	45	L	−52	38	8	480	6.1373
		R	48	36	8	32	4.1754
ACC	24	L	−6	6	28	221	5.6654
Superior Frontal Gyrus	24	R	20	−2	50	587	5.6515
Insula	13	R	38	2	−4	549	5.2156
Precentral Gyrus	6	R	46	−4	46	87	4.5354
		L	−8	−14	68	22	3.7848
Fusiform Gyrus	20	L	−34	−6	−36	26	4.3679
Subcallosal Cortex	25	R	12	12	−16	24	4.287
Superior Temporal Gyrus	41	R	44	−34	8	34	4.2175
Amygdala	–	R	24	−4	−22	40	4.1342
Postcentral Gyrus	5	R	14	−46	66	41	4.0461
Thalamus	–	R	4	−4	12	11	3.6324
Superior Frontal Gyrus	32	L	−20	6	48	13	3.4618
Episodic > Semantic							
PCC	23	L	−6	−52	24	2309	−7.71
Angular Gyrus	39	L	−44	−60	24	1164	−6.1458
Middle Temporal Gyrus/Temporal Pole	21	L	−58	−14	−18	290	−5.824
		L	−42	2	−38	135	−5.1566
		R	56	0	−30	57	−4.481
Middle Frontal Gyrus	8	L	−46	14	44	183	−5.3155
		L	−22	26	38	107	−4.0018
Lateral Occipital Cortex	39	R	58	−60	20	196	−5.1933
Frontal Pole/ Medial Frontal Gyrus	10	L	−12	52	−6	474	−5.1711
		R	8	56	4	60	−4.4575
		L	−46	56	−8	30	−3.994
Lingual Gyrus	30	R	18	−46	0	52	−4.0374
	18	L	−8	−78	−10	53	−3.9951
Cerebellum	–	L	−2	−56	−32	20	−3.8654
Paracingulate Gyrus	9	L	−8	44	36	14	−3.6984
Frontal Orbital Cortex	47	L	−46	32	−14	19	−3.6853
Intracalcarine Cortex	17	R	14	−80	4	22	−3.6638

contrast, eCFT were associated with activity in areas including left vmPFC (BA 10; DN), bilateral PCC (DN), bilateral occipital pole (visual network), bilateral middle temporal gyri (DN) and bilateral superior frontal gyri (DN; Table 3; Fig. 5C).

CFT-only PLS analysis

To investigate neural differences in CFT specific to episodicity and/or plausibility, a mean-centered PLS analysis with only the CFT conditions was conducted. This analysis revealed one significant latent variable (CFT-only_LV1; $p = 0.02$) accounting for 49.8% of the crossblock variance. CFT-only_LV1 separated semantic counterfactuals from episodic counterfactuals (Fig. 6A). Similar to the pattern of activity associated with eCFT from LV3, CFT-only_LV1 indicated that both Implausible and Plausible eCFT correlated with activity in left PCC (DN), bilateral occipital pole (visual network), bilateral angular gyri (DN), bilateral temporal pole and middle temporal gyri (DN and limbic network), bilateral middle occipital gyri (visual network), left insula (DN and ventral attention network), and right angular gyrus (DN). By contrast, sCFT correlated with activity in regions including bilateral inferior temporal gyri (frontoparietal and dorsal attention networks), bilateral inferior frontal gyri (frontoparietal network), bilateral occipital cortex (frontoparietal and dorsal attention networks), and bilateral frontal pole (frontoparietal network; Table 4; Fig. 6C).

Table 3
fMRI clusters associated with LV3, Implausible sCFT vs. Episodic CFT, from the PLS analysis. Conventions: BA = Brodmann's area in which the peak of the cluster lies; Hemi. = cerebral hemisphere (L – left or R – right); R-L and L-R indicate that the cluster crosses the midsaggital line, but the peak voxel is in the hemisphere listed first; MNI = Montreal Neurological Institute standard space; BSR = bootstrap ratio derived from the partial least squares analysis; PCC = posterior cingulate cortex.

