NeuroImage 109 (2015) 12–26

Contents lists available at ScienceDirect

NeuroImage

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

Neural activity associated with self, other, and object-based counterfactual thinking

Felipe De Brigard [a,b,c,*], R. Nathan Spreng [d,e], Jason P. Mitchell [f,g], Daniel L. Schacter [f,g]

[a] Department of Philosophy, Duke University, USA
[b] Center for Cognitive Neuroscience, Duke University, USA
[c] Duke Institute for Brain Sciences, USA
[d] Department of Human Development, Cornell University, USA
[e] Human Neuroscience Institute, Cornell University, USA
[f] Center for Brain Science, Harvard University, USA
[g] Department of Psychology, Harvard University, USA

ARTICLE INFO

Article history:
Accepted 29 December 2014
Available online 8 January 2015

Keywords:
Counterfactual thinking
Default mode network
Partial least squares
Mental simulation
Self
Other

ABSTRACT

Previous research has shown that autobiographical episodic counterfactual thinking—i.e., mental simulations about alternative ways in which one's life experiences could have occurred—engages the brain's default network (DN). However, it remains unknown whether or not the DN is also engaged during impersonal counterfactual thoughts, specifically those involving other people or objects. The current study compares brain activity during counterfactual simulations involving the self, others and objects. In addition, counterfactual thoughts involving others were manipulated in terms of similarity and familiarity with the simulated characters. The results indicate greater involvement of DN during person-based (i.e., self and other) as opposed to object-based counterfactual simulations. However, the involvement of different regions of the DN during other-based counterfactual simulations was modulated by how close and/or similar the simulated character was perceived to be by the participant. Simulations involving unfamiliar characters preferentially recruited dorsomedial prefrontal cortex. Simulations involving unfamiliar *similar* characters, characters with whom participants identified personality traits, recruited lateral temporal gyrus. Finally, our results also revealed differential coupling of right hippocampus with lateral prefrontal and temporal cortex during counterfactual simulations involving familiar similar others, but with left transverse temporal gyrus and medial frontal and inferior temporal gyri during counterfactual simulations involving either oneself or unfamiliar dissimilar others. These results suggest that different brain mechanisms are involved in the simulation of personal and impersonal counterfactual thoughts, and that the extent to which regions associated with autobiographical memory are recruited during the simulation of counterfactuals involving others depends on the perceived similarity and familiarity with the simulated individuals.

Introduction

We spend a substantial amount of our lives entertaining mental simulations about situations beyond our temporally and spatially present surroundings.[1] Some of these situations are real but long gone, as when we remember specific episodes from our personal past. But some of these situations are hypothetical, as when we imagine ourselves in a possible future scenario—a kind of mental simulation that has come to be known as *episodic future thinking* (Atance and O'Neill, 2001; for reviews, see Schacter et al., 2012; Szpunar, 2010). The last decade of research in the cognitive neuroscience of both episodic memory and episodic future thinking has revealed striking commonalities between the neural mechanisms underlying both kinds of mental simulations (Okuda et al., 2003; Addis et al., 2007; Hassabis et al., 2007b; Szpunar et al., 2007). Moreover, these studies have revealed that the brain regions commonly engaged by episodic memory and episodic future thinking are part of what it is now known as the brain's default network (DN), a set of functionally connected brain regions 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 (Buckner et al., 2008).

* Corresponding author at: Duke University, 203A West Duke Building, Durham, NC 27708-0743, USA.
E-mail address: felipe.debrigard@duke.edu (F. De Brigard).