Region(s)	BA	Hemi.	Peak MNI Coordinates			# of Voxels	BSR
			X	Y	Z		
Implausible Semantic CFT > Episodic CFT							
Inferior Frontal Gyrus	45	L	−48	24	12	98	5.5506
Precentral Gyrus	9	L	−42	2	24	92	4.9808
Inferior Temporal Gyrus	37	L	−54	−52	−12	43	4.9182
Lateral Occipital Cortex	39	L	−30	−66	36	142	4.676
	19	R	32	−68	40	35	3.7675
Middle Frontal Gyrus	6	R	28	12	48	44	3.7556
Episodic CFT > Implausible Semantic CFT							
Frontal Pole/ Medial Frontal Gyrus	10	L	−4	58	−2	1796	−7.6092
PCC	23	R - L	2	−22	32	607	−6.7952
	30	L - R	−8	−50	24	1136	−6.5143
Supramarginal Gyrus	22	R	64	−44	12	345	−5.1706
Middle Temporal Gyrus	21	R	52	2	−22	86	−5.0715
		L	−62	−22	−12	202	−4.7234
		R	68	−20	−6	34	−4.0499
	20	R	46	−28	−14	57	−3.7654
Occipital Pole	18	L	−24	−98	16	259	−4.9692
	19	R	10	−92	34	246	−4.9643
	18	R	8	−90	18	21	−3.8065
Cingulate Cortex	24	L	−16	−6	46	15	−4.6174
Middle Frontal Gyrus	9	L	−54	16	34	47	−4.5672
Temporal Pole	38	R	60	18	−10	11	−4.4474
	21	R	36	4	−42	35	−4.1574
			56	10	−38	50	−4.0658
Inferior Occipital Gyrus	18	L	−46	−82	−18	73	−4.3941
Superior Frontal Gyrus	8	L	−10	38	52	92	−4.2753
	6	R	6	20	60	14	−3.8118
	9	L	−24	40	42	33	−3.7783
			−2	48	38	79	−3.5914
Angular Gyrus	13	R	38	−48	24	66	−4.2642
Lingual Gyrus	19	L	−12	−46	−8	191	−4.1559
Middle Occipital Gyrus	19	R	36	−78	12	98	−4.0401
		L	−38	−90	2	78	−3.7938
Cerebellum	–	L	−6	−48	−42	14	−3.9559
		R	8	−54	−48	23	−3.8626
		R	46	−66	−28	23	−3.8604
Fusiform Gyrus	19	R	26	−76	−20	44	−3.9493
	37	R	24	−50	−12	21	−3.8003
Superior Temporal Gyrus	22	L	−68	−36	18	12	−3.9026
			−60	−58	10	41	−3.7808
	39	L	−46	−56	12	81	−3.7183
Claustrum	–	L	−28	6	12	12	−3.8099
Insula	13	L	−38	−12	2	11	−3.6693
Subcallosal Cortex	25	L	−4	12	−10	16	−3.5785

ROI analysis

Since previous studies have revealed mixed findings regarding the degree of hippocampal activity both during eCFT as well as plausible versus implausible hypothetical simulations, hippocampal ROI analyses were conducted. For the right hippocampal ROI, only one contrast returned significant differences in activity levels: All Episodic simulations – All Semantic simulations ($M = 0.06$, $SEM = 0.03$, $p = .046$). For the left ROI, however, this analysis revealed significant differences in hippocampal activity in five contrasts of interest: Episodic All – Semantic All ($M = 0.08$, $SEM = 0.03$, $p = .027$), Episodic Memory (eM) – baseline ($M = 0.18$, $SEM = 0.05$, $p < .001$), Implausible eCFT – Implausible sCFT ($M = 0.07$, $SEM = 0.03$, $p = .022$), Implausible eCFT – baseline ($M = 0.13$, $SEM = 0.04$, $p = .004$), and Implausible sCFT – baseline ($M = 0.07$, SEM: 0.04, $p = .018$). Bilaterally, episodic simulations recruited the hippocampus more than semantic simulations, though the left hippocampus was more engaged during implausible relative to plausible CFT simulations, for both eCFT and sCFT (Fig. 7; see also Supplementary Information).

Discussion

Research on the cognitive neuroscience of mental simulation has spurred interest in the neural underpinnings of counterfactual thinking. Early findings suggested that eCFT, like eM and episodic future thinking, recruited core regions of the DN (Addis et al., 2009; Van Hoeck et al., 2013). However, further studies qualified this observation by noting different patterns of brain activity depending on whether the CFT simulations were perceived as plausible or implausible, self-versus other-based, or if they involved imagining alternatives to semantic and/or object-based factual information (De Brigard et al., 2013, 2015; Kulakova et al., 2013). To systematically explore differences in episodicity and plausibility, the current project investigates neural activity associated with both plausible and implausible eCFT and sCFT, and compares it with neural activity associated with eM and sM. The experimental design employed enabled us to test three aims. Using a multivariate approach, we sought to uncover the neural differences between CFT and memory reactivation and determine how episodicity and plausibility of the alternate event affected DN activity. Finally, ROI analyses allowed us to focus on the hippocampus in order to elucidate previous conflicting results regarding its involvement in episodic hypothetical simulations and their perceived plausibility.

Though memory and CFT simulation share a common network of brain activity (Addis et al., 2009; Benoit and Schacter, 2015), we predicted reliable differences in the pattern of brain activation across all counterfactual conditions relative to both memory conditions. Our data-driven analysis confirmed this hypothesis by identifying a reliable group of brain regions that segregated both episodic and semantic memory retrieval from all four types of counterfactual simulations (i.e., both implausible and plausible eCFT and sCFT). This result suggests that

episodic and semantic simulations share common activity during memory reactivation and during CFT, despite differences between the memory types. Consistent with previous work on episodic memory reactivation (Benoit and Schacter, 2015; Spreng and Grady, 2010; St. Jacques, Carpenter, Szpunar and Schacter, 2017), the memory conditions activated pure, core regions of the DN more so than the CFT conditions. By contrast, increased activity in dlPFC, vlPFC, and temporal cortex was preferentially associated with all four CFT conditions. Recent research on the cognitive neuroscience of CFT has associated these regions with neural networks engaged in goal-directed activity, semantic knowledge, and cognitive control (St. Jacques et al., 2017; Van Hoeck, Watson and Barbey, 2015). In agreement with this framework, we interpret activity in these regions as reflecting processing of both existing and novel information in working memory during the creation of mental simulations involving alternative scenarios.

Recalling a memory and creating a CFT are inherently different tasks; one involves retrieving what is known and true while the other escapes reality through an alternative version of a true event or statement. Previous work on neural responses to fictional and realistic scenarios has found that, while reading about scenarios with real characters, precuneus and PCC, major DN regions, were more active, while during fictional scenarios, there was greater activation in inferior frontal and occipital regions (Abraham et al., 2008). Consistent with this finding, our results suggest that, even when the fictional event is developed from a real event and could have possibly occurred, the brain still distinguishes these concocted scenarios from actual events.

The differences between memory recollection and CFT may also suggest that multiple regions may be differentially recruited during mental simulation, with each region activated to different levels, and within different functional ensembles, depending upon the specific simulation. A recent meta-analysis suggests that, during memory retrieval, the DN, dorsal attentional network, and ventral attentional network were all active (Kim, 2010). However, DN activity was higher for trials that were fully recalled rather than experienced as merely familiar, and the ventral attention network is more active with increasing familiarity. By contrast, the dorsal network, closely related to both the dorsal attention network and the frontoparietal network in the 7-parcellation model (Yeo et al., 2011), showed the opposite result, with higher activity during familiarity relative to recollection. The contrast between memory versus CFT from our first analysis could also be interpreted within the recollection/familiarity framework. Since creating a CFT involves the retrieval of information stored in memory with the purpose of