[1] The notion of 'simulation' has been traditionally employed as an alternative to the so-called "'theory'–theory" in the literature on mentalizing. However, nowadays the term has acquired a wider scope, becoming essentially a shorthand to refer to the cognitive process of generating coherent imaginations involving scenes (for discussion see, Schacter et al., 2008). In a recent comprehensive volume on mental simulation, and in line with this more general definition, Markman, Klein and Suhr (2008) defined 'simulation' simply as "the act of imagination and generation of alternative realities" (p. vii). Our use of 'simulation' is consistent with this broader definition. We thank an anonymous reviewer for inviting us to clarify this issue.

http://dx.doi.org/10.1016/j.neuroimage.2014.12.075
1053-8119/

Importantly, other kinds of mental simulations about hypothetical scenarios have been shown to engage core regions of the DN as well. For instance, both mental navigation, or our capacity to mentally simulate the spatial surroundings from someone's point of view (Maguire et al., 1998), and mentalizing, or our capacity to mentally simulate another person's perspective (Saxe and Kanwisher, 2003; Mitchell, 2009), have shown to activate core regions of the DN (Spreng et al., 2009). To account for these convergent results, Buckner and Carroll (2007) suggested that core regions of the DN may be commonly activated during these cognitive processes because the DN plays a critical functional role in the generation and support of stimulus-independent simulations in which we project ourselves onto hypothetical situations.

Further support for this view comes from studies on another kind of hypothetical thought which, up until very recently, had not received much attention in the cognitive neuroscience of mental simulation: *counterfactual thinking*, our tendency to think about alternative ways in which things might have occurred in the past but did not (Roese, 1997). Counterfactual thoughts play a central role in human emotion and decision-making, and have been extensively studied in philosophy and linguistics (Goodman, 1947; Lewis, 1973) as well as social psychology and behavioral economics (Roese and Olson, 1995; Mandel, Hilton and Catellani, 2005; Epstude and Roese, 2008).[2] Thus, given how many of our counterfactual simulations involve projecting ourselves onto possible pasts that could have occurred but did not, it is not unreasonable to hypothesize that core regions of the DN would be engaged during counterfactual thinking, which also constitutes a kind of self-generated thought (Andrews-Hanna et al., 2014)

This hypothesis was recently supported by two studies (De Brigard et al., 2013a; Van Hoeck et al., 2013) in which participants engaged in *episodic* counterfactual thinking: counterfactual simulations about alternative ways in which past *personal* (i.e., self-involving) events could have occurred but did not (De Brigard and Giovanello, 2012). Although both studies showed significant engagement of core regions of DN during episodic counterfactual thinking, De Brigard et al. (2013a) also found that the engagement of such regions was modulated by the perceived likelihood of the counterfactual thought. Specifically, they found that the more likely the counterfactual alternative was perceived, the greater the engagement of the DN. Of note, this effect was most clear in certain core regions of the DN, such as the hippocampus and the vMPFC, which were parametrically modulated by perceived likelihood of the episodic counterfactual thought.

Why is there differential engagement of DN regions during episodic counterfactual simulations? One hypothesis is that likely episodic counterfactuals were perceived by the participants as more personally relevant for social interactions. This hypothesis is consistent with much research in the social psychology of counterfactual thinking, suggesting that our tendency to engage in episodic counterfactual simulations may be a goal-oriented cognitive strategy to help us to modify future behavior in the context of social interactions (Johnson and Sherman, 1990; Markman and McMullen, 2003; Epstude and Roese, 2008). Indirect evidence in support of this hypothesis comes from a recent study in which Van Hoeck et al. (2014) found significant overlap in brain activation during false-belief and counterfactual tasks involving possible social interactions. Critically, some of this overlap occurred in temporo-parietal junction and precuneus, which have been associated with the DN. However, this suggestive result only speaks indirectly to the above hypothesis, as they did not employ episodic counterfactual simulations based upon actual autobiographical events, and did not directly manipulate the personal relevance (for the participant) of the characters involved in the vignettes.

On the other hand, the hypothesis that involvement of the DN during autobiographically-based episodic counterfactual thoughts is associated with perceived personal relevance of the content of the simulation for social interaction is also consistent with recent proposals suggesting a critical role of the DN supporting socially relevant goal-oriented cognition (Andrews-Hanna, 2012; Andrews-Hanna et al., 2014). In line with these results, we conjecture that if the involvement of core DN regions during counterfactual thinking is modulated by the personal and social relevance of the simulated event, then it is likely that impersonal and non-socially relevant counterfactual simulations would engage processes outside of the DN, whereas personal and socially relevant episodic counterfactual simulations would mainly engage core regions in the DN.