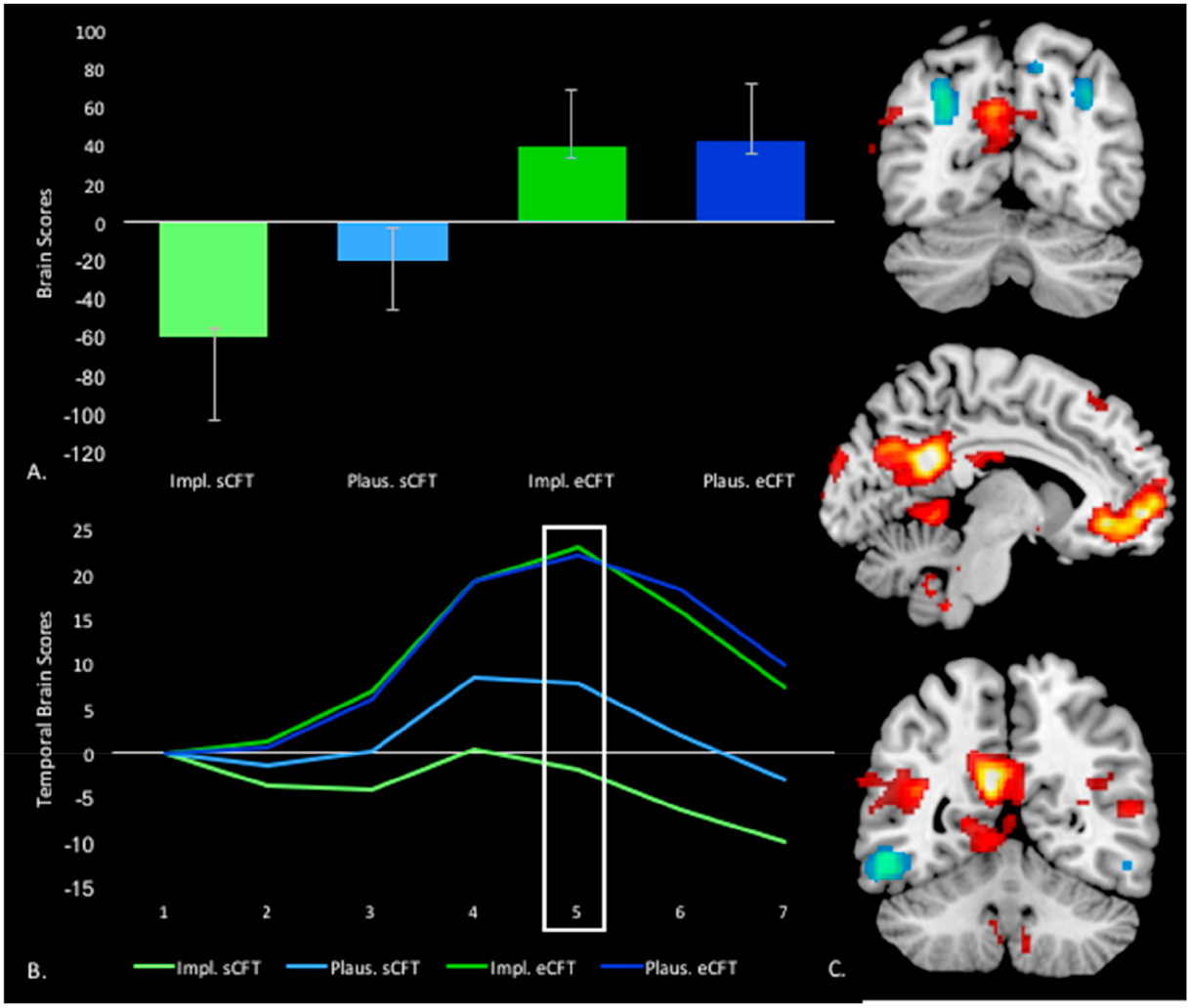

Fig. 6. Results from Latent Variable 1 (CFT-only_LV1) from the CFT-only mean-centered PLS analysis. A) Plot of brain scores for each condition. Error bars represent confidence intervals from bootstrapping. B) Plot of temporal brain scores indicating weighed average of activation across all voxels in all participants during the length of the mental simulation. The white box represents the TR chosen for maximal separation between conditions. C) Regions with negative saliences in brain scores, depicted here in cool colors, were engaged during Impl. and Plaus. eCFT, while regions with positive saliences, shown here in warm colors, were engaged during Plaus. and Impl. sCFT. Conventions: sM = Semantic Memory; Impl. = Implausible; Plaus. = Plausible; sCFT = semantic counterfactual thinking; eCFT = episodic counterfactual thinking; eM = episodic memory.

Table 4
fMRI clusters associated with LV1 from the CFT-only PLS analysis, eCFT vs. sCFT. Conventions: BA = Brodmann's area in which the peak of the cluster lies; Hemi. = cerebral hemisphere (L – left or R – right); MNI = Montreal Neurological Institute standard space; BSR = bootstrap ratio derived from the partial least squares analysis; PCC = posterior cingulate cortex.

Region(s)	BA	Hemi.	Peak MNI Coordinates			# of Voxels	BSR
			X	Y	Z		
Episodic CFT > Semantic CFT							
PCC	31	L	−8	−52	22	1781	9.0617
	23	L	0	−22	32	463	7.558
Frontal Pole	10	L	−8	56	−2	2051	8.3609
	9	L	−2	58	30	106	4.8663
Occipital Pole	18	L	−24	−98	18	236	6.3672
		R	22	−100	16	240	4.5122
		L	−12	−102	4	23	3.747
Lingual Gyrus	19	L	−10	−48	−4	415	5.4449
Angular Gyrus	22	L	−40	−56	20	597	5.3794
	13	R	38	−48	24	149	5.1624
	22	R	64	−46	12	210	4.8658
Middle Temporal Gyrus	21	L	−58	−18	−14	341	5.1845
		R	52	2	−22	97	4.765
		R	52	−22	−14	65	3.9457
Middle Frontal Gyrus	9	L	−44	12	42	104	4.6027
			−22	28	36	31	3.5486
Temporal Pole	21	L	−38	10	−40	17	4.5579
		R	54	12	−40	46	4.4113
		R	36	4	−42	37	3.9551
Insula	13	L	−32	10	−14	28	4.4413
Superior Frontal Gyrus	8	L	−4	32	56	62	4.0816
	9	L	−24	38	42	26	3.9152
Cerebellum	–	R	6	−52	−38	73	4.0768
		R	26	−76	−20	31	3.7848
Fusiform Gyrus	37	R	24	−50	−12	23	3.9736
Subcallosal Cortex	25	L	−2	14	−10	22	3.7245
Middle Occipital Gyrus	18	L	−38	−84	−4	27	3.6825
	19	R	36	−78	12	22	3.4257
Semantic CFT > Episodic CFT							
Inferior Temporal Gyrus	37	L	−54	−52	−12	295	−6.9339
	20	R	52	−52	−14	29	−3.6983
Inferior Frontal Gyrus	9	L	−44	2	24	136	−5.1531
	46	L	−48	26	14	48	−3.8537
Lateral Occipital Cortex	39	L	−30	−68	36	322	−4.9497
	19	R	32	−68	38	182	−4.3671
Fusiform Gyrus	36	L	−34	−36	−26	35	−4.8689
Precuneus	7	R	10	−66	50	36	−4.1323
Middle Frontal Gyrus	6	R	28	8	52	65	−3.9497
Frontal Pole	45	L	−50	38	10	17	−3.8437
		R	48	36	8	13	−3.5881
Supramarginal Gyrus	40	L	−40	−44	38	24	−3.8275

manipulating rather than reinstating it, as occurs during memory recall, the increased dorsal frontal and parietal activity during CFT may possibly be signaling familiarity with the memorial contents. By contrast, the increased activation in DN and ventral attention during memory recall may be indexing a higher degree of recollection. This interpretation is also consistent with previous findings suggesting that CFT simulations more closely resembling the actual memory should activate DN to a greater extent, as they involve less manipulation of memorial contents and greater degree of memory reinstatement, as in recollection. Finally, although the engagement of these different networks is consistent with brain activity associated with autobiographical memory (Addis et al., 2009; Benoit and Schacter, 2015; St. Jacques and De Brigard, 2015), it is important to note that our first PLS analysis also significantly separates eM and sM from one another (Fig. 2A and B). This differential activation for eM and sM suggests differential contributions to factual autobiographical versus non-autobiographical knowledge during memory retrieval. Further research is needed to fully understand the precise contributions of autobiographical versus non-autobiographical information to the generation of CFT and its relation to increased DN activity.

Second, we hypothesized a reliable difference in the pattern of brain activation across all three semantic simulations (sM, plausible sCFT, and implausible sCFT) relative to all three episodic simulations (eM, Plausible eCFT and Implausible eCFT). Based on prior work, we expected episodic simulations to activate more DN than semantic simulations. Our analyses mostly confirmed this hypothesis; our multivariate analyses separated 1) all episodic simulations from sM and plausible sCFT and 2) eCFT from sCFT. Only our hippocampal ROI analysis identified significant differences in activity when contrasting the episodic simulations from all three semantic simulations. Nonetheless, as expected, whole-brain activity during episodic simulations corresponded with DN core regions, particularly PCC and vmPFC. This result is consistent with a recent review of neural activity during eCFT (Van Hoeck et al., 2015), the only difference being ITG activity, which Van Hoeck et al. suggest is involved in semantic processing and, in our results, was preferentially associated with sCFT rather than eCFT. Other brain regions that showed increased activation during the semantic conditions included clusters in superior parietal/supramarginal gyri, inferior frontal gyrus (IFG), and frontal pole, as well as small clusters in MTL near the amygdala and entorhinal/perirhinal cortices. Critically, many of these regions are not typically associated with DN, further indicating their preferential engagement during the manipulation of non-autobiographical information and, by contrast, further supporting the self-referentiality and episodicity of the mental simulations associated with activity in DN.

This study provided unpredicted findings with regard to semantic simulations. Previous work on sCFT found activity in caudate and cuneus during semantic simulations (Kulakova et al., 2013), neither of which appear in our results. Though this null result does not guarantee complete lack of semantic activity in these regions, the discrepancy may be due to the difference in our paradigms. We compared semantic simulations to episodic ones (Figs. 4, 6 and 7), while Kulakova et al. (2013) looked at contrasts of sCFT against hypothetical statements that were not counterfactual in nature. These non-CFT hypotheticals, e.g. "If it rained then the street was wet" (Kulakova et al., 2013, pg. 265), also contained semantic information and thus only isolated the CFT aspects of the linguistic interpretation and not the semantic nature of sCFT. Nonetheless, the regions activated in our analyses during semantic simulations were consistent with literature on semantic memory: neuropsychology research in semantic dementia (Hodges and Graham, 2001; Hodges and Patterson, 2007) and neuroimaging of item memories (Davachi, 2006) demonstrate that inferior temporal areas and perirhinal, parahippocampal, and entorhinal cortices are critical during semantic processing.

The third hypothesis pertained to the identification of reliable differences between neural activity associated with plausible relative to implausible simulations. However, this hypothesis was not supported by the data, as neither of our multivariate approaches reliably differentiated implausible vs. plausible CFT across both episodic and semantic memory. Instead, it separated neural activity associated with implausible sCFT from both implausible and plausible eCFT. Consistent with our contrasts comparing eCFT to semantic simulations, both kinds of eCFT engaged regions typically associated with eCFT and DN, including vmPFC, bilateral PCC, and bilateral superior temporal/IPL. By contrast, implausible sCFT engaged regions including left IFG, left ITG, left precentral gyrus, and right mFG, mirroring findings for sCFT from our previous contrasts. These regions are not typically associated with the DN, perhaps because implausible sCFT simulation is the most different from eM reactivation. Based on participants' self-reported ratings, semantic implausible counterfactuals were considered the least plausible (or most implausible) scenarios. Given these behavioral results, it is not surprising that regions activated by implausible sCFT correspond with regions activated when imagining fictional events, particularly inferior frontal and occipital areas (Abraham et al., 2008).

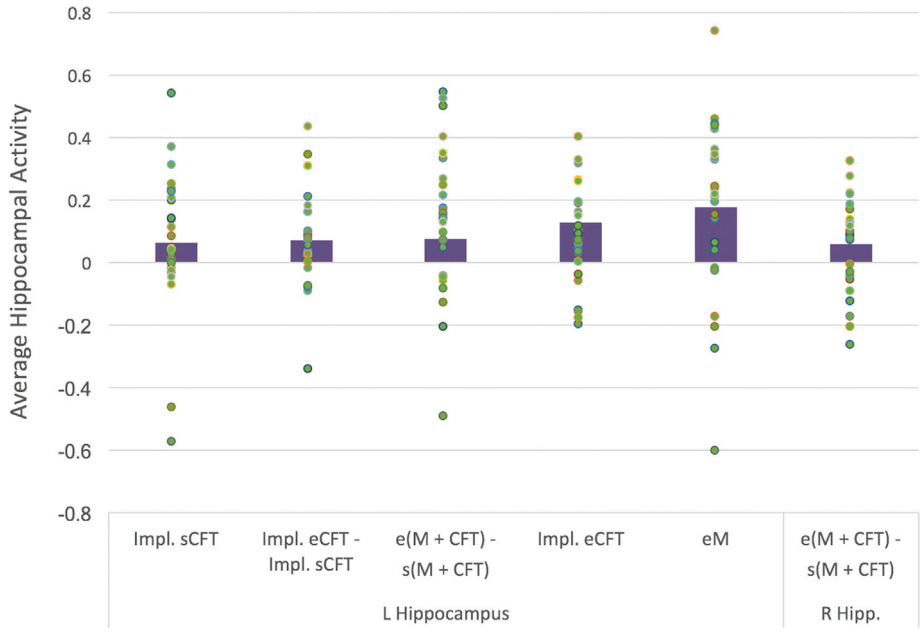

Fig. 7. Significant contrasts from the Hippocampal ROI analysis. Individual subject's contrast values (n = 27) are spread as a scatter plot around the plotted average. Conventions: sM = Semantic Memory; Impl. = Implausible; sCFT = semantic counterfactual thinking; eCFT = episodic counterfactual thinking; eM = episodic memory; Hipp. = hippocampus. Single condition items are contrasted against an implicit baseline.