To explore this general hypothesis, the current study was designed to extend our understanding of the involvement of regions of the DN during personal and socially relevant counterfactual simulations in three ways. First, this study investigates whether or not core regions of the DN are engaged during mental simulations of impersonal counterfactual thoughts pertaining to either objects or people other than oneself. Participants were asked to simulate counterfactuals that either involved themselves, other people, or objects. Given recent neuroimaging results showing significant overlap in DN regions during episodic memory and theory of mind tasks (Spreng and Grady, 2010; Mitchell, 2009), and greater involvement of DN during simulations that involve primarily autobiographical details rather than tasks involving non-autobiographical processing of objects (Addis et al., 2007; Addis et al., 2009; Hassabis et al., 2007b), we expected to see greater involvement of DN during person-based (i.e., self and other) relative to object-based counterfactual simulations. Indeed, two recent fMRI studies exploring neural correlates of semantic evaluation of non-autobiographical hypothetical and counterfactual statements show relatively little involvement of DN regions (Nieuwland, 2012; Kulakova et al., 2013), further suggesting that object-based counterfactual simulations may primarily recruit processes outside the DN.

On the other hand, given previous research showing differential MPFC recruitment for self- relative to other-based mental simulations (Denny et al., 2012; Hassabis et al., 2014; Wagner et al., 2012), we also expected to find differences in prefrontal activation between self versus other-based counterfactual simulations. Thus, a second way in which the current study seeks to investigate the involvement of DN in personal and socially relevant counterfactual simulations, is by way of contrasting the recruitment of DN regions during personal and socially relevant counterfactual thoughts (i.e., self-based) versus impersonal and non-socially relevant counterfactual simulations (i.e., object-based), on the one hand, and impersonal yet socially relevant counterfactual simulations (i.e., other-based), on the other.

Finally, since certain DN regions recruited during theory of mind tasks—e.g., MPFC, anterior cingulate cortex (ACC), and hippocampus—are differentially engaged depending on whether or not the simulated character is personally known (i.e., familiar) and/or perceived to be similar in personality by the participant (Mitchell et al., 2006; Krienen et al., 2010), we also expected to find neural differences when other-based counterfactuals involved either familiar and/or similar characters.

[2] Although related, the expression "counterfactual" as it is used in psychology does not square precisely with the way in which the notion of "counterfactual" is used in philosophy and linguistics. Philosophers and linguists tend to be interested in the semantics of counterfactual statements; that is, they seek to understand how to assign truth values to conditional statements whose antecedents are false by virtue of referring to (or, less controversially, expressing) events that are contrary-to-fact. Psychologists, on the other hand, understand "counterfactual" as a psychological term, employed in reference to the cognitive process of thinking about alternative ways in which a thought-to-be-true fact could have occurred differently. As such, it is possible for a counterfactual thought, understood psychologically, to be semantically factual. If I think "Had I left the door open, the dog wouldn't have left", because I wrongly believe that I closed the door, I am entertaining a counterfactual thought that may not qualify as a counterfactual, in the semantic sense, because the antecedent could very well be true, namely if I did, in fact, leave the door open. Moreover, early canonical uses of the term "counterfactual simulation" restricted its use to imagined alternative ways in which *past* events could have occurred (Kahneman and Miller, 1986; Roese, 1997; McMullen, 1997). Now, though, psychologists tend to use the term "counterfactual simulation" in a more encompassing way, referring to the process of actively constructing and maintaining a mental image or scene in which one or several known facts are altered. Our use of the term "counterfactual simulation" is consistent with this latter construal, although we are sensitive to the fact that, semantically, counterfactual simulations may best be called hypothetical (De Brigard, 2014). We thank an anonymous reviewer for inviting us to clarify this issue.