One reason the data-driven approach may not have identified a difference in CFT plausibility may be due to the uniqueness of implausible sCFT. Aforementioned multivariate results comparing implausible sCFT to all eCFT, and even the contrast comparing all eCFT to all sCFT, separate implausible sCFT from all other simulations (see Fig. 5a and b and 6a,b). It activates a unique set of regions in superior occipital and lateral frontal areas. Therefore, the multivariate analysis could not find commonalities between implausible sCFT and implausible eCFT. Thus, we failed to see a clear neural separation based on plausibility in the current study, as was seen previously (De Brigard et al., 2013). The previous study found that more plausible counterfactuals activated more DN core regions than less plausible counterfactuals. Instead, our results showed us that eCFT recruit more DN regions than implausible sCFT. One explanation is that plausibility is driving this result as well, as both plausible and implausible eCFT are rated as more plausible than implausible sCFT (Fig. 2). However, it is difficult to disentangle this result from the episodicity hypothesis, as eCFT and implausible sCFT also differ on this criterion. Furthermore, our CFT-only PLS analysis (Fig. 6) does not support this hypothesis; there, both sCFT are separated from both eCFT, and our behavioral plausibility ratings do not mimic this separation.

Finally, to investigate whether the degree of change in hippocampal activity during counterfactual simulation was due to episodicity or perceived plausibility, an ROI analysis was conducted. Previous research (De Brigard et al., 2013) reported increased hippocampal activity as a function of increase in perceived plausibility during eCFT. This suggests that it may be the plausible nature of the simulation that may drive the engagement of hippocampal activity. By contrast, Van Hoeck et al. (2015), replicating results from Addis et al. (2009), reported increased hippocampal activation during both episodic future and eCFT relative to a baseline condition. Consistent with this work, bilateral hippocampal activity was greater for episodic relative to semantic simulations. However, unlike the findings from De Brigard et al. (2013), left hippocampus was significantly more engaged during implausible CFT relative to baseline regardless of whether such simulations were semantic or episodic-based—though implausible eCFT did show increased hippocampal activity relative to implausible sCFT. We interpret this finding in light of recent proposals suggesting that increased activity in the hippocampus during mental simulation may reflect an increase in processing effort during the construction of simulations that involve more arbitrary relational binding (Van Hoeck et al., 2015; Weiler et al., 2010). Specifically, since implausible CFT requires connections between items that may not normally be associated, the hippocampus is more likely to be engaged than during the novel but realistic simulations of plausible CFT. Furthermore, it is thought that when information is less consistent with prior knowledge, the hippocampus is more involved (Henke, 2010). These findings explain why the hippocampus is engaged during implausible CFT more than plausible CFT, but not why it is also more activated during episodic simulations as a whole. However, recent work on autobiographical future planning suggests that the hippocampus is sensitive to both level of detail and the abstractness of a plan (Spreng et al., 2015). Specifically, novel plans with more detail but also more abstract or farther-off plans engaged the hippocampus. Consistent with these results, our implausible CFT were rated as more abstract than plausible CFT, but episodic CFT were not consistently rated as more detailed than semantic CFT (Fig. 2). Further research needs to be done to support this finding in a CFT context.

One limitation to our study is that the results might be influenced by differing levels of effort or difficulty required by each condition. Because there are differences in how long it takes participants to construct the simulation for each condition, one could argue that the implausible eCFT condition is the most effortful and the sM condition the easiest to simulate, and that these differences may play a role in the fMRI results. If this were the case, however, we expect that our analysis would have identified at least one significant pattern of activity in which conditions are separated according to effort or difficulty, that is, a pattern whereby memory conditions, plausible conditions, and implausible conditions are separated from each other, in that order. However, not a single output from our data-driven analyses, statistically significant or not, conformed to that pattern of results. Consequently, although we acknowledge that effort and difficulty may differ across conditions, they do not seem to be the main factor driving the neural differences in our findings.

To our knowledge, the current study is the first to look at semantic CFT in comparison to semantic memory, demonstrating that sCFT activates brain regions associated with mental simulation, but to a lesser extent than does eCFT. This finding is consistent with the semantic scaffolding hypothesis, according to which semantic knowledge provides

a necessary stepping stool for episodic memory retrieval and future simulation (Irish and Piguet, 2013). Further work could explore whether semantic knowledge is equally necessary for eCFT. Furthermore, the role of plausibility in CFT is yet to be determined. The hippocampus seems to favor implausible CFT, but no other regions clearly distinguished both episodic and semantic implausible from plausible CFT. Future studies could further dissect the effects of plausibility. It would also be critical to study sCFT that include autobiographical content. Many naturally occurring semantic counterfactual thoughts involve self-relevant information, e.g. "if only I were taller." The current experiment intentionally avoided self-relevant facts to create a clear separation between episodic and semantic simulations, especially when creating alternative versions, as elaborating on an autobiographical fact may bring episodic events to mind. However, this category of stimuli may have a unique blend of neural activity that would be worth studying.

Conclusion

Overall, the findings of the current study suggest that activity during episodic and semantic simulations differ in levels of DN activation by comparing CFT activity to memory reactivation. We further demonstrated that episodic simulations, regardless of plausibility, may be more similar to one another than semantic simulations, which differ greatly from sM to plausible sCFT to implausible sCFT. Finally, we found that the hippocampus may be active during semantic as well as episodic CFT, particularly implausible sCFT, due to its abstract nature and greater processing need. These results push the growing field of mental simulation by demonstrating how different memorial content differentially contributes to distinct types of counterfactual thinking.

Acknowledgements

This research was funded by a grant from the Duke Institute for Brain Sciences, Duke University. NP was independently funded with Government support under and awarded by DoD, Air Force Office of Scientific Research, National Defense Science and Engineering Graduate (NDSEG) Fellowship, 32 CFR 168a throughout her work on this project.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.neuroimage.2018.05.053.

References

Abraham, A., Von Cramon, D.Y., Schubotz, R.I., 2008. Meeting George Bush versus meeting Cinderella: the neural response when telling apart what is real from what is fictional in the context of our reality. J. Cognit. Neurosci. 20 (6), 965–976.

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. https://doi.org/10.1016/j.neuropsychologia.2008.10.026.

Addis, D.R., Schacter, D.L., 2008. Constructive episodic simulation: temporal distance and detail of past and future events modulate hippocampal engagement. Hippocampus 18 (2), 227–237. https://doi.org/10.1002/hipo.20405.

Addis, D.R., Schacter, D.L., 2012. The Hippocampus and imagining the future: where do We Stand? Front. Hum. Neurosci. 5. https://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. https://doi.org/10.1016/j.neuropsychologia.2006.10.016.

Andrews-Hanna, J.R., Smallwood, J., Spreng, R.N., 2014. The default network and self-generated thought: component processes, dynamic control, and clinical relevance. Ann. N. Y. Acad. Sci. 1316 (1), 29–52. https://doi.org/10.1111/nyas.12360.

Benoit, R.G., Schacter, D.L., 2015. Specifying the core network supporting episodic simulation and episodic memory by activation likelihood estimation. Neuropsychologia 75, 450–457. https://doi.org/10.1016/j.neuropsychologia.2015.06.034.

Buckner, R.L., Andrews-Hanna, J.R., Schacter, D.L., 2008. The Brain's default network. Ann. N. Y. Acad. Sci. 1124 (1), 1–38. https://doi.org/10.1196/annals.1440.011.

Buhrmester, M., Kwang, T., Gosling, S.D., 2011. Amazon's Mechanical Turk: a new source of inexpensive, yet high-quality, data? Perspect. Psychol. Sci. 6 (1), 3–5. https://doi.org/10.1177/1745691610393980.

Byrne, R.M.J., 2005. The Rational Imagination: How People Create Alternatives to Reality. MIT Press.

Davachi, L., 2006. Item, context and relational episodic encoding in humans. Curr. Opin. Neurobiol. 16 (6), 693–700. https://doi.org/10.1016/j.conb.2006.10.012.

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. https://doi.org/10.1016/j.neuropsychologia.2013.01.015.

De Brigard, F., Giovanello, K.S., 2012. Influence of outcome valence in the subjective experience of episodic past, future, and counterfactual thinking. Conscious. Cognit. 21 (3), 1085–1096. https://doi.org/10.1016/j.concog.2012.06.007.

De Brigard, F., Giovanello, K.S., Stewart, G.W., Lockrow, A.W., O'Brien, M.M., Spreng, R.N., 2016. Characterizing the subjective experience of episodic past, future, and counterfactual thinking in healthy younger and older adults. Q. J. Exp. Psychol. 69 (12), 2358–2375. https://doi.org/10.1080/17470218.2015.1115529.

De Brigard, F., Rodriguez, D.C., Montañés, P., 2017. Exploring the experience of episodic past, future, and counterfactual thinking in younger and older adults: a study of a Colombian sample. Conscious. Cognit. 51 (Suppl. C), 258–267. https://doi.org/10.1016/j.concog.2017.04.007.

De Brigard, F., Spreng, R.N., Mitchell, J.P., Schacter, D.L., 2015. Neural activity associated with self, other, and object-based counterfactual thinking. NeuroImage 109, 12–26. https://doi.org/10.1016/j.neuroimage.2014.12.075.

Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., et al., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage 31 (3), 968–980. https://doi.org/10.1016/j.neuroimage.2006.01.021.

Goodman, N., 1947. The problem of counterfactual conditionals. J. Philos. 44 (5), 113–128. https://doi.org/10.2307/2019988.

Henke, K., 2010. A model for memory systems based on processing modes rather than consciousness. Nat. Rev. Neurosci. 11 (7), 523. https://doi.org/10.1038/nrn2850.

Hodges, J.R., Graham, K.S., 2001. Episodic memory: insights from semantic dementia. Phil. Trans. Roy. Soc. Lond. 356 (1413), 1423–1434.

Hodges, J.R., Patterson, K., 2007. Semantic dementia: a unique clinicopathological syndrome. Lancet Neurol. 6 (11), 1004–1014.

Irish, M., Piguet, O., 2013. The pivotal role of semantic memory in remembering the past and imagining the future. Front. Behav. Neurosci. 7. https://doi.org/10.3389/fnbeh.2013.00027.

Irish, M., Piolino, P., 2016. Impaired capacity for prospection in the dementias – theoretical and clinical implications. Br. J. Clin. Psychol. 55 (1), 49–68. https://doi.org/10.1111/bjc.12090.

Kahneman, D., Tversky, A., 1982. Variants of uncertainty. Cognition 11 (2), 143–157. https://doi.org/10.1016/0010-0277(82)90023-3.

Kim, H., 2010. Dissociating the roles of the default-mode, dorsal, and ventral networks in episodic memory retrieval. NeuroImage 50 (4), 1648–1657. https://doi.org/10.1016/j.neuroimage.2010.01.051.

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

Lakens, D., 2013. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front. Psychol. 4. https://doi.org/10.3389/fpsyg.2013.00863.

Lewis, D., 1973. Counterfactuals and comparative possibility. J. Phil. Logic 2 (4), 418–446. https://doi.org/10.1007/BF00262950.

McIntosh, A.R., Chau, W.K., Protzner, A.B., 2004. Spatiotemporal analysis of event-related fMRI data using partial least squares. NeuroImage 23 (2), 764–775. https://doi.org/10.1016/j.neuroimage.2004.05.018.

Özbek, M., Bohn, A., Berntsen, D., 2016. Examining the Phenomenology and Functions of Episodic Memories, Counterfactual Thoughts and Future Simulations.

Özbek, M., Bohn, A., Berntsen, D., 2017. Imagining the personal past: episodic counterfactuals compared to episodic memories and episodic future projections. Mem. Cognit. 45 (3), 375–389. https://doi.org/10.3758/s13421-016-0671-2.

Roese, N.J., 1997. Counterfactual thinking. Psychol. Bull. 121 (1), 133–148. https://doi.org/10.1037/0033-2909.121.1.133.

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. https://doi.org/10.1016/j.neuron.2012.11.001.

Spreng, R.N., Gerlach, K.D., Turner, G.R., Schacter, D.L., 2015. Autobiographical planning and the brain: activation and its modulation by qualitative features. J. Cognit. Neurosci. 27 (11), 2147–2157. https://doi.org/10.1162/jocn_a_00846.

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. Cognit. Neurosci. 22 (6), 1112–1123. https://doi.org/10.1162/jocn.2009.21282.

St. Jacques, P.L., Carpenter, A.C., Szpunar, K.K., Schacter, D.L., 2017. Remembering and imagining alternative versions of the personal past. Neuropsychologia. https://doi.org/10.1016/j.neuropsychologia.2017.06.015.

St. Jacques, P.L., De Brigard, F., 2015. Neural correlates of autobiographical memory. In: Addis, D.R., Barense, M., Duarte, A. (Eds.), The Wiley Handbook on the Cognitive Neuroscience of Memory. John Wiley & Sons, Ltd, pp. 265–286. https://doi.org/10.1002/9781118332634.ch13.

Stanley, M.L., Parikh, N., Stewart, G.W., De Brigard, F., 2017. Emotional intensity in episodic autobiographical memory and counterfactual thinking. Conscious. Cognit. 48 (Suppl. C), 283–291. https://doi.org/10.1016/j.concog.2016.12.013.

Szpunar, K.K., 2010. Episodic future thought: an emerging concept. Perspect. Psychol. Sci. 5 (2), 142–162. https://doi.org/10.1177/1745691610362350.

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

Van Hoeck, N., Watson, P.D., Barbey, A.K., 2015. Cognitive neuroscience of human counterfactual reasoning. Front. Hum. Neurosci. 9. https://doi.org/10.3389/fnhum.2015.00420.

Wager, T.D., Keller, M.C., Lacey, S.C., Jonides, J., 2005. Increased sensitivity in neuroimaging analyses using robust regression. NeuroImage 26 (1), 99–113. https://doi.org/10.1016/j.neuroimage.2005.01.011.

Weiler, J.A., Suchan, B., Daum, I., 2010. Foreseeing the future: occurrence probability of imagined future events modulates hippocampal activation. Hippocampus 20 (6), 685–690. https://doi.org/10.1002/hipo.20695.

Yarkoni, T., Poldrack, R.A., Nichols, T.E., Essen, D.C.V., Wager, T.D., 2011. Large-scale automated synthesis of human functional neuroimaging data. Br. J. Pharmacol. 8 (8), 665–670. https://doi.org/10.1038/nmeth.1635.

Yeo, B.T.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., et al., 2011. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J. Neurophysiol. 106 (3), 1125–1165. https://doi.org/10.1152/jn.00338.2011